

---

# Quantitative Plate Tectonics



---

Antonio Schettino

# Quantitative Plate Tectonics

Physics of the Earth - Plate  
Kinematics – Geodynamics

Antonio Schettino  
School of Science & Technology – Geology Division  
University of Camerino  
Camerino, Italy

Additional material can be downloaded from <http://extras.springer.com>

ISBN 978-3-319-34594-9      ISBN 978-3-319-09135-8 (eBook)  
DOI 10.1007/978-3-319-09135-8  
Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2015

Softcover reprint of the hardcover 1st edition 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

*Cover image:* part of a Landsat 5 image of Sierra de Velasco Mountains of northern Argentina (February 28, 1985). From the collection: Earth as Art 3. Credit: U.S. Geological Survey

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*To Annalisa,  
my inspiring goddess*



---

## Preface

I consider *plate tectonics* as the Earth Sciences discipline that describes the geodynamic evolution of the lithosphere, coupled with the underlying “fluid” mantle, over long time intervals, of the order of million years. Such a description, which includes plate kinematics, mantle dynamics, and the geology of plate boundaries, is based on a variety of data from marine geophysics, paleomagnetism, seismology, structural geology, stratigraphy, paleontology, and geochemistry. Therefore, although traditionally plate tectonics has been considered either as synonymous with plate kinematics (by most geophysicists) or a discipline describing the structural geology of plate boundaries (by many geologists), I will try to present an integrated approach to this science, introducing concepts from most of the research fields listed above. In fact, despite that the modern society drives young scientists toward exasperated specialization, it is my firm opinion that a true expertise in plate tectonics should be based on a *holistic* view of the geological processes.

This textbook evolved from a series of courses that I taught over the last years at the University of Camerino, Italy, to students enrolled in the MSc programme in Geoenvironmental Resources and Risks. These students had a background in geology, geography, engineering, or physics, a level ranging from advanced undergraduate to graduate, and came from four continents. For many geology, geophysics, and environmental engineering students, knowledge of plate tectonics is generally *qualitative* and limits to a few consolidated principles and to some rudiments about plate boundary processes. This book was born from the idea that a number of students could be stimulated to undertake a more thorough study of plate tectonics after having acquired basic skills in introductory courses. For these people, attaining an in-depth knowledge of the physics of plate tectonic processes through a rigorous approach based on mathematical methods will be a pleasant adventure, not a boring exercise. In fact, any student that strives for understanding the nature of the forces that drive the large-scale geological processes is naturally led to consider mathematics as a fundamental method for describing the natural processes in their generality, not only as a tool for solving practical problems. This book can also be useful for researchers specialized in one of the several disciplines of geosciences, who have the necessity to learn techniques of plate tectonic modelling or simply desire to expand their knowledge. Unfortunately, with the exception of the classic volume published by Le Pichon et al. in 1973, there are no advanced books where a student can learn quantitative methods of plate

tectonics and find a unified description of the laws that govern the motion of tectonic plates. This book attempts to fill the gap by first exploring the principles of plate kinematics in the first part (Chaps. 1, 2, 3, 4, 5, and 6), then trying to link plate motions to physical processes occurring in the lithosphere and the mantle in the second part (Chaps. 7, 8, 9, 10, 11, 12, 13, and 14). In order to be self-contained, the book includes three chapters (Chaps. 8, 9, and 10) that illustrate the basic principles of seismology, because this discipline represents a fundamental source of data for plate tectonics. Throughout the book, it is assumed that the reader has an adequate background in geology, geochemistry, and classic physics, while skills in computer programming are required for solving some exercises. With a few exceptions, reading this book does not require more mathematical background than the customary undergraduate courses in advanced calculus and vector analysis. As an aid to reading, two electronic appendices introduce vector analysis and algorithms are included at the end of this book. Finally, a basic understanding of plate tectonics at the level of the classic book of Cox and Hart (1986) is desirable. Some exercises have been designed to be solved with the help of professional software, either freeware or commercial. These software tools include two free computer programs that I designed for the analysis of marine magnetic anomalies and for making plate tectonic reconstructions, respectively *Magan* and *PlaKin*. Readers can freely download the software and the solutions to the exercises from <http://extras.springer.com> either to learn some techniques or for their own research purposes.

The special emphasis I give to computer methods shows through some chapters of this book and is clearly a consequence of my heterogeneous scientific background. Although I studied theoretical physics, my passion for computer programming led me to start working as a software developer and consultant in Milan, Italy, soon after my graduation. However, even in those days I matured a strong scientific interest in the application of advanced algorithms to geosciences, in particular to plate kinematics. During that period, I designed and developed *PCME* (Paleo–Continental Map Editor) (Schettino 1998), an interactive computer program for making plate reconstructions. This event led me to get in touch with C. R. Scotese, a scientist who pioneered the application of computer methods to plate kinematics. At that time he was at the University of Texas at Arlington, and I started collaborating with him to the construction of a new atlas of plate tectonic reconstructions for the Mesozoic and the Cenozoic in the context of the Paleomap Project, as well as to the implementation of a series of advanced software tools for plate tectonic modelling. Starting from the beginning of the new millennium, my research interests focused on paleomagnetism, marine geophysics, plate kinematics at global and regional scale, and on the dynamics of subduction. The general conclusion at which I arrived during my studies, which is now a guiding principle in my approach to the analysis of plate tectonic processes, can be summarized as follows. For time intervals of several million years, tectonic plates move at constant angular velocity about stationary axes of rotation in any geocentric reference frame fixed with respect to the Earth's spin axis. Therefore, conditions of dynamic equilibrium must normally exist between driving and resistive

torques exerted on the lithosphere. Such a global equilibrium is broken when new plate boundaries form, or existing ones become extinct, so that a new system of tectonic plates is established. Then, the plates of this new system start moving about new, generally different, rotation axes and with different angular velocities. New plate boundaries always form by propagation of lithospheric discontinuities from a source region toward existing active boundaries, where they will give rise to additional triple junctions. Occasionally, plate motions proceed in non-equilibrium conditions, because active asthenospheric flows driven by horizontal pressure gradients in the mantle exert an excess basal drag on the overlying lithosphere, determining significant variations of angular velocity and accelerated motion. In this instance, high spreading and subduction rates accompany the plate motions. The ultimate cause of such episodes of non-equilibrium plate kinematics is the presence of upper mantle thermo-chemical heterogeneities, in particular mantle plumes and subducted slabs. Plate motions that proceed in conditions of perturbed equilibrium are associated with spectacular and sometimes unexplained geological phenomena, but these events are uncommon in the history of global plate motions and must be viewed as the exception rather than the rule. Therefore, the equilibrium between driving and resistive torques exerted on the lithosphere, and the resulting invariance of the angular velocity vectors of tectonic plates for long time intervals, should be considered as one of the most fundamental laws of plate tectonics.

I would like to thank my students of the Earth Physics course, a.a. 2013/2014, for the many errors found in a draft of Chap. 2 and for their stimulating questions. I would also thank colleagues with whom I worked for several years on plate kinematics, especially C. R. Scotese and Eugenio Turco. I am honored to have learnt so many things from these people. I am also grateful to Giorgio Ranalli (Carleton University), Marco Ligi (ISMAR–CNR Bologna), and Eugenio Turco (University of Camerino) for the help I received through their constructive criticism and useful suggestions during the review of this book. If errors persist in this edition, it is not their fault but a consequence of my negligence. Finally, I am indebted to the editor, Petra van Steenbergen, and the editorial staff of Springer for their assistance.

March 2014  
Camerino, Italy

Antonio Schettino

---

## References

- Cox A, Hart RB (1986) Plate tectonics: how it works. Blackwell Scientific Publications, Palo Alto, 392 pp
- Le Pichon X, Francheteau J, Bonnin J (1973) Plate tectonics. Elsevier, Amsterdam, 300 pp
- Schettino A (1998) Computer aided paleogeographic reconstructions. *Comput Geosci* 24(3):259–267



---

# Contents

## Part I Plate Kinematics

<b>1</b>	<b>Composition of the Crust and the Mantle</b>	3
1.1	Crust and Mantle Minerals	3
1.2	Continental Crust	4
1.3	Oceanic Crust	7
1.4	Lithospheric Mantle	14
1.5	Asthenosphere	16
1.6	Transition Zone	20
1.7	Lower Mantle	22
	Problems	25
	References	25
<b>2</b>	<b>Plate Motions</b>	29
2.1	The Continuum Mechanics Representation	29
2.2	Euler's Theorem and Rigid Rotations	30
2.3	Reference Frames	34
2.4	Plate Boundaries	36
2.5	Triple Junctions	45
2.6	Tectonic Elements	49
2.7	Plate Circuits and Rotation Models	54
2.8	Plate Tectonic Reconstructions	59
2.9	Current Plate Motions	69
	Problems	77
	References	78
<b>3</b>	<b>Magnetization and Magnetic Minerals</b>	81
3.1	Electric Currents	81
3.2	Magnetic Moments	83
3.3	Maxwell's Equations for the Magnetic Field	85
3.4	Magnetization	87
3.5	Magnetic Properties of Rocks	92
	Problems	101
	References	101
<b>4</b>	<b>The Geomagnetic Field</b>	103
4.1	Source of the Main Geomagnetic Field	103
4.2	The Geodynamo	107

4.3	Secular Variation of the Core Field . . . . .	112
4.4	Polarity Inversions, Chrons, and Geomagnetic Timescales . . . . .	115
4.5	Ionosphere and Magnetosphere . . . . .	118
4.6	Crustal Magnetic Field . . . . .	123
4.7	The Geomagnetic Potential . . . . .	126
4.8	Spherical Harmonic Expansion of the Geomagnetic Field and the IGRF . . . . .	130
	Problems . . . . .	140
	References . . . . .	140
<b>5</b>	<b>Marine Magnetic Anomalies . . . . .</b>	<b>143</b>
5.1	Magnetic Anomalies . . . . .	143
5.2	Acquisition and Pre-processing of Raw Magnetic Data . . .	145
5.3	Levelling Techniques . . . . .	148
5.4	Modelling of Marine Magnetic Anomalies . . . . .	149
5.5	Forward Modelling Procedures . . . . .	157
5.6	Construction of Isochron Maps . . . . .	164
5.7	Determining Finite Rotations . . . . .	168
5.8	Data Transformations . . . . .	170
	Problems . . . . .	175
	References . . . . .	176
<b>6</b>	<b>Paleomagnetism and Earth History . . . . .</b>	<b>177</b>
6.1	Néel's Theory of Single-Domain TRM . . . . .	177
6.2	Paleomagnetic Sampling and Cleaning Procedures . . . . .	182
6.3	Paleomagnetic Directions . . . . .	185
6.4	Paleopoles and Apparent Polar Wander Paths . . . . .	188
6.5	Paleomagnetic Reference Frames . . . . .	199
6.6	True Polar Wander . . . . .	203
6.7	Velocity Fields and Acceleration Fields During the Cenozoic and the Mesozoic . . . . .	205
6.8	Non-dipole Paleomagnetic Fields . . . . .	217
	Problems . . . . .	221
	References . . . . .	221

## **Part II Dynamics of the Lithosphere and the Mantle**

<b>7</b>	<b>Stress and Strain . . . . .</b>	<b>227</b>
7.1	The Stress Tensor . . . . .	227
7.2	Displacement Fields and Strain . . . . .	232
7.3	Cauchy Momentum Equation . . . . .	235
7.4	Basic Rheological Models and Constitutive Equations . . .	237
	Problems . . . . .	244
	References . . . . .	244
<b>8</b>	<b>Elasticity of the Earth . . . . .</b>	<b>245</b>
8.1	Hooke's Law . . . . .	245
8.2	Equations of Motion for Elastic Media . . . . .	246

8.3	Seismic Waves.....	247
8.4	Seismic Energy.....	253
	Problems.....	255
	References.....	255
<b>9</b>	<b>Seismic Rays.....</b>	<b>257</b>
9.1	The Eikonal Equation.....	257
9.2	Geometrical Spreading.....	259
9.3	Snell's Law.....	262
9.4	1-D Velocity Models.....	263
9.5	Travel Time Curves.....	268
9.6	Low-Velocity Zones.....	270
9.7	Seismic Phases Nomenclature.....	272
	Problems.....	278
	References.....	278
<b>10</b>	<b>Earthquakes.....</b>	<b>279</b>
10.1	Reid's Model.....	279
10.2	Faults and Focal Mechanisms.....	284
10.3	Moment Tensor.....	292
10.4	Earthquake Magnitude.....	297
10.5	Gutenberg-Richter Law.....	297
	Problems.....	298
	References.....	299
<b>11</b>	<b>Seismic Deformation of the Lithosphere.....</b>	<b>301</b>
11.1	Kostrov's Formula.....	301
11.2	The Asymmetric Strain Tensor.....	303
11.3	Global Pattern of Lithospheric Deformation.....	304
11.4	Bending of Slabs.....	306
	Problems.....	314
	References.....	315
<b>12</b>	<b>Heat Flow and Thermodynamics of the Lithosphere.....</b>	<b>317</b>
12.1	Fourier's Law.....	317
12.2	Continental Geotherms.....	320
12.3	Non-steady State Heat Conduction.....	323
12.4	Cooling of the Oceanic Lithosphere.....	326
12.5	Driving Mechanism of Plate Tectonics: Slab Pull and Ridge Push.....	332
	Problems.....	335
	References.....	336
<b>13</b>	<b>Flow and Fluid Behaviour of the Mantle.....</b>	<b>337</b>
13.1	Continuity Equation.....	337
13.2	Navier-Stokes Equations.....	339
13.3	Energy Balance.....	341
13.4	Flow in the Asthenosphere.....	344

13.5	Stream Functions and Two-Dimensional Flows . . . . .	348
13.6	Rayleigh-Bénard Convection . . . . .	354
	Problems . . . . .	361
	References . . . . .	361
<b>14</b>	<b>Gravity Field of the Earth . . . . .</b>	<b>363</b>
14.1	Gravity Field and Geopotential . . . . .	363
14.2	Spherical Harmonic Expansion of the Geopotential: The Geoid . . . . .	364
14.3	Geoid and Ellipsoid . . . . .	365
14.4	MacCullagh's Formula . . . . .	370
14.5	Gravity Measurements and Reduction of Gravity Data . . .	372
14.6	Isostasy and Dynamic Topography . . . . .	375
	Problems . . . . .	377
	References . . . . .	377
	<b>Appendix 1: Vector Calculus . . . . .</b>	<b>379</b>
A1.1	Scalar Fields . . . . .	379
A1.2	Vector Fields . . . . .	379
A1.3	Integrals Theorems . . . . .	382
A1.4	Spherical Coordinates . . . . .	384
A1.5	Index Notation . . . . .	385
	<b>Appendix 2: Algorithms and Data Structures . . . . .</b>	<b>387</b>
A2.1	Algorithms . . . . .	387
A2.2	Data Structures . . . . .	388
A2.3	Trees . . . . .	392
	<b>Index . . . . .</b>	<b>395</b>

---

## Part I

### Plate Kinematics



## Abstract

The external layers of the solid Earth, from the crust to the lower mantle, are the main actors involved in the plate tectonics drama. In this chapter, I describe their chemical composition and introduce the principal geodynamic processes occurring within and between these layers.

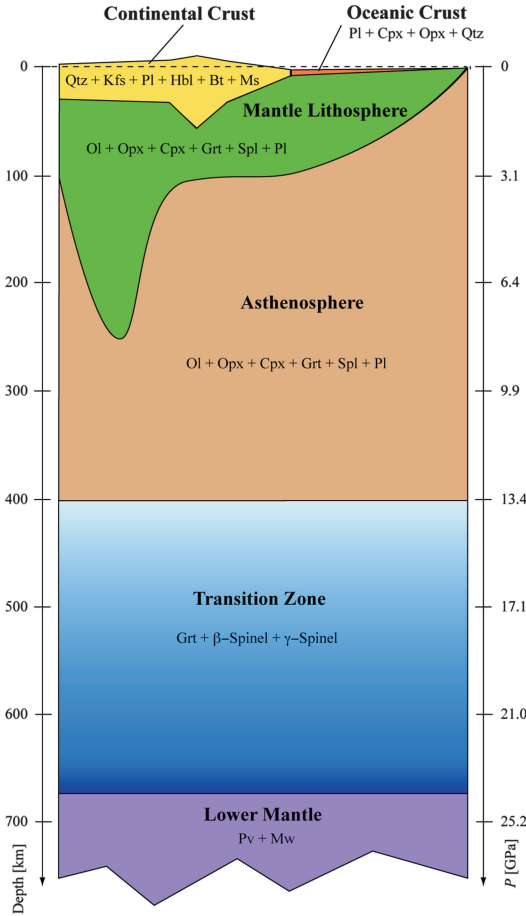
## 1.1 Crust and Mantle Minerals

Plate tectonics focuses on the dynamics and kinematics of the most external layers of the solid Earth: the relatively thin *oceanic crust* (7–10 km thickness), the *continental crust* (5–70 km thickness), the *mantle lithosphere*, whose lower boundary can be found at depths ranging between 80 and 250 km, the underlying *asthenosphere* (up to 410 km depth), the *transition zone* (410–670 km depth), and the very huge *lower mantle*, which extends to a depth of  $\sim 2,900$  km (Fig. 1.1). All these rock layers are formed by solid-state mixtures of minerals that are chemically and structurally stable only within determined intervals of pressure and temperature. Rocks can eventually contain liquid phases in the existing pores between grains or in cracks.

In general, chemical composition, crystalline structure, and the physical state of Earth materials change from point to point, so that rocks are strongly heterogeneous systems. However, any rock system can be ideally resolved into a finite number of *phases* that are physically and chemically homogeneous.

For example, seismology studies and research on high-temperature (*HT*) and high-pressure (*HP*) minerals indicate that the upper mantle mostly consists of two phases:  $\text{Mg}_2\text{SiO}_4$  (the forsterite end-member of olivine) and  $\text{Mg}_2\text{Si}_2\text{O}_6$  (orthopyroxene). The chemical composition of mineral assemblages is usually described by the so-called *components*, which represent the minimum number of chemical formulae that are needed to describe the set of phases composing the rock. For example, both olivine and orthopyroxene can be represented by a mixture of  $\text{MgO}$  and  $\text{SiO}_2$ , because  $\text{Mg}_2\text{SiO}_4 = 2\text{MgO} + \text{SiO}_2$  and  $\text{Mg}_2\text{Si}_2\text{O}_6 = 2\text{MgO} + 2\text{SiO}_2$ . We emphasize that this decomposition is arbitrary and represents only a useful way to describe the *chemical* composition of a rock through its constitutive elements, independently from the real crystalline structure of the mineral phases included in the solid state mixture.

A basic mineralogical characterization and stratigraphy of the external layers of the solid Earth is illustrated in Fig. 1.1, while composition, density, and other physical parameters of the main crustal and mantle minerals are listed in Table 1.1. The meaning of these mechanical



**Fig. 1.1** Geometry and mineral composition of the crust and the mantle (Mineral abbreviations are explained in Table 1.1.  $P$  is the lithostatic pressure)

parameters will be discussed in detail in Chap. 8. An apparent feature of the subdivision shown in Fig. 1.1 is the trend towards a more complicated chemical composition and geometry as we approach the Earth's surface. This greater complexity is only partially real, and mainly arises from our increased and often direct knowledge of the most external layers, whereas starting from the transition zone we must rely on a combination of theoretical modelling (both thermodynamic and geochemical), experimental petrology laboratory results, and indirect geophysical (mainly seismological) constraints. However, it is possible to affirm that most of the complicate geologic structures that we observe at crustal scale, as well as the considerable lateral chemical differentiation, are unique features that

result from the global plate tectonics process, which accompanied the origin and evolution of life on the Earth during the last 2.5–3 Ga (billion year b.p.).

The stratification shown in Fig. 1.1 was originated during the last stages of formation of the solid Earth by a process of gravitational separation of materials according to their melting points and densities. In the next pages, we shall build a continuum mechanics representation of these layers, in order to establish a rigorous quantitative basis for the description of the geophysical entities that we call *tectonic plates*.

## 1.2 Continental Crust

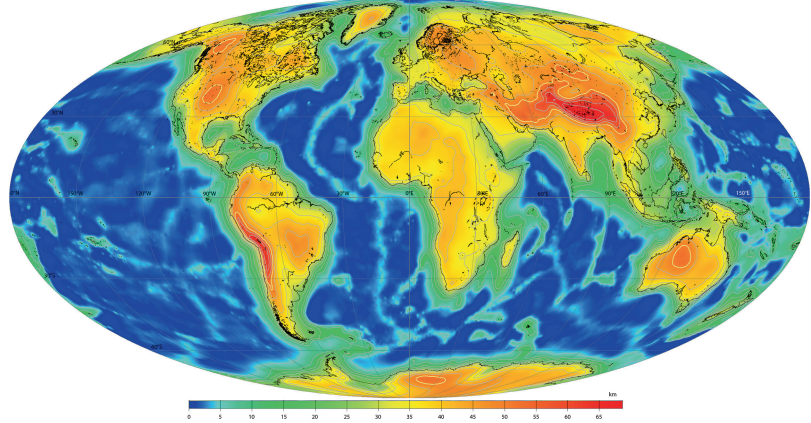
The major part of the continental crust most probably formed between 4 and 2 Ga during two main episodes of differentiation at  $\sim 1.9$  Ga ago and  $\sim 3.3$  Ga (Hawkesworth and Kemp 2006). Geochemical modeling suggests that at  $\sim 4$  Ga basaltic magma, having a composition similar to a mixture of 92 % present day oceanic island arc basalts (IAB) and 8 % ocean island basalts (OIB), was extracted from the primitive mantle. Cooling of this magma ocean led to the formation of an early relatively thin (30–45 km, Herzberg and Rudnick 2012) basaltic crust envelope of the entire globe. Further differentiation occurred sometime later, at  $\sim 3.3$  and  $\sim 1.9$  Ga, and consisted into localized processes of re-melting of the original basaltic layer, with formation of a  $\sim 13$  km thick and lighter *upper continental crust* and an underlying stratum having much greater thickness and density. The latter became gravitationally unstable as a consequence of phase transitions that produced dense minerals such as garnet. Therefore, the original *deep continental crust* resulted from further gravitational differentiation, with foundering of a relevant high-density part of the residual layer into the mantle. After this initial phase, concentrated in two pulses during the Meso-Archean and the Paleo-Proterozoic, slow and more regular growth of the continental crust occurred at the expenses of the surrounding oceanic crust, and was essentially driven by plate tectonic processes, in particular by arc magmatism.

**Table 1.1** Main crust and mantle minerals and mechanical parameters at *STP* conditions ( $T = 25^\circ\text{C}$ ,  $P = 0.1\text{ MPa}$ )

Mineral	Sym	Formula	$\rho$ [ $\text{Kg m}^{-3}$ ]	$v_p$ [ $\text{ms}^{-1}$ ]	$v_s$ [ $\text{ms}^{-1}$ ]	$K_S$ [GPa]	$\mu$ [GPa]	Ref.
Quartz	Qtz	$\text{SiO}_2$	2,646	5,980	4,040	37	43	2
<i>K</i> -Feldspar	Kfs	$\text{KAlSi}_3\text{O}_8$	2,543	6,080	3,290	57	28	2
Plagioclase	Pl	$\text{NaAlSi}_3\text{O}_8 - \text{CaAl}_2\text{Si}_2\text{O}_8$	2,751	6,820	3,580	81	35	2
Hornblende	Hbl	$(\text{Ca},\text{Na})_{2-3}(\text{Mg},\text{Fe},\text{Al})_5\text{Si}_6(\text{Al},\text{Si})_2\text{O}_{22}(\text{OH})_2$	3,224	7,090	4,030	92	52	2
Biotite	Bt	$\text{K}(\text{Mg},\text{Fe})_3\text{AlSi}_3\text{O}_{10}(\text{OH})_2$	2,765	5,210	2,770	47	21	2
Muscovite	Ms	$\text{KAl}_2(\text{Si}_3\text{Al})\text{O}_{10}(\text{OH})_2$	2,804	5,520	3,180	48	28	2
Clinopyroxene	Cpx	$\text{Ca}(\text{Mg},\text{Fe})\text{Si}_2\text{O}_6$	3,277 + 380XFe	7,710	4,370	113 + 7XFe	67 – 6XFe	1
Orthopyroxene	Opx	$(\text{Mg},\text{Fe})\text{SiO}_3$	3,204 + 799XFe	7,870	4,700	104	77 – 24XFe	1
Olivine	Ol	$(\text{Mg},\text{Fe})_2\text{SiO}_4$	3,222 + 1,182XFe	8,310	4,800	129	82 – 31XFe	1
Garnet	Grt	$(\text{Fe},\text{Mg})_3\text{Al}_2(\text{SiO}_4)_3$	3,562 + 758XFe	8,657	4,821	175 + 1XFe	90 + 8XFe	1
Garnet	Grt	$\text{Ca}_3(\text{Al},\text{Fe})_2(\text{SiO}_4)_3$	3,595 + 265XFe	9,250	5,450	169 – 11XFe	104 – 14XFe	1
$\beta$ -Spinel	$\beta$ -Sp	$(\text{Mg},\text{Fe})_2\text{SiO}_4$	3,472 + 1,240XFe	9,410	5,480	174	114 – 41XFe	1
$\gamma$ -Spinel	$\gamma$ -Sp	$(\text{Mg},\text{Fe})_2\text{SiO}_4$	3,548 + 1,300XFe	9,530	5,540	184	119 – 41XFe	1
Spinel	Spl	$\text{FeCr}_2\text{O}_4$	5,090	8,209	4,542	203	105	3
Spinel	Spl	$\text{MgAl}_2\text{O}_4$	3,585	9,670	5,440	197	108	3
Spinel	Spl	$\text{FeAl}_2\text{O}_4$	4,280	8,630	4,400	210	84	3
Perovskite	Pv	$(\text{Mg},\text{Fe})\text{SiO}_3$	4,104 + 1,070XFe	10,066	5,743	266	153	1
Perovskite	Pv	$\text{CaSiO}_3$	4,130	9,763	5,501	227	125	1
Magnesiowüstite	Mw	$(\text{Mg},\text{Fe})\text{O}$	3,583 + 2,280XFe	7,732	4,425	163 – 8XFe	131 – 77XFe	1

$\rho$  is density.  $v_p$  and  $v_s$  are, respectively, the  $P$  and  $S$  waves velocity.  $K_S$  is the adiabatic bulk modulus.  $\mu$  is the rigidity modulus.  $X_{Fe}$  is the molar fraction of iron:  $X_{Fe} = [\text{Fe}]/([\text{Mg}] + [\text{Fe}])$ . Ref.: [1] = Duffy and Anderson (1989); [2] = Hacker and Abers (2004); [3] = Lee (2003).  $v_p$  and  $v_s$  of mantle minerals are either from Anderson (1989) or determined by elastic moduli and density

**Fig. 1.2** Global crustal thicknesses in the model SECT (After Bagherbandi 2012)



The continental crust is characterized by lateral and vertical chemical heterogeneity, as well as by variable thickness. These rocks have a bulk low-density ( $\rho \sim 2,700 \text{ kg m}^{-3}$ ) “andesitic” (intermediate) composition. This implies that a description of the average chemical composition of the continental crust in terms of oxide components coincides with that of a typical andesite, an extrusive rock that is characteristic of subduction zones and is made principally by plagioclase, mafic (that is, having high Mg and Fe content) minerals (hornblende, clinopyroxene, and orthopyroxene), and possibly quartz and biotite (see Table 1.1). The thickness varies from a few kilometers along the ultra-thinned continent-ocean boundaries (COBs) to more than 80 km in regions of continental collision. Seismology studies indicate that the greatest crustal thicknesses can be found in the Himalayas ( $\sim 80 \text{ km}$ ) and along the western Cordillera of the Andes (70–74 km). Figure 1.2 shows the recent  $1^\circ$  resolution global model proposed by Bagherbandi (2012), which furnishes a good representation of the crustal geometry especially in the continental areas.

In general, at least three layers with different petrologic and physical properties are necessary to describe accurately the continental crust. The upper continental crust is formed by sedimentary and granitic rocks, with a bulk granodioritic composition. In terms of components, it includes 66.6 %  $\text{SiO}_2$ , 15.4 %  $\text{Al}_2\text{O}_3$ , 5.0 %  $\text{FeO}_T$  (total Fe computed as FeO), 3.6 % CaO, 3.3 %  $\text{Na}_2\text{O}$ , and 2.8 %  $\text{K}_2\text{O}$  (Rudnick and Gao 2003). The deep

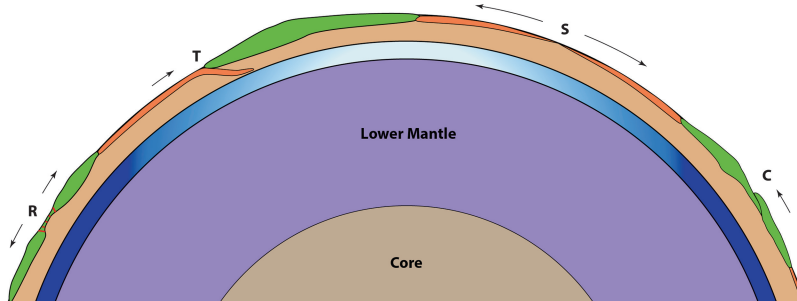
continental crust can be generally divided into *middle* (between 10 and 20 km depth) and *lower crust* (below 20–25 km). The former is dominated by amphibolites (metamorphic rocks which have experienced temperature in excess of  $500^\circ\text{C}$  and pressure between 0.3 and 0.8 GPa) and has an overall trace-element pattern that is very similar to the upper crust, which indicates that also this layer is formed by products of the original phase of intra-crustal differentiation. Its bulk composition is intermediate and includes 63.5 %  $\text{SiO}_2$ , 15.0 %  $\text{Al}_2\text{O}_3$ , 6.0 %  $\text{FeO}_T$ , 5.3 % CaO, 3.6 % MgO, 3.4 %  $\text{Na}_2\text{O}$ , and 2.3 %  $\text{K}_2\text{O}$  (Rudnick and Gao 2003). Therefore, with respect to the upper crust this layer has lower  $\text{SiO}_2$  and  $\text{K}_2\text{O}$  concentrations and higher FeO, MgO, and CaO concentrations. The lower continental crust is believed to be mainly formed by granulites, metamorphic rocks that are associated with *HT* ( $T > 650^\circ\text{C}$ ) and low (*LP*) to moderate pressure conditions ( $0.3 < P < 1.2 \text{ GPa}$ ). The *P*-waves seismic velocity is subject to relevant lateral variations, but is generally high ( $v_P \sim 7 \text{ km s}^{-1}$ ), suggesting a dominance of mafic granulite and/or amphibolite lithologies. However, continental arcs and some Archean cratons are characterized by slower seismic velocity, which is indicative of more evolved composition (that is, higher  $\text{SiO}_2$  content). The lower crust bulk composition includes 53.4 %  $\text{SiO}_2$ , 16.9 %  $\text{Al}_2\text{O}_3$ , 8.6 %  $\text{FeO}_T$ , 9.6 % CaO, 7.2 % MgO, 2.7 %  $\text{Na}_2\text{O}$ , and 0.8 %  $\text{TiO}_2$  (Rudnick and Gao 2003).

### 1.3 Oceanic Crust

We now consider the *oceanic crust*. With respect to the continental counterpart, this is a thin layer (Fig. 1.1) with different and much more homogeneous composition, and different mechanical properties (in terms of density and seismic velocities). It formed much later at the expenses of the originary undifferentiated basaltic layer mentioned above and can be considered as the most evident product of the global plate tectonics process, which is historically known as *Wilson cycle* (Dewey and Burke 1974). This process governs the formation of new continents by aggregation of continental masses after episodes of collision (orogenic phases). It also governs their splitting and subsequent dispersal, with formation of new oceanic seaways, as a consequence of extensional forces. Figure 1.3 shows schematically the four constitutive elements of the Wilson cycle. A quantitative description of these processes is the ultimate objective of this book. The Wilson cycle implies that oceanic crust is continuously accreted at *spreading centers* that are localized along the *mid-ocean ridges*, and continuously destroyed at *subduction zones*, where it bends downwards and sinks passively into the asthenosphere. The morphological expression of these bending lines is represented by the *oceanic trenches*. The global age distribution of the mod-

ern oceanic sea floor (Müller et al. 2008) suggests that a complete recycling of the oceanic crust may take more than 200 Myrs. Recent estimates suggest  $\sim 3$  Ga as a probable starting time of this process (Shirey and Richardson 2011; Dhuime et al. 2012), although other authors claim that plate tectonics did not initiate during the Archean (e.g., Hamilton 1998). Therefore, the starting time of the Wilson cycle could be as young as 2.5 Ga.

New oceanic crust is generated by crystallization of a mid-ocean ridge basaltic (MORB) magma in the empty space that is continuously created at a spreading center from the horizontal displacement of two oceanic plates that are moving apart. Several lines of evidence suggest that mid-ocean ridges have deep “roots” in the mantle asthenosphere. Although this layer is made by solid-state peridotite rocks having considerable density (between 3,400 and 3,500 kg m<sup>-3</sup>), we shall prove in Chap. 13 that it has a distinct fluid behaviour, with relatively low viscosity of the order of 10<sup>20</sup> Pa s. The lateral divergent motion of two oceanic plates induces passive upwelling of asthenosphere material, because the separating plates exert viscous drag on the underlying hot mantle, determining the continuous formation of void that must be filled by a vertical flow. The asthenospheric upwelling beneath mid-ocean ridges has a lower bound velocity of the order  $v \sim 10$  mm yr<sup>-1</sup> and can be traced downwards to a depth of 250–300 km on the basis of seismic



**Fig. 1.3** The Wilson cycle paradigm. Oceanic crust and underlying mantle lithosphere are shown in orange. Continental crust and underlying mantle lithosphere are shown in green. The asthenosphere and transition zone are shown in light brown and blue respectively. Four distinct interconnected processes contribute to the Wilson cycle: *R Rifting*, where continents are split apart by extensional

force fields, *S Spreading*, where new oceanic crust is created by cooling of basaltic magma associated with asthenosphere upwelling; *T Subduction*, where old oceanic crust and mantle lithosphere sink into the asthenosphere as a consequence of gravitational instability; and *C Collision*, where continental masses join together and form mountain belts after complete closure of the intervening oceans

tomography techniques (Ritsema et al. 2004). The velocity of this flow and the typical thermal diffusivity of peridotite ( $\kappa \cong 0.6 \text{ mm}^2 \text{ s}^{-1}$  at  $T = 1,400^\circ\text{K}$ ,  $p = 7 - 10 \text{ GPa}$ ) prevent substantial transfer of heat to the surrounding mantle, because for a path length  $L = 300 \text{ km}$  the dimensionless quantity  $\nu L/\kappa \cong 1.59 \times 10^2$  (*Peclet number*), which measures the relative importance of advection with respect to conduction, is much greater than unity. Therefore, we can consider this as an adiabatic process, which is dominated by the advection of heat rather than by thermal conduction.

Let us consider a small rising volume element of the asthenosphere beneath a mid-ocean ridge. During its ascent, the temperature  $T$  and the ambient pressure  $p$  change in such a way that no heat is lost or gained ( $dQ = 0$ ). This process is generally considered to be reversible to a first approximation, so that  $dQ = TdS$  and the entropy is invariant too ( $dS = 0$ , isentropic process). In these conditions, it is simple to determine how the temperature changes as the volume element moves towards the surface. By the cyclic relations of thermodynamics, we have that:

$$\left(\frac{\partial T}{\partial p}\right)_S = -\left(\frac{\partial T}{\partial S}\right)_p \left(\frac{\partial S}{\partial p}\right)_T \quad (1.1)$$

Furthermore, Maxwell's relations require that:

$$\left(\frac{\partial S}{\partial p}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_p \quad (1.2)$$

The rate of variation of the volume with temperature can be expressed through the *coefficient of thermal expansion*. This quantity is given by:

$$\alpha = \frac{1}{V} \left(\frac{\partial V}{\partial T}\right)_p \quad (1.3)$$

Substituting into Eq. (1.1) we obtain:

$$\left(\frac{\partial T}{\partial p}\right)_S = \alpha V \left(\frac{\partial T}{\partial S}\right)_p \quad (1.4)$$

The last derivative in Eq. (1.4) can be calculated through the *specific heat at constant pressure* and the usual Maxwell relations:

$$c_p = \frac{T}{\rho V} \left(\frac{\partial S}{\partial T}\right)_p = \frac{T}{\rho V (\partial T / \partial S)_p} \quad (1.5)$$

where  $\rho$  is the density. Therefore, the rate of change of the temperature with pressure will be given by:

$$\left(\frac{\partial T}{\partial p}\right)_S = \frac{\alpha T}{\rho c_p} \quad (1.6)$$

In order to determine the variation of  $T$  with depth, we can use the simple relationship expressing the variation of the hydrostatic pressure with depth:

$$\frac{\partial P}{\partial z} = \rho g \quad (1.7)$$

where  $g$  is the acceleration of gravity and  $z$  is the depth. Therefore, the adiabatic rate of change of the temperature with depth will be given by:

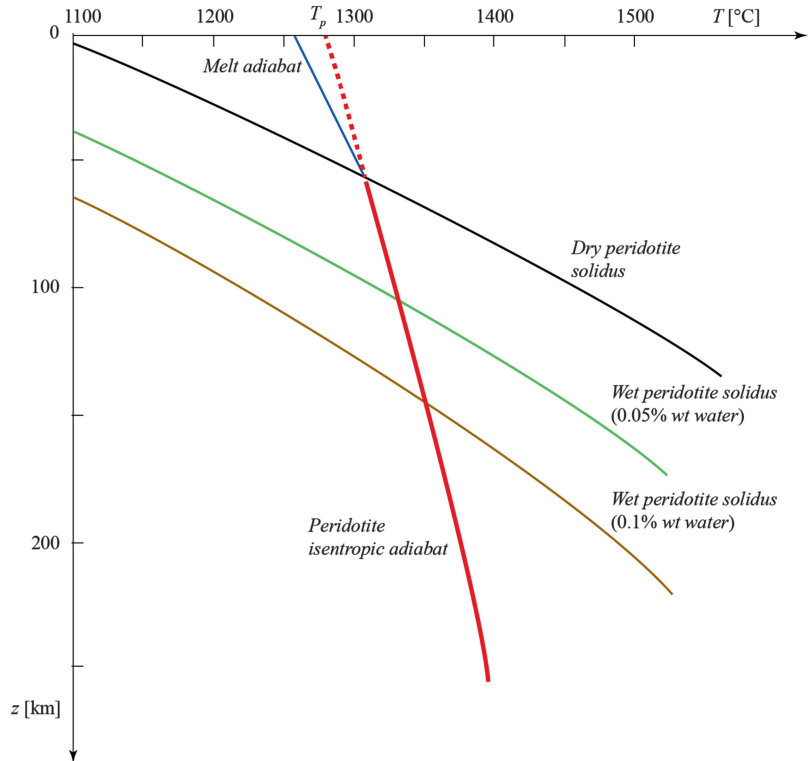
$$\left(\frac{\partial T}{\partial z}\right)_S = \frac{\alpha T g}{c_p} \quad (1.8)$$

The coefficient of thermal expansion varies between  $\alpha = 5 \times 10^{-5} \text{ K}^{-1}$  at  $p = 0.1 \text{ MPa}$ ,  $T = 1,700^\circ\text{K}$  (an adiabatically decompressed mantle) and  $\alpha = 1 \times 10^{-5} \text{ K}^{-1}$  at the core-mantle boundary (Chopelas and Boehler 1992). For  $T = 1,700^\circ\text{K}$ ,  $p = 10 \text{ GPa}$  ( $\sim 300 \text{ km}$  depth) this quantity assumes the value:  $\alpha \cong 3.5 \times 10^{-5} \text{ K}^{-1}$ . The specific heat at constant pressure for the mantle is  $c_p \cong 1.2 \text{ kJ K}^{-1} \text{ kg}^{-1}$  (Stacey 2010). Therefore, using these values in Eq. (1.8), the adiabatic gradient of temperature in the uppermost asthenosphere is approximately:

$$\left(\frac{\partial T}{\partial z}\right)_S \cong 0.5^\circ\text{K km}^{-1} \quad (1.9)$$

Equations (1.6) and (1.8) describe the dependence of temperature from pressure,  $T = T(p)$ , and depth,  $T = T(z)$ , beneath ridges. These *isentropic adiabat geotherms* have great importance in geodynamics. A key observation is that an isentropic adiabat has a less steep slope  $dT/dp$  (or  $dT/dz$ ) than the *peridotite solidus*, which is the set of  $(p, T)$  pairs where these rocks start melting.

**Fig. 1.4** Isentropic adiabat upwelling geotherms for an asthenosphere (*red line*) with potential temperature  $T_p = 1,280^\circ\text{C}$  (see text) and for MORB (in *blue*). Dry and wet peridotite solidi are also shown



Thus, the two curves intersect at a point and partial melting becomes a necessary consequence of the passive upwelling of asthenosphere, as shown in Fig. 1.4.

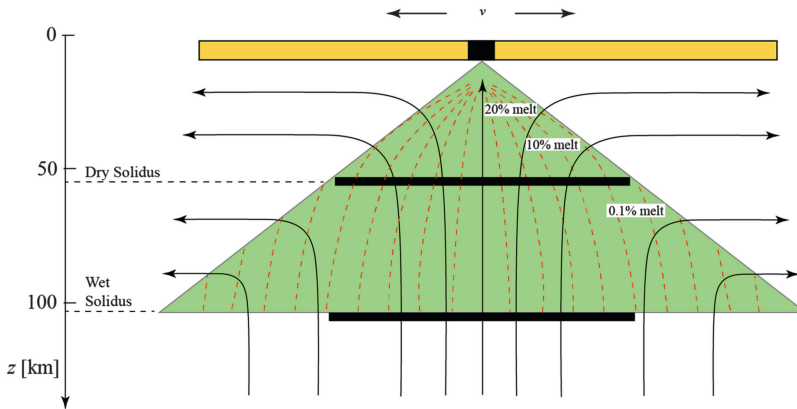
The concept of *potential temperature*,  $T_p$ , is related to the necessity to perform comparisons of the heat content of a volume element at different depths (hence at different  $p$  and  $T$  conditions). If the material is incompressible, then a difference of heat content,  $dQ$ , between two volume elements at different  $(p, T)$  will be proportional to the difference of temperature,  $dT$ . However, if it is not, then  $dQ$  will be proportional to  $dT$  only if the two volume elements will be brought to the same pressure through an isentropic path. The reason is that if the material is compressed or stretched by a variation of pressure, so that its volume is reduced or expanded, it will be also heated or cooled as a result of the work done by the external or internal pressure fields. Therefore, it is a common practice in geodynamics to introduce the quantity  $T_p$ , which is the temperature that a volume element would have if it were brought

isentropically to a reference pressure  $p_0$  (usually 1 atm). After this reduction, differences of temperature measure variations in heat content, thereby  $T_p$  only changes when the entropy of the material changes. To find the potential temperature associated with the actual temperature  $T$  at depth  $z$  we simply integrate Eq. (1.8) between  $z$  and the Earth's surface:

$$\int_T^{T_p} \frac{dT'}{T'} = \frac{\alpha g}{c_p} \int_z^0 dz' \quad (1.10)$$

$$T_p = T(z) e^{-\alpha g z / c_p} \quad (1.11)$$

McKenzie and Bickle (1988) estimated a potential temperature of the asthenosphere of  $1,280^\circ\text{C}$  to generate a 7 km thick oceanic crust. More recent estimates suggest that  $T_p$  beneath ridges varies linearly between  $1,280$  and  $1,400^\circ\text{C}$  as a function of the MgO content in primary magma (10–13 % wt MgO, e.g. McKenzie et al. 2005; Herzberg et al. 2007).



**Fig. 1.5** Asthenosphere flow beneath a mid-ocean ridge (black lines) and melting regime (green triangle). Red dashed lines represent melt migration flow lines. The oceanic crust is shown in yellow. The small black rectangle at the spreading center represents the volume of new

crust generated per unit time. The large black rectangles at dry and wet solidi depths show the volumes of mantle entering respectively the higher-degree and lower degree melting regime zones during the same time interval, which are necessary to produce that amount of new crust

Figure 1.4 shows the mantle adiabatic geotherm for a potential temperature  $T_p = 1,280^\circ\text{C}$ . In this example, the melt temperature ranges from  $\sim 1,310^\circ\text{C}$  at the dry peridotite solidus (55 km depth) to  $\sim 1,260^\circ\text{C}$  at the surface. We note that the melt adiabatic geotherm has different slope and potential temperature. The reason is that  $(\partial T/\partial z)_S \cong 1.0^\circ\text{K/km}$  for a basaltic magma, approximately twice the slope of the mantle geotherm.

Let us consider now the solidus curves of peridotite rocks (Fig. 1.4). Hirschmann (2000) showed that the dry solidus of peridotite between 0 and 10 GPa can be constrained experimentally to follow a simple parabolic law:

$$T_s(p) = ap^2 + bp + c \quad (1.12)$$

where  $a \cong -5.14^\circ\text{C GPa}^{-2}$ ,  $b \cong 132.90^\circ\text{C GPa}^{-1}$ , and  $c \cong 1,120.66^\circ\text{C}$ . This solidus is strongly affected by the presence of volatiles (water,  $\text{CO}_2$ , etc.), which always reduce the melting point temperature of these rocks. For example, Fig. 1.4 shows the effect of the addition of water to a dry peridotite system (Katz et al. 2003). Therefore, partial melting may start at considerable depth beneath a spreading center. Dasgupta and Hirschmann (2006) have recently shown that deep melting must occur at depths up

to 330 km, producing 0.03–0.3 % carbonatite liquids, as a consequence of the addition of small amounts of  $\text{CO}_2$  to peridotite. However, most of melts form at depths shallower than  $\sim 55$  km, where the mantle adiabat crosses the dry solidus of peridotite (Fig. 1.4). The zone of upwelling and melting can be as wide as 100 km at 50–70 km depth. At shallower depths melting proceeds at enhanced rate, with extensive extraction of water from the solid phase. Figure 1.5 illustrates the flows within and around the so-called *melting regime*, which is the region below a mid-ocean ridge where melts are generated and carried to the surface.

The melting regime can be defined as the region above the peridotite wet solidus where the asthenosphere flow has a vertical component of velocity (Plank and Langmuir 1992). Therefore, its lateral boundaries are marked by the zone where the mantle flows almost horizontally, so that adiabatic decompression ceases or heat conduction prevails. In stationary conditions, it can be shown that the maximum velocity of upwelling is  $u_{\max} = v/\pi$ , where  $v$  is the spreading rate (Phipps Morgan 1987). Therefore, if  $w$  is the width of the higher-degree melting regime zone and  $\bar{F}$  is the average degree of melting, then the generation of a sliver of oceanic crust having thickness  $H$  requires a melting regime




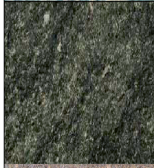

having a minimum width  $w = \pi H / \bar{F}$ . For example, for  $H = 7$  km and  $\bar{F} = 20$  %, a 110 km wide melting regime is necessary, granted that the entire amount of melt is transported to the ridge axis. In general, the degree of melt focusing to the spreading center is elevated in the higher-degree melting regime zone. The pattern of melt migration depends from the anisotropy of permeability of the peridotite rocks, which determines the preferred directions of channeling of the melt between mineral grains. Such anisotropy results from the process of deformation of these rocks within the melting regime, where the upwelling asthenosphere flow is deviated to the horizontal. Figure 1.5 shows the pattern of melt focusing in the model of Phipps Morgan (1987). It is clear that a small amount of melt is always retained, even in the higher-degree melting zone, because melts that cross the lateral boundaries of the melting regime will be incorporated within the horizontal mantle flow and will never reach the ridge. Therefore, the asthenosphere outside the melting regime always contains a small amount of melt, which contributes to decrease the viscosity of this layer.

The average degree of melting beneath normal ocean ridges,  $\bar{F}$ , increases with increasing spreading rate. The compositions of the oceanic crust and the residual asthenosphere depend from the process of mixing and equilibration of the melts that are produced during the mantle upwelling. An important feature of the quantity  $\bar{F}$  is that it reflects the degree of enrichment of the oceanic crust in highly *incompatible elements*. These are elements, such as K, Rb, Sr, U, and rare-earth elements (REE), that do not easily fit into the crystal lattice structures of minerals such as olivine, pyroxene, spinel, and garnet, and thereby are the first to be partitioned into the melt. Conversely, the residual asthenospheric column leaving the melting regime will be depleted in these elements. The final product of sea floor spreading is represented by a 6–10 km thick oceanic crust layer made by extrusive MORBs and intrusive gabbros. Mid-ocean ridge basalts are fine-grained rocks, glassy to porphyritic, whose sequence of crystallization starts at depths shallower than  $\sim 18$  km (Grove

et al. 1992). This sequence is: olivine ( $\pm$ Mg–Cr Spinel), olivine + plagioclase ( $\pm$ Mg–Cr Spinel), olivine + plagioclase + clinopyroxene. The average composition of MORBs includes 50.5 %  $\text{SiO}_2$ , 15.3 %  $\text{Al}_2\text{O}_3$ , 10.4 %  $\text{FeO}_T$ , 11.3 %  $\text{CaO}$ , 7.6 %  $\text{MgO}$ , 2.7 %  $\text{Na}_2\text{O}$ , and 1.6 %  $\text{TiO}_2$  (Hofmann 1988).

The oceanic crust has a characteristic layered structure, which can be investigated both seismically and by on-land observation of *ophiolite sequences*. These rock assemblages represent remnants of oceanic crust that has been obducted onto a continental margin by plate tectonic processes. For example, if an oceanic plate is subducting beneath another oceanic plate, and if it also carries continental crust in the rear, then a collision between the continent and the intra-oceanic subduction zone (comprising the associated island arc) is unavoidable, as soon as the oceanic crust has been entirely destroyed. In this instance, subduction will be rapidly stopped, because the low-density continental crust cannot be subducted. However, a small sliver of the overriding oceanic plate will be eventually transported (obducted) onto the continental margin, where it will form an ophiolite sequence. Figure 1.6 shows a typical cross-section of oceanic crust, based on refraction and reflection seismology experiments (e.g., White et al. 1992; Christeson et al. 2012), and direct field observation of ophiolite sequences.

The sequence starts (in top-down direction) with a thin layer of deep-sea sediments, usually less than 0.5 km, which can be formed by carbonate oozes, radiolarites, or argillites depending from sea floor depth, distance from the ridge, and latitude. Then, we find the effusive high-porosity MORB layer 2,  $\sim 0.75$  km thick, formed by ovoidal masses resembling pillows (pillow lavas). This stratum can be further divided into two layers, 2A and 2B, with a boundary that is observed at  $\sim 400$ – $600$  m below the sea floor away from the ridge axis (Christeson et al. 2012). The 2A layer has  $P$ -wave velocities of  $2$ – $3$  km  $\text{s}^{-1}$  in the upper 250–300 m, followed by a high-gradient region, with velocity increasing linearly up to  $\sim 4.7$  km  $\text{s}^{-1}$  at 500 m depth. The transition from such high-gradient region to velocities of

	Layer	Sym	$H$	$v_p$
	Deep-sea sediments	1	< 0.5	1.7 – 2.0
	Pillow lava basalts	2A	0.5	3.0 – 3.2
		2B	0.25	4.7 – 4.9
	Sheeted dike complex	2C	1.5	6.6 – 6.7
	Isotropic gabbro	3A	2 – 5	7.2
	Layered gabbro			

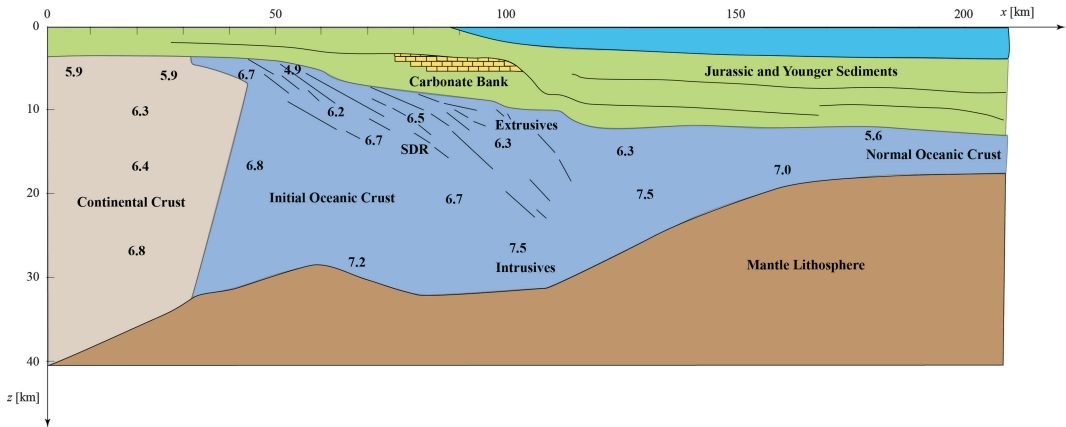
**Fig. 1.6** Layering of the oceanic crust.  $H$  is the thickness in km,  $v_p$  is the average  $P$ -waves velocity in  $\text{km s}^{-1}$ .  $Sym$  is the symbol of the corresponding layer

4.7–4.9  $\text{km s}^{-1}$  marks the top of the seismic layer 2B, which can be considered as a transition zone characterized by the first appearance of dikes. The underlying low-porosity sheeted dike complex, layer 2C, is  $\sim 1.5$  km thick and has  $P$ -wave velocities between 6.6 and 6.7  $\text{km s}^{-1}$ . This layer provides the most direct evidence of sea floor spreading, because it is formed by dikes that have intruded older dikes in so far as the divergent motion of two tectonic plates has created new space to be filled. Therefore, these dikes are generally arranged in a regular temporal sequence, with the youngest ones located close to the ridge and the oldest placed along the continental margins. The presence of a well-developed 2C layer is also evidence of approximate balance between spreading rate and magma supply (Robinson et al. 2008). The sequence of rocks forming the oceanic crust is completed by the pair of intrusive gabbroid layers 3A and 3B, having thickness 2–5 km and  $P$ -wave velocity  $v_p \sim 7.2$   $\text{km s}^{-1}$ . The isotropic gabbros that can be found just beneath the sheeted dike complex grade into gabbros having a weakly developed near-vertical layering. Such layering becomes more developed and acquires shallower

dip toward the base of the sequence, where dips are sub-horizontal (Nicolas 1989).

Except for plate boundaries, only COBs interrupt the considerable petrologic and mechanical regularity of the oceanic crust. These zones, which mark the transition to the continental domain, can be narrow or broad, and characterized by the presence of ultra-thinned blocks of “transitional” continental crust embedded into an oceanic “matrix”, exhumed mantle not capped by oceanic crust, or thickened “transitional” oceanic crust. A modern and more rigorous term (with respect to the acronym COB) to indicate the interface region between oceanic and continental crust is *ocean-continent transition zone*, or simply OCT. We shall conclude this section dedicated to the oceanic crust with a short discussion about the principal characteristics of these interesting regions.

The formation of an oceanic basin is always preceded by a phase of continental rifting that involves extensional faulting and thinning of the continental crust. A modern example of this process can be observed along the West African Rift zone, which extends from the region of Afar in Ethiopia to South Africa. The onset of sea floor spreading is not necessarily synchronous along the rift axis, but more generally occurs along discrete axial cells that form and grow independently for some million years until they join into linear spreading segments, as observed in the Red Sea (Bonatti 1985). At the end of this stage, the geologic setting of the region at the interface between the thinned continental margin and the truly oceanic domain depends from many factors, the most important being: (a) the velocity of rifting, (b) the presence of thermal anomalies, (c) the fertility of the asthenosphere beneath the rift area, and (d) the presence of small-scale convective currents in the asthenosphere (Ligi et al. 2011, 2012). Two end-members can be used to illustrate the possible range of situations. In the case of *volcanic passive margins*, the onset of sea floor spreading is associated with intense volcanism, eventually related to the presence of a *mantle plume*. This can be viewed as an anomalously hot or fertile region of the asthenosphere or, more in general, as an upper-mantle area characterized by



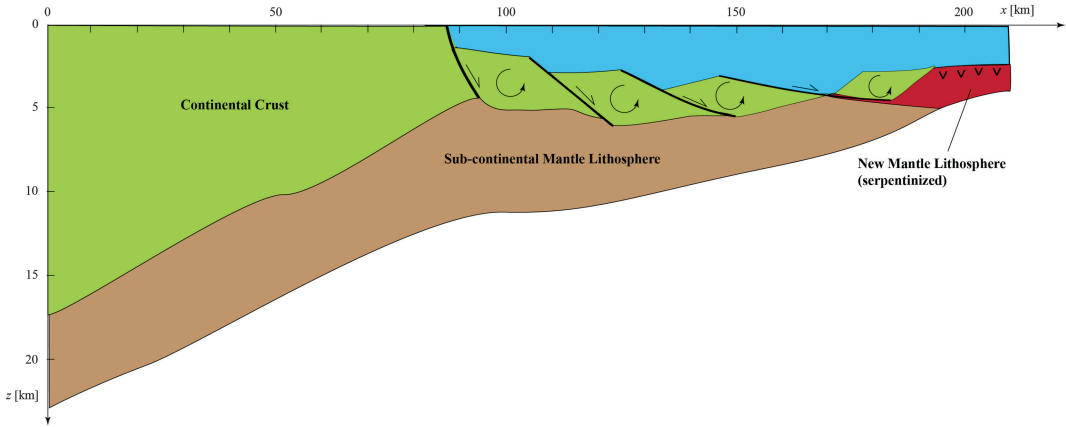
**Fig. 1.7** Seismic reflection-refraction cross-section along the eastern US Atlantic margin (Modified from Talwani and Abreu 2000. Numbers represent  $P$ -waves velocity in  $\text{km s}^{-1}$ . SDR = Seaward dipping reflectors)

anomalous composition or volatile content (e.g., Anderson and Natland 2005). In this instance, the acceleration that accompanies the transition from the rifting stage to drifting may induce a “spurt” of *active* upwelling, with anomalous production of melt and formation of compressive structures along the continental margins. For example, the break-up of the supercontinent Pangaea at  $\sim 200$  Ma was accompanied by an event of extensive volcanism, which gave rise to the *Central Atlantic Magmatic Province* (CAMP). This event was also responsible for the formation of a thick initial oceanic crust along the eastern margin of North America (Talwani and Abreu 2000). At the same time, the extensional structures associated with the rift basins were inverted as a consequence of horizontal compression (Schlische et al. 2002).

A similar mechanism is probably incipient in the present-day northern Red Sea (Ligi et al. 2011), thereby the magmatism of the Afar region could be only the “epicenter” of a future larger magmatic pulse. Figure 1.7 shows a combined seismic reflection–refraction cross-section along the Atlantic margin of the United States (Talwani and Abreu 2000), which illustrates the main features of volcanic margins. We note the presence of a more than 20 km thick atypical oceanic crust, whose upper part has higher velocities with respect to both the continental crust and the extrusive layers (2A, 2B) of the normal oceanic crust. The considerable thickness of such

initial oceanic crust is a direct consequence of the anomalously high potential temperature of the asthenosphere beneath these rifts ( $100\text{--}200^\circ\text{C}$  above normal  $T_p$ , White and McKenzie 1989). A distinctive feature of this kind of OCT is the presence of *seaward dipping reflectors* (SDR), which are surfaces of discontinuity within the extrusive region, having a characteristic dip towards the ocean (Fig. 1.7). Volcanic OCTs belong to the world’s *Large Igneous Provinces* (LIPs, Coffin and Eldholm 1994). They are also the magmatic expression of catastrophic events that have determined the continental break-up and huge volcanism. Well-known examples are the East Coast of the US from Georgia to Connecticut, the western Indian margin, the conjugate margins of South America and the South African craton, Greenland and Eurasia, Eritrea and Yemen.

At the opposite of the quite common volcanic rifts, there exist only a few examples of *non-volcanic passive margins*, which are characterized by low magma supply and ultraslow velocity of the separating plates. An abrupt decrease in the production of melt can be observed when the full spreading rates are below  $\sim 20 \text{ mm year}^{-1}$  (White et al. 2001). At these very slow spreading rates, the upwelling of asthenosphere is not adiabatic, so that conductive cooling prevails and melting is inhibited. This favors the growth of new lithospheric mantle beneath the rift, according to a thermodynamic process that will be clarified in Chap. 12. Conversely, the absence of MORB



**Fig. 1.8** Typical cross-section of a non-volcanic OCT. The continental margin is formed by a sequence of blocks that have been tilted and rotated by listric faults. The growth of new mantle lithosphere not capped by oceanic crust occurs towards the spreading center (*brown area*).

The olivine in mantle peridotite is subject to hydration at the contact with the oceanic seawater, which induces a metamorphic transformation that converts this mineral to serpentine (a hydrous Mg-silicate), magnetite, and other minerals

magma prevents the formation of oceanic crust at the ridge axis, thereby the new oceanic seaway will be floored directly by mantle peridotites. Figure 1.8 illustrates the main features of this kind of OCTs. Well-known examples are the western margin of Iberia and its conjugate (the Newfoundland margin), the northernmost segments of the central Atlantic passive margins, the conjugate margins of Labrador and western Greenland, and the paleo-margins of the western Tethys. These OCTs, which can be hundreds of kilometers wide, generally have a seaward termination represented by a *peridotite ridge*, a basement high partly buried by post-rift sediment (Boillot and Froitzheim 2001). Although both the peridotite ridge and the serpentinized mantle exhumed along the OCT may have a strong magnetic signal, the truly oceanic domain, characterized by MORB accretion and well-developed seafloor spreading magnetic anomalies (Chap. 5), starts farther seaward.

## 1.4 Lithospheric Mantle

The *lithospheric mantle* composes the strong outermost part of the Earth's mantle. However, it is often described as the lower portion of a more heterogeneous external layer, the

*lithosphere*, which also includes the continental and oceanic crust. In most treatises on plate tectonics, this composite lithospheric layer is considered as the true protagonist of the game, because the usual definition of *tectonic plates* refers to a subdivision of the entire lithosphere (including the crustal and mantle components) into a set of blocks that move independently each other and behave, to a first approximation, like elastic laminae. Therefore, it is generally supposed that the lithosphere has coherent mechanical behaviour despite the lateral and vertical variability of chemical composition. In reality, both direct observations of seismicity and laboratory experiments indicate that only the upper portion of a lithospheric plate, the *elastic-ductile lithosphere*, can resist static shear deformation through an elastic or plastic response, whereas the lower part will flow just like the underlying asthenosphere, although with much higher viscosity. For this reason, such lower lithospheric layer cannot be considered as a permanent part of a tectonic plate. Experimentally, it is found that the boundary between the elastic-ductile lithosphere and the lower fluid layer approximately coincides with the 650 °C isotherm and is marked by a sharp cutoff of seismicity (e.g., Bodine et al. 1981; Anderson 1995).

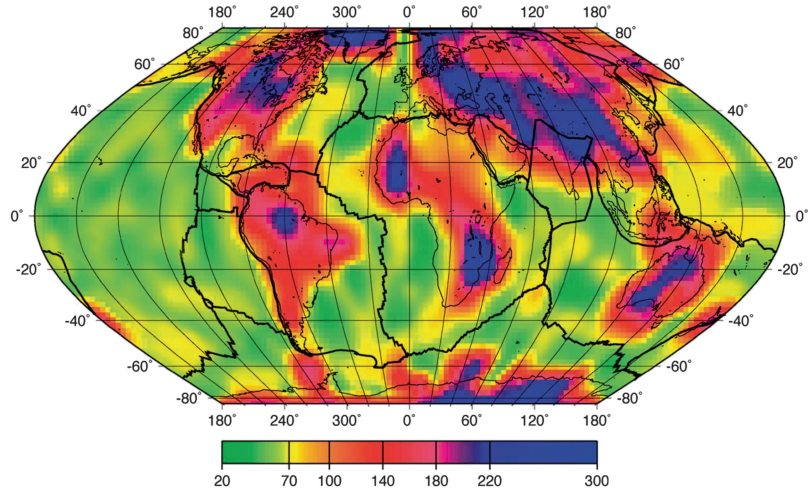
The limit between crust and lithospheric mantle is shallower than the thermo-mechanical boundary discussed above, and is represented by a sharp surface of discontinuity in the seismic velocities. Across this discontinuity, the  $P$ -wave velocity jumps from 7.0–7.2 to 8.1 km s<sup>-1</sup>, while the  $S$ -wave velocity changes from an average 3.9 to 4.5 km s<sup>-1</sup>. This surface is called *Mohorovičić discontinuity*, or simply *Moho*. It marks the transition from the basaltic rocks of the oceanic crust, or from the granulite facies of the lower continental crust, to the upper mantle *peridotites*. An example of mineral composition of peridotite rocks in the mantle lithosphere is: 71.6 % Ol, 23.7 % Opx, 3.6 % Cpx, and 1.1 % Sp (Eggins et al. 1998). In terms of oxide components, these ultramafic rocks may include: 44.1 % SiO<sub>2</sub>, 44.5 % MgO, 7.9 % FeO<sub>T</sub>, 1.3 % Al<sub>2</sub>O<sub>3</sub>, 0.9 % CaO, and 0.5 % Cr<sub>2</sub>O<sub>3</sub>. Although the chemical compositions of the mantle lithosphere and the asthenosphere are similar, there is a substantial difference in the degree of fertility, hydration, and presence of melt. When fertile and wet asthenosphere melts at a spreading ridge by adiabatic decompression (Fig. 1.5), the residual column of asthenospheric material leaving the melting regime is depleted in incompatible elements and dehydrated. Furthermore, it may contain retained melts in the lower part, at the wet solidus boundary. While this column is dragged horizontally by the overlying oceanic crust, it is also subject to cooling by conductive loss of heat through the Earth's surface. The details of this process will be given in Chap. 12. For the moment, it is sufficient to observe that at any time we can divide the column into an upper part, where the *potential temperature* is fallen below the asthenosphere  $T_p$  ( $\sim 1,280$  °C, Fig. 1.4), and a lower hotter zone, which has not yet lost a significant amount of heat. The potential temperature at any depth  $z$  within this conductive *thermal boundary layer* (TBL) can be calculated using Eq. (1.11), starting from the effective temperature  $T(z)$ . The base of the TBL coincides with the lower boundary of the *thermal lithosphere*, and is generally used to define the *lithosphere-asthenosphere* boundary (LAB) also beneath the continents, although it

must be emphasized that the material between the 650 °C isotherm and the LAB behaves as a fluid and lacks any elastic or plastic strength.

Although the definition of TBL does not explicitly mention the chemical nature of the material within this layer, there is a substantial difference between the sub-continental and the oceanic mantle lithospheres. A first important distinction is that the first one most likely formed during the Archean (3.0–3.5 Ga) together with the overlying crust (Carlson et al. 2005), whereas new oceanic mantle lithosphere is continuously created by the conductive cooling of residual asthenospheric columns. A second distinction concerns the TBL height. The thickness of an oceanic TBL within a residual column increases progressively at the expenses of underlying asthenosphere, as it moves away from the ridge and cools. In Chap. 12, we shall show that a simple relation exists between age of a residual column and thickness of the TBL, which attains  $\sim 125$  km after 80 Myrs. Conversely, there is no simple relationship between age and thickness of the TBL in the case of the sub-continental mantle lithosphere. Here the depth to the LAB shows considerable variability and might reach  $\sim 300$  km beneath the Precambrian shields (Fig. 1.9).

Figure 1.9 shows a global estimate of the depth to the LAB, based on seismic anisotropy data (Plomerová et al. 2002). In this kind of studies, the starting point is the observation that the crystallographic axes of olivine and orthopyroxene aggregates acquire non-random orientations (lattice-preferred orientations, or LPO) in response to shear deformation. For example, the orientations of the crystallographic axes of minerals in the asthenosphere will be determined by the direction of the present day flows within this layer (Tanimoto and Anderson 1984). In particular, the  $a$ -axes will cluster about the flow direction, the  $a$ -axes and  $c$ -axes will concentrate in the flow plane, while the  $b$ -axes tend to be aligned with the normal to the plane of flow. This behavior clearly determines *anisotropy* in the propagation of seismic waves (Chap. 9), thereby the wavefront velocity will be different along different directions of propagation. Such anisotropy can be detected through a variety of techniques. In any case, the

**Fig. 1.9** Depth to the LAB (After Plomerová et al. (2002). Depths are in km)



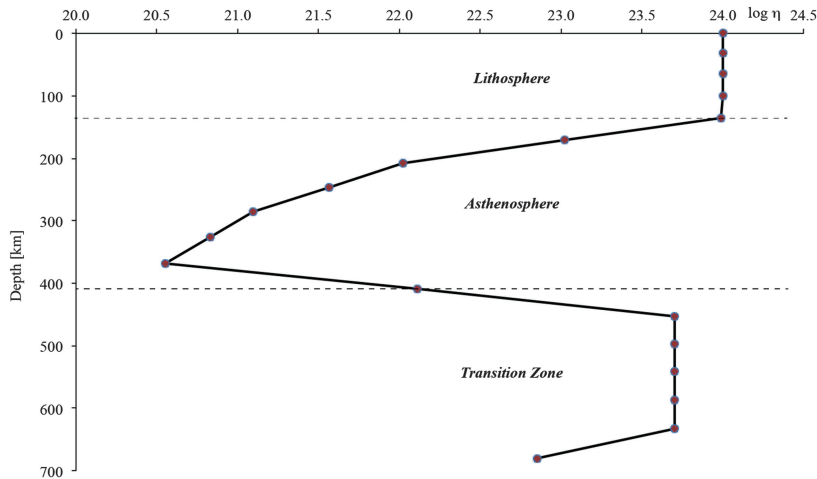
fact that the rate of shear deformation increases abruptly passing from the mantle lithosphere to the underlying asthenosphere implies a vertical variation of seismic anisotropy that can be detected and used to constrain the LAB topography.

An interesting feature of the lithosphere beneath the continental cratons is the presence of thick “keels” of highly refractory peridotite (Fig. 1.9), which are characterized by high seismic velocities. Furthermore, while the base of the lithosphere is marked by a sharp reduction of seismic velocities beneath the oceanic basins, possibly associated with retained melts, such a reduction seems to be absent beneath the old Precambrian shields. In general, the seismic velocities are higher in the sub-continental mantle lithosphere and correlate with the age of tectono-magmatic activity at regional scale. Another important source of data about the sub-continental mantle lithosphere, which has undoubtedly helped to clarify the differences with respect to the oceanic counterpart, is represented by *xenoliths*. These are fragments of mantle lithosphere carried to the surface by explosive volcanic rocks such as kimberlites, whose systematic study has produced a large knowledge base about the physical properties, the chemistry, and the geochronologic history of the sub-continental mantle lithosphere. For example, xenoliths have revealed that this lithosphere is anomalously enriched in some highly incompatible trace elements, such as potassium,

uranium, and thorium. Such enrichment explains the relatively high production of radiogenic heat of the continental regions with respect to the oceans. Perhaps the most important conclusion of these studies has been that the cratonic mantle lithosphere is less dense than the oceanic TBL, as a consequence of high degree partial melting and melt removal during the Archean. Such primary melt depletion gave to the sub-continental mantle peridotites a compositional buoyancy that must be considered as a key factor preserving both the continental crust and the underlying mantle from sinking into the asthenosphere (e.g., Carlson et al. 2004). Conversely, the dry, chemically depleted, oceanic lithosphere becomes gravitationally unstable in so far as its thickness and average density increase with the cooling. Therefore, the fate of the old oceanic lithosphere is to bend downwards and sink into the asthenosphere, possibly after an episode of horizontal compression, forming a *slab*. In this event, phase transitions, such as dehydration or MORB metamorphism to eclogite facies, will determine an increase of density, which facilitates passive sinking into the mantle.

## 1.5 Asthenosphere

The *asthenosphere* is a mechanically weak solid state fluid layer just beneath the lithosphere (Fig. 1.1). Its base is defined by the 410-km



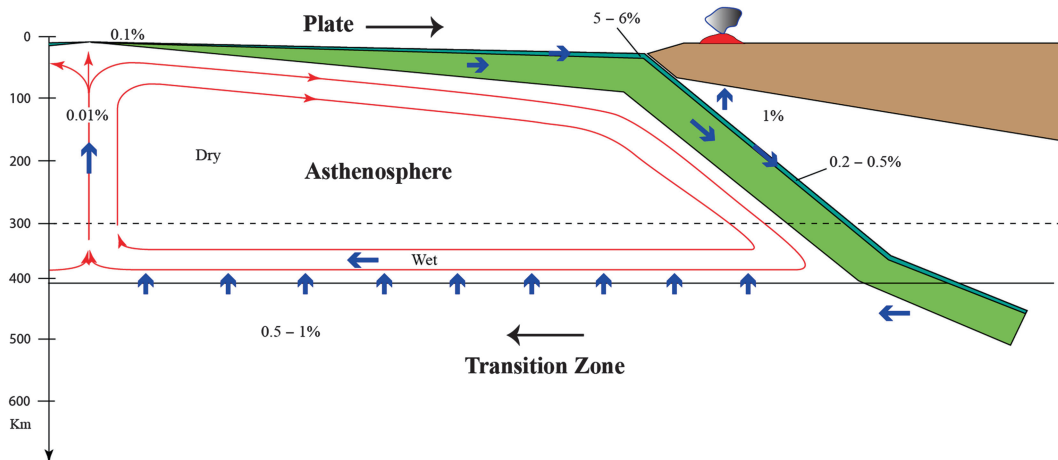
**Fig. 1.10** Logarithm of viscosity,  $\eta$ , across the upper mantle in the model of Walzer et al. (2004)

*discontinuity*, a surface of sharp increase of seismic velocities and density associated with the chemical transition of the olivine to the  $\beta$ -spinel structure. The upper part of this layer, down to  $\sim 220$  km depth, forms a *low-velocity zone* (LVZ), where the seismic velocities are slightly smaller with respect to those of the overlying lithosphere and the underlying lower asthenosphere (Gutenberg 1959). This phenomenon is probably related to the presence of melts, either microscopically as thin intergranular films or macroscopically as narrow dikes or sills (Anderson 1989). The LVZ is very thin beneath the cratons, whereas it approaches the 410-km discontinuity under the East Pacific Rise.

There are two key features that characterize the asthenospheric layer. The first one is represented by its *local* inhomogeneity, which is clearly determined by the continuous injection of density, chemical, and thermal anomalies through the subduction process. Such inhomogeneity induces both vertical and horizontal variations in the upper mantle geotherms, so that the range of temperatures could be as high as  $\sim 400$  °C (Anderson 2000). The second key feature is the capability of this layer to flow and deform promptly in response to external forces. In Chap. 13 we shall show that these flows fall in three basic categories: (a) Laminar flows, associated with the drag exerted by an overlying

tectonic plate that is moving towards a subduction zone; (b) Pressure-driven flows, which are generated by lateral variations of pressure; and (c) Temperature-driven Rayleigh-Bénard thermal convection, which arises spontaneously as a consequence of the vertical variability in the temperature distribution and the basal heating of this layer. The local presence of melts further increases the complexity of the asthenosphere and consequently the difficulties to find an adequate geodynamic description of this layer. Figure 1.10 illustrates a possible trend of the viscosity in the upper mantle (Walzer et al. 2004). It is evident in such a model that the asthenosphere is characterized by a viscosity profile with values that decrease with depth by more than three orders of magnitude, from  $\eta \sim 10^{24}$  Pa s at the LAB to  $\eta \cong 3.6 \times 10^{20}$  Pa s. Only close to the base of this layer, at  $\sim 367$  km depth, the viscosity starts increasing, reaching the value  $\eta \sim 10^{22}$  Pa s at its lower boundary. However, most recent estimations of upper mantle viscosity point to even lower values. In Chap. 13, we shall see that the average upper mantle viscosity is between  $0.5$  and  $1.0 \times 10^{21}$  Pa s, while the asthenosphere viscosity could be between  $0.5$  and  $1 \times 10^{20}$  Pa s.

Some lines of evidence suggest that the lower part of the asthenosphere, approximately below  $\sim 300$  km depth, may be the source region of



**Fig. 1.11** Large-scale asthenosphere circulation in the oceanic domain and cycle of water in the upper mantle. Blue lines represent transport of H<sub>2</sub>O, while numbers indicate the estimated water content (% wt). Red lines show

upwellings, downwellings, flows, and counterflows in the asthenosphere. Light and dark green regions represent, respectively, mantle lithosphere and crust that independently carry variable amounts of water

MORB. When the oceanic lithosphere moves away from a ridge axis, it generates laminar flows within the upper asthenosphere by exerting viscous drag forces. These shallow flows, which diverge from mid-ocean ridges, must be compensated by counterflows within the deep asthenosphere that are directed towards the mantle upwelling zones (Chase 1979). Although Anderson and Bass (1986) argued that the sources of MORB reside within the transition zone (between the 410 and the 670 km discontinuities), there are three independent classes of observation suggesting a shallower source. The first one is that the viscosity in the transition zone is too high to allow an effective counterflow. In other words, even assuming a participation of this layer to the return flow, the viscosity contrast with the asthenosphere is so high that most of the counterflow will be concentrated anyway in the lower asthenosphere, as we shall prove in Chap. 13. The second class of observations is represented by seismic tomography models of the asthenosphere beneath ridges, which do not show evidence of vertical motion below ~300 km (e.g., Gu et al. 2005). Finally, the distribution of water within the mantle and the balance of water flow between crust and mantle lead to conclude that the probable source of MORB is just above the

410-km discontinuity in the lower asthenosphere. This point will be discussed in detail below.

In the Earth's mantle, H<sub>2</sub>O can either be stored in solid minerals or be present as hydrous fluids or melts. Both the instantaneous distribution and the large-scale transport of this element represent key aspects of plate tectonics. The distribution is clearly controlled by the flows, but it also depends from the storage capacity of the mantle minerals. The storage capacity of the asthenosphere is between 0.1 and 1 wt.% (Hirschmann 2006), that of the lower mantle is much lower (<20 ppm H<sub>2</sub>O). Conversely, the transition zone may have a relevant storage capacity of 0.5 wt.% at 1,600 °C and transition zone pressures (Ohtani et al. 2001). Regarding the water transport, diffusion cannot be a viable mechanism, because the diffusion distance is only ~10 km Gy<sup>-1</sup>. Therefore, the relevant mechanisms of transport are the subduction of oceanic lithosphere and partial melting. In general, the former brings water downwards to the asthenosphere and the transition zone, the latter drives water upwards to the crust, as shown in Fig. 1.11.

Water can be considered as an incompatible element during partial melting, having strong similarity with Ce. Therefore, the measured ratio H<sub>2</sub>O/Ce in undegassed basalts can be

combined with known concentrations of Ce in abyssal peridotites, which are representative of MORB source regions (Workman and Hart 2005), to provide an estimate of the water content. Such studies indicate that the MORB source region has water content between 50 and 200 ppm. Some authors have assumed this concentration as representative of the whole asthenosphere, with the exception of the subduction zones, where the concentration could be as high as 1 wt.% as a consequence of slab dehydration (e.g., Hirschmann 2006; Karato 2011). The best way to estimate the effective water content across the asthenosphere is through analysis of electrical conductivity, because this quantity is much sensitive to the presence of hydrogen (Karato 1990). Studies of the electrical conductivity of the upper mantle in the Pacific region (e.g., Shimizu et al. 2010) are compatible with the *dry* asthenosphere conductivity profile computed by Karato (2011) and *not* with the observed H<sub>2</sub>O concentration of the MORB source region. Below the 410-km discontinuity, the theoretical and observed conductivity profiles are in good agreement if the concentration of H<sub>2</sub>O is  $\sim 1$  wt.%. Therefore, the distribution of water in the upper mantle suggests a MORB source region close to the 410-km discontinuity. In this instance, the only regions of the oceanic asthenosphere above  $\sim 300$  km where the concentration of H<sub>2</sub>O rises to  $\sim 0.01$  wt.% will be the upwelling zones beneath mid-ocean ridges. This model predicts a moderately wet MORB source region confined to the lower asthenosphere and an essentially dry upper oceanic asthenosphere (Fig. 1.11). It also implies hydration of the lower asthenosphere counterflow by advection of wet material across the 410-km discontinuity (Hirschmann 2006).

The high solubility of H<sub>2</sub>O in the transition zone minerals suggests that this layer plays a key role in the global water circulation. Despite the shallow dehydration of slabs at subduction zones, which determines a large H<sub>2</sub>O concentration ( $\sim 1$  % wt, Fig. 1.11) in the wedge just above the slab, the rapidly sinking oceanic lithosphere introduces a considerable amount of water at greater depths, directly in the transition

zone. Fresh MORBs contain  $\sim 0.1$  % wt H<sub>2</sub>O (Green et al. 2010), whereas the underlying mantle lithosphere peridotites are essentially dehydrated. However, the water content of both crust and lithospheric mantle increases progressively with the age, because of hydrothermal infiltration of seawater within the oceanic crust or serpentinization of mantle peridotites along fracture zones and transform faults. Therefore, when the lithosphere bends and starts sinking its water content could be as high as 5–6 % wt at a depth of 10–20 km (Schmidt and Poli 1998). A consistent part of this reservoir will be extracted at shallow depth, between 90 and 150 km, determining extensive partial melting of the mantle wedge and arc volcanism (Fig. 1.11). Schmidt and Poli (1998) estimated that the degree of dehydration at this stage is between 18 and 37 %. A fraction of the remaining part of the original reservoir will be extracted at greater depth, where the bulk water content of the slab could decrease to 0.2–0.5 wt.%. Therefore, the subduction process continuously injects H<sub>2</sub>O within the transition zone, where it can be temporarily stored in high-pressure polymorphs of olivine, wadsleyite, and ringwoodite. The steady equilibrium of the Wilson cycle and the geological evidence of stationary oscillations of the sea level clearly exclude a progressive decrease of ocean water at the Earth's surface. Thus, an upward water flow from the transition zone to the overlying asthenosphere, where it will be conveyed towards upwelling flows, is necessary to obtain a global mass balance.

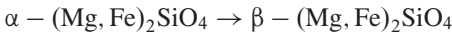
The subduction process itself anyway requires the upward advection of possibly wet material across the 410-km discontinuity. When cold oceanic lithosphere penetrates the transition zone, it is generally subject to upward bending and flattening just above the base of this layer. These stagnant slabs tend to increase the total volume of the transition zone, thereby a corresponding volume will be pushed upwards and will cross the 410-km discontinuity. In this instance, the high H<sub>2</sub>O content will exceed the storage capacity of asthenosphere minerals, determining hydrous melting (Hirschmann 2006). This mechanism may explain the

low-velocity anomalies that have been observed just above the 410-km discontinuity, with a shear-wave velocity drop of  $\sim 5\%$ .

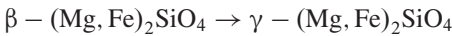
## 1.6 Transition Zone

The *transition zone* forms the lowermost layer of the upper mantle. We have discussed in the previous section its important role in the global water balance and mantle circulation. Here we shall focus on some key features regarding the chemistry and thermodynamics of this region. There are two principal models for the composition of the transition zone. In the classic *pyrolite model* of Ringwood (1975), the two major seismic discontinuities in the Earth's mantle, which bound the transition zone respectively at 410 and 670 km depth, are isochemical phase transformations of a hypothetical garnet peridotite composed mainly by olivine and pyroxene (hence the term “pyrolite”) (Fig. 1.12).

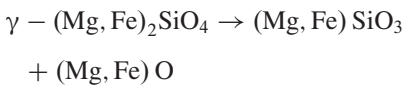
In this instance, the 410-km discontinuity is associated with the pressure-induced transformation of the olivine phase in peridotite rocks to wadsleyite:



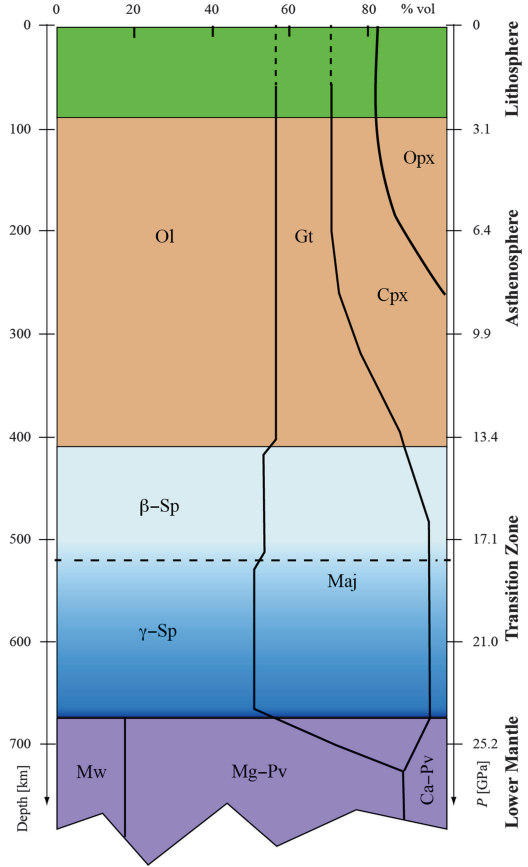
Similarly, a second-order discontinuity at  $\sim 520$  km depth ( $\sim 18$  GPa) determines new collapse of this mineral to ringwoodite:



Finally, the 670-km discontinuity at the base of the mantle transition zone ( $\sim 23$  GPa) is associated with a complete breakdown of the mineral phase to form perovskite and magnesio-wüstite:



The other two phases of the upper mantle mineral assemblage, pyroxenes and garnet, are subject to more gradual transformations. For example, pyroxene gradually dissolves into the

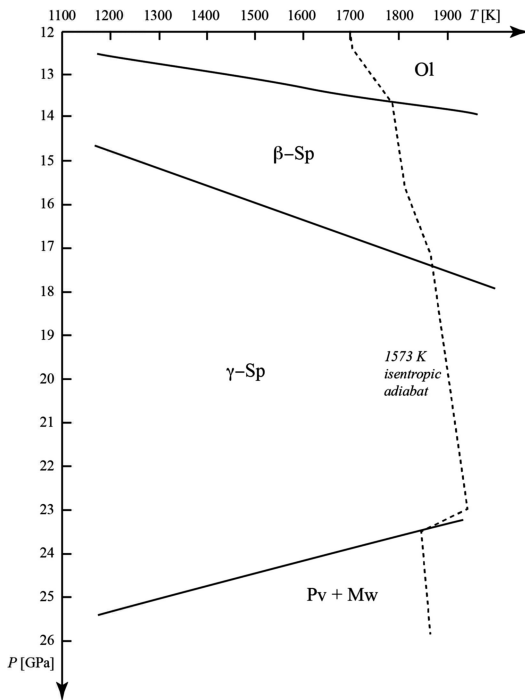


**Fig. 1.12** Chemical composition of the mantle (% vol) in the pyrolite model of Ringwood (1975). Maj is majorite garnet, Mg-Pv = Mg-perovskite, Ca-Pv = Ca-perovskite, the other symbols are explained in Table 1.1

garnet structure starting from  $\sim 300$  km depth ( $\sim 10$  GPa). Majorite garnet, in turn, dissolves into the perovskite structure between 670 and 750 km depth (Fig. 1.12).

In the *piclogite model* of Bass and Anderson (1984) the composition of the mantle transition zone is essentially based on the match of theoretical seismic velocities with observed data and a model of chemical differentiation of the primitive Earth (Anderson and Bass 1986). In this instance, the transition region is assumed garnet-rich rather than olivine-rich and it is viewed as the source region of MORBs.

An important prediction of the pyrolite model arises from the *Clapeyron slopes*,  $dp/dT$ , of the phase transitions at 410 and



**Fig. 1.13** Phase diagram of pyrolite in  $P - T$  space, showing phase transitions at the 410-, 520-, and 670 km discontinuities (Modified from Christensen 1995)

670 km depth. The magnitude of  $dp/dT$  for the transition of olivine to wadsleyite is *positive*, and a recent estimate gives  $+3.1 \text{ MPa K}^{-1}$  at 13.4 GPa and 1,400 K (Akaogi et al. 2007). Conversely, the breakdown of the ringwoodite to perovskite and magnesiowüstite has a *negative* Clapeyron slope of  $-2.6 \pm 0.2 \text{ MPa K}^{-1}$  at 670 km depth (Akaogi et al. 2007). Therefore, a negative thermal anomaly across the transition zone, for example associated with the presence of subducted material, will determine simultaneously a downward deflection of the 670-km discontinuity and an upward deflection of the 410-km discontinuity, as illustrated in Fig. 1.13. Consequently, we expect that in general the topography of these primary surfaces of discontinuity be negatively correlated.

For example, the thermal anomaly associated with a stagnant slab in the transition zone could be as high as 700 K. Helffrich and Wood (2001) estimated that such a difference of temperature would determine  $\sim 30 \text{ km}$  upward offset of the

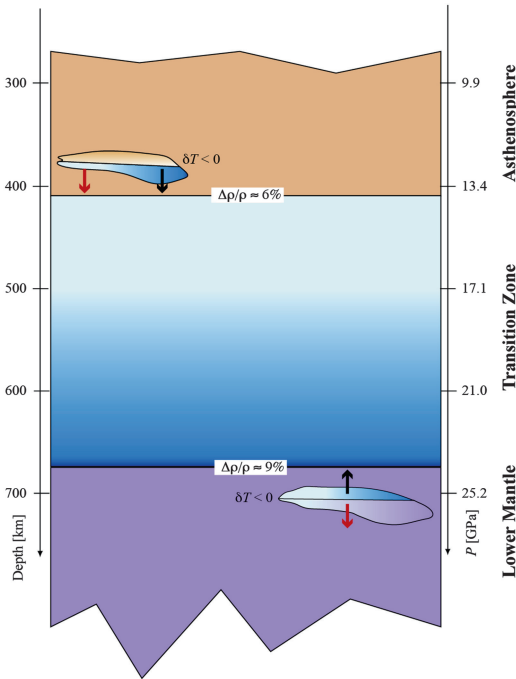
410-km discontinuity and  $\sim 60 \text{ km}$  downward depression of the 670-km discontinuity. The diverse amounts of displacement are clearly a consequence of the different magnitudes of the Clapeyron slopes associated with the transformations of  $\alpha$ -olivine and  $\gamma$ -spinel. These quantities were calculated assuming pure isochemical phase transitions at the transition zone boundaries, in agreement with the pyrolite model. Although some seismic studies have confirmed the existence of these deflections (e.g., Shearer and Masters 1992), their negative correlation and the predicted relative magnitude are not always observed experimentally (Bina and Helffrich 1994). However, there is a general consensus about the validity of a model based on phase transitions rather than chemical boundaries. Part of the discrepancies can be explained by taking into account that the transformations of majorite garnet and pyroxene to perovskite at the base of the transition zone and in the uppermost lower mantle have *positive* Clapeyron slopes (Weidner and Wang 1998), thereby their effect is opposite to that of spinel.

Another important consequence of the pyrolite model is represented by the effect of phase transitions on buoyancy and mantle convection, in so far as they determine variations of density and heat transfer (e.g., Christensen 1995). According to the Clausius-Clapeyron relation, the Clapeyron slope of a single-component system is given by:

$$\gamma = \frac{dp}{dT} = -\frac{\Delta S}{\Delta V} = \frac{\rho^2 \Delta S}{\Delta \rho} = \frac{\rho^2 Q_L}{T \Delta \rho} \quad (1.13)$$

where  $\Delta S$  and  $\Delta V$  are respectively the variations of entropy and specific volume,  $\rho$  is the mean density of the two phases,  $\Delta \rho$  is the density contrast between the phases, and  $Q_L$  is the latent heat. Therefore, the phase transition is exothermic if  $\gamma > 0$ , otherwise it will be endothermic.

The sketch in Fig. 1.14 illustrates the predicted behaviour of negative thermal anomalies placed in the vicinity of the 410- and 670-km discontinuities. At 410 km depth, the transition of olivine to wadsleyite has positive  $\gamma$ , so that it is exothermic. Therefore, a small downgoing body with negative thermal anomaly ( $\delta T < 0$ ) just above the



**Fig. 1.14** Buoyancy of negative thermal anomalies in proximity of the mantle transition zone boundaries. Black arrows indicate buoyancy associated with phase transitions, red arrows indicate buoyancy arising from thermal anomalies

discontinuity will be subject to phase transition to wadsleyite at shallower depth, and its density will increase by  $\sim 6\%$  after transformation. The region will acquire further negative buoyancy with respect to the surrounding asthenosphere, in addition to the original negative buoyancy associated with its negative thermal anomaly. Consequently, the region will accelerate sinking until it is assimilated by the transition zone. Although the effect of the exothermic release of latent heat is to *increase* the temperature by  $\sim 100$  K, so that it would oppose sinking, if this increment of temperature also occurs in the surrounding mantle then the contribution of the latent heat to the forces balance is negligible.

Let us consider now a region characterized by negative thermal anomaly, which has travelled across the upper mantle and is now located just below the 670-km discontinuity. In this instance, the phase transition has negative  $\gamma$  and is endothermic. The downgoing blob, which

has negative buoyancy because it is colder than the surrounding mantle, will be subject to phase transition from ringwoodite to perovskite and magnesiowüstite at a depth greater than 670 km, and only after this transformation its density will increase by  $\sim 9\%$  (Fig. 1.14). Therefore, in this instance the anomalous lower mantle region will be partially formed by less dense transition region minerals, which will give a *positive* contribution to the total buoyancy. If such a positive contribution balances the negative thermal buoyancy, then the body will stop sinking. Consequently, the effect of the 670-km phase transition on buoyancy can potentially prevent deep penetration of subducting slabs and constitutes at least an obstacle to whole mantle convection.

## 1.7 Lower Mantle

The 670-km discontinuity potentially represents a *thermal barrier* separating the *lower mantle* from the outer shells of the Earth. In the previous section, we have shown that this discontinuity hinders the downward injection of cold material coming from the upper mantle. Following a similar reasoning, it is easy to show that it may also impede the upwelling of hot (less dense) material from the lower mantle. As a consequence, it is conceivable that convective motions and large-scale circulation within the upper mantle proceed separately and independently from any potential lower mantle convection, so that the interaction between the two layers would occur essentially through conductive transfer of heat across the 670-km discontinuity. In general, the possibility of layered convection depends on whether or not the magnitude of positive buoyancy associated with the endothermic phase transition is greater than the negative thermal buoyancy of a slab reaching the base of the transition zone. Therefore, the possibility of layered convection within the mantle cannot be affirmed on the basis of simple qualitative estimates. For example, Christensen and Yuen (1985) showed that the critical value of the Clapeyron slope for triggering layered convection depends from the Rayleigh number,  $Ra$ , a dimensionless geodynamic parameter

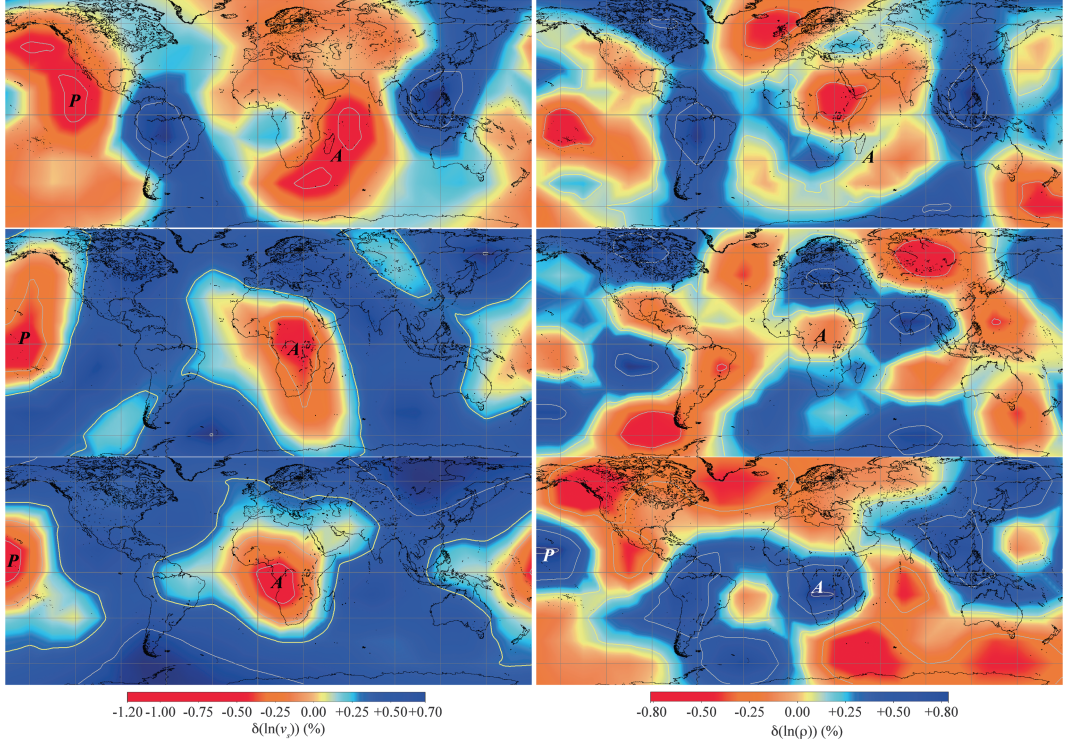
describing the relative importance of convection with respect to conduction (Chap.13). A common view is that the 670-km phase transition favors the accumulation of cold material at the base of the transition zone. When the negative thermal buoyancy of the accumulated material overcomes the positive component of buoyancy, then the slab rapidly penetrates into the lower mantle, determining the formation of a descending column of cold material and simultaneously allowing the upward injection of hot material across the 670-km discontinuity (e.g., Zhong and Gurnis 1994). In this scenario, mantle convection is normally layered, but catastrophic events of mass exchange (and associated whole-mantle convection) periodically occur to restore the equilibrium between upper and lower mantle. As pointed out by Christensen (1995), such a “hybrid” regime requires values of  $Ra$  greater than  $10^6$ , whereas for  $Ra < 10^6$  layered convection would prevail.

An alternative viewpoint has been proposed by Don Anderson and others since the early 1980s (e.g., Anderson and Natland 2005). In this model, the 670-km discontinuity represents a *chemical* boundary and convection is essentially confined to the upper mantle. However, the major point of controversy is probably the interpretation of mantle plumes. In the original formulation of Morgan (1971), these are hot and narrow regions of active upwelling that originate as instabilities within hot thermal boundary regions at the base of the mantle. This model was formulated to explain the age progression and linear arrangement of volcanic island chains and aseismic ridges, which would originate from the motion of tectonic plates over *hot-spots* fixed at the base of the lower mantle. In many modern models of mantle circulation, plumes involve large mass transport from the base of the mantle to the Earth’s surface. Together with deep slab penetration, they represent an essential feature of *whole-mantle convection* (van der Hilst et al. 1997), in which layered circulation is considered as a local transient phenomenon. The alternative scenario proposed by Don Anderson and colleagues presents a lower mantle characterized by sluggish convection (very low  $Ra$ ), where any deep mantle plume would be suppressed by the effects of

pressure on viscosity, thermal conductivity and thermal expansion (Anderson 2002). Conversely, the asthenosphere is considered as a layer characterized by local chemical heterogeneity, as well as variations of fertility and melting point, that are consequent to subduction of young plates, aseismic ridges, and seamounts, and the delamination of lower continental crust (Anderson and Natland 2005). In this view, all upwellings are passive and anomalous plume magmatism only reflects higher fertility, not higher temperature (Anderson 2006). Therefore, plumes are ultimately viewed as thermo-chemical heterogeneities, which can be transported passively by upper mantle currents (e.g., Hawaii) or sampled by migrating ridges (e.g., Iceland, Galapagos).

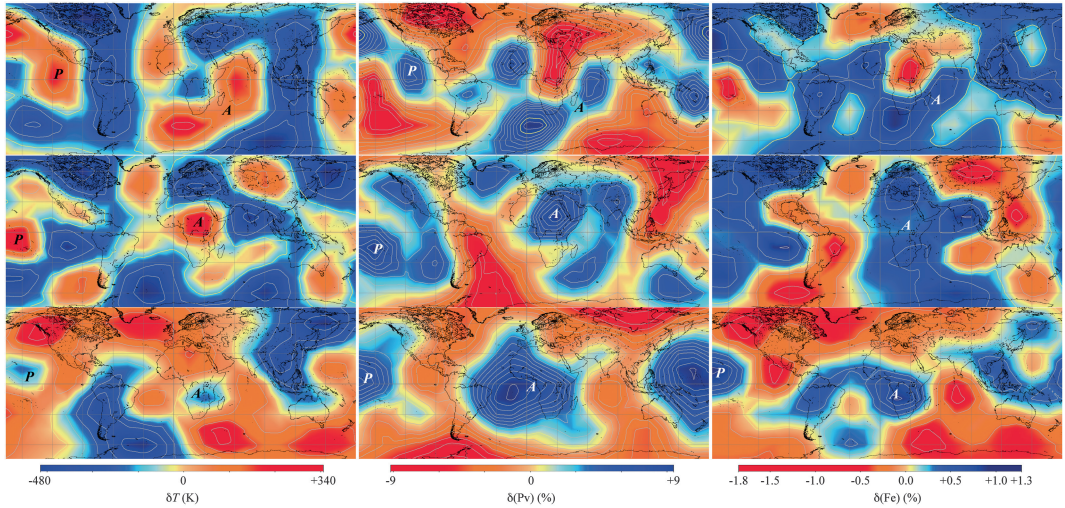
Recent advances in mantle tomography and the study of seismic scattering in the lower mantle have confirmed the presence of heterogeneities within the entire mantle. For example, Shearer and Earle (2004) have found that the upper mantle has strong wave scattering, which determines 3–4 % rms velocity heterogeneity at 4-km scale length, whereas the lower mantle heterogeneity would be only 0.5 % rms at 8-km scale length. Most importantly, three independent studies published in 2004 confirmed the existence of two large nearly antipodal thermochemical “superplumes” in the lower mantle, which are now known, respectively, as the *Pacific* and the *African superplumes* (Trampert et al. 2004; Ishii and Tromp 2004; Ritsema et al. 2004). These deep structures represent large low-shear-velocity provinces (LLSVP), which are characterized by anomalously high density *and* temperature (Fig. 1.15), which implies that they constitute *chemically* distinct regions of the lower mantle.

The chemical diversity of these regions can be described in terms of anomalously high perovskite and iron content (Fig. 1.16). The contrasting effect of high density and high temperature on the buoyancy of LLSVPs determines complicated geodynamic behaviour of these regions. As the thermal expansivity decreases with increasing depths, in the lowermost 1,000 km of the lower mantle the positive thermal buoyancy of the LLSVPs is surmounted by the negative chemical buoyancy, thereby the lower part of these regions



**Fig. 1.15** Shear wave velocity ( $v_s$ ) and density ( $\rho$ ) anomalies in the lower mantle, based on the seismic tomography model of Trampert et al. (2004). The lower mantle has been subdivided in three layers, respectively

from 2,000 to 2,891 km (*lower panels*), from 1,200 to 2,000 km (*middle panels*), and from 670 to 1,200 km (*top*). *A* and *P* are, respectively, the African and the Pacific superplumes



**Fig. 1.16** Thermal ( $\delta T$ ) and composition ( $\delta P_v$  and  $\delta Fe$ ) anomalies in the lower mantle, based on the seismic tomography model of Trampert et al. (2004). The subdivi-

sion of the lower mantle is the same of Fig. 1.15.  $\delta P_v$  and  $\delta Fe$  are, respectively, the relative variation of perovskite and iron content

is negatively buoyant (Trampert et al. 2004). In the upper part of the lower mantle, thermal and chemical buoyancies are equally important, so that the LLSVPs have neutral buoyancy. Finally, Trampert and colleagues found scarce evidence of correlation between the subducting slabs and plumes hypothesized in whole-mantle convection models and positively or negatively buoyant regions of the lower mantle.

Curiously, some succeeding studies reinterpreted the LLSVPs in terms of whole-mantle convection. For example, Burke et al. (2008) yet acknowledging that LLSVPs do not represent regions of upwelling, proposed a geometrical correlation between the locations of major hotspot volcanoes and LIPs and narrow belts surrounding the LLSVPs close to the core-mantle boundary (CMB). To this purpose, the LIPs were moved from their present day location to the position that they occupied at the time of eruption with respect to the Earth's rotation axis. These authors claimed that hot spots and reconstructed LIPs form two clusters lying vertically above the peripheries of the superplumes, which were thereby viewed as long-lived "plume generation zones" (PGZ) for at least the last  $\sim 300$  Myrs, while the remaining of the lower mantle was considered to be the "graveyard" of slabs.

In summary, the controversy between whole-mantle convection models, supported by interpreted seismic tomography data, and alternative theories of layered convection or, possibly, convection limited to the upper mantle, is still alive, and it is likely that a new quantitative approach to the analysis of seismic tomography data will be necessary to establish which model is correct. Furthermore, more data from the lowermost mantle could also contribute to assess the hypothesis that this region forms a thermal boundary layer hosting hot spots. In fact, the 250–350 km thick layer just above the CMB represents an enigmatic region where complex transformations occur. This region, which is known as the *D'' layer*, presents an anomalously low gradient of shear-wave velocity accompanied by seismic discontinuity and scattering. Furthermore, lateral variations of the depth to the *S*-wave velocity discontinuity suggests a phase transition of the

perovskite to a polymorph called *post-perovskite*. Finally, the observation of isolated pockets of ultralow seismic velocity may indicate the presence of magma chambers just above the CMB (e.g., Garnero and McNamara 2008).

---

## Problems

1. Determine analytically the melt temperature at surface for a wet peridotite having 0.05 % water content;
2. Solve Eq. (1.8) assuming a linear decrease of the coefficient of thermal expansion  $\alpha$  with the depth  $z$ ;
3. Determine the thickness of new oceanic crust formed from a 100 km wide melting regime, assuming that the degree of melting increases linearly from nearly zero at  $z = 50$  km to 24 % close to the Earth's surface and that the fraction of retained melt is 0.1 %/km;

---

## References

- Akaogi M, Takayama H, Kojitani H, Kawaji H, Atake T (2007) Low-temperature heat capacities, entropies and enthalpies of  $\text{Mg}_2\text{SiO}_4$  polymorphs, and  $\alpha$ – $\beta$ – $\gamma$  and post-spinel phase relations at high pressure. *Phys Chem Miner* 34:169–183. doi:10.1007/s00269-006-0137-3
- Anderson DL (1989) *Theory of the Earth*, 1st edn. Blackwell Scientific, Oxford, 366 pp
- Anderson DL (1995) Lithosphere, asthenosphere, and perisphere. *Rev Geophys* 33(1):125–149. doi:10.1029/94RG02785
- Anderson DL (2000) The thermal state of the upper mantle; no role for mantle plumes. *Geophys Res Lett* 27(22):3623–3626
- Anderson DL (2002) Plate tectonics as a far-from-equilibrium self-organized system. In: Stein S, Freymueller JT (eds) *Plate boundary zones*, vol 30, *Geodyn.* AGU, Washington, DC, pp 411–425. doi:10.1029/GD030p0411
- Anderson DL (2006) Speculations on the nature and cause of mantle heterogeneity. *Tectonophysics* 416:7–22
- Anderson DL, Bass JD (1986) Transition region of the Earth's upper mantle. *Nature* 320:321–328
- Anderson DL, Natland JH (2005) A brief history of the plume hypothesis and its competitors: concept and controversy. In: Foulger GR, Natland JH, Presnall DC, Anderson DL (eds) *Plates, plumes, and paradigms*, Geological Society of America Special Paper, GSA, 388., pp 119–145

- Bagherbandi M (2012) Combination of seismic and an isostatic crustal thickness models using Butterworth filter in a spectral approach. *J Asian Earth Sci* 59: 240–248
- Bass JD, Anderson DL (1984) Composition of the upper mantle: geophysical tests of two petrological models. *Geophys Res Lett* 11:229–232. doi:10.1029/GL011i003p00229
- Bina CR, Helffrich G (1994) Phase transition Clapeyron slopes and transition zone seismic discontinuity topography. *J Geophys Res* 99(B8):15853–15860. doi:10.1029/94JB00462
- Bodine JH, Steckler MS, Watts AB (1981) Observations of flexure and the rheology of the oceanic lithosphere. *J Geophys Res* 86(B5):3695–3707. doi:10.1029/JB086iB05p03695
- Boillot G, Froitzheim N (2001) Non-volcanic rifted margins, continental break-up and the onset of sea-floor spreading: some outstanding questions. *Geol Soc Lond Spec Publ* 187:9–30. doi:10.1144/GSL.SP.2001.187.01.02
- Bonatti E (1985) Punctiform initiation of seafloor spreading in the Red Sea during transition from a continental to an oceanic rift. *Nature* 316:33–37
- Burke K, Steinberger B, Torsvik TH, Smethurst MA (2008) Plume generation zones at the margins of large low shear velocity provinces on the core–mantle boundary. *Earth Planet Sci Lett* 265:49–60
- Carlson RW, Irving AJ, Schulze DJ, Hearn BC Jr (2004) Timing of Precambrian melt depletion and Phanerozoic refertilization events in the lithospheric mantle of the Wyoming Craton and adjacent Central Plains Orogen. *Lithos* 77:453–472
- Carlson RW, Pearson DG, James DE (2005) Physical, chemical, and chronological characteristics of continental mantle. *Rev Geophys* 43, RG1001. doi:10.1029/2004RG000156
- Chase CG (1979) Asthenospheric counterflow: a kinematic model. *Geophys J R Astron Soc* 56:1–18
- Chopelas A, Boehler R (1992) Thermal expansivity in the lower mantle. *Geophys Res Lett* 19(19):1983–1986
- Christensen U (1995) Effects of phase transitions on mantle convection. *Ann Rev Earth Planet Sci* 23: 65–88
- Christensen UR, Yuen DA (1985) Layered convection induced by phase transitions. *J Geophys Res* 90(B12):10291–10300. doi:10.1029/JB090iB12p10291
- Christeson GL, Morgan JV, Warner MR (2012) Shallow oceanic crust: full waveform tomographic images of the seismic layer 2A/2B boundary. *J Geophys Res* 117:B05101. doi:10.1029/2011JB008972
- Coffin MF, Eldholm O (1994) Large igneous provinces: crustal structure, dimensions, and external consequences. *Rev Geophys* 32(1):1–36. doi:10.1029/93RG02508
- Dasgupta R, Hirschmann MM (2006) Melting in the Earth's deep upper mantle caused by carbon dioxide. *Nature* 440:659–662. doi:10.1038/nature04612
- Dewey JF, Burke K (1974) Hot spots and continental break-up: implications for collisional orogeny. *Geology* 2:57–60
- Dhuime B, Hawkesworth CJ, Cawood PA, Storey CD (2012) A change in the geodynamics of continental growth 3 billion years ago. *Science* 335:1334–1336. doi:10.1126/science.1216066
- Duffy TS, Anderson DL (1989) Seismic velocities in mantle minerals and the mineralogy of the upper mantle. *J Geophys Res* 94(B2):1895–1912
- Eggin SM, Rudnick RL, McDonough WF (1998) The composition of peridotites and their minerals: a laser-ablation ICP–MS study. *Earth Planet Sci Lett* 154: 53–71
- Garnero EJ, McNamara AK (2008) Structure and dynamics of Earth's lower mantle. *Science* 320:626–628. doi:10.1126/science.1148028
- Green DH, Hibberson WO, Kovács I, Rosenthal A (2010) Water and its influence on the lithosphere–asthenosphere boundary. *Nature* 467:448–451. doi:10.1038/nature09369
- Grove TL, Kinzler RJ, Bryan WB (1992) Fractionation of mid-ocean ridge basalt (MORB). In: Morgan JP, Blackman DK, Sinton JM (eds) *Mantle flow and melt generation at mid-ocean ridges*, vol 71, *Geophys. Monogr. Ser. AGU*, Washington, DC, pp 281–310. doi:10.1029/GM071p0281
- Gu YJ, Lerner-Lam AL, Dziewonski AM, Ekström G (2005) Deep structure and seismic anisotropy beneath the East Pacific Rise. *Earth Planet Sci Lett* 232: 259–272
- Gutenberg B (1959) The asthenosphere low-velocity layer. *Ann Geophys* 12(4):439–460
- Hacker BR, Abers GA (2004) Subduction factory 3: an excel worksheet and macro for calculating the densities, seismic wave speeds, and H<sub>2</sub>O contents of minerals and rocks at pressure and temperature. *Geochem Geophys Geosyst* 5(1), Q01005. doi:10.1029/2003GC000614
- Hamilton WB (1998) Archean magmatism and deformation were not products of plate tectonics. *Precambrian Res* 91:143–179
- Hawkesworth CJ, Kemp AIS (2006) Evolution of the continental crust. *Nature* 443:811–817. doi:10.1038/nature05191
- Helffrich GR, Wood BJ (2001) The Earth's mantle. *Nature* 412:501–507
- Herzberg C, Rudnick R (2012) Formation of cratonic lithosphere: an integrated thermal and petrological model. *Lithos* 149:4–15
- Herzberg C, Asimow PD, Arndt N, Niu Y, Leshner CM, Fitton JG, Cheadle MJ, Saunders AD (2007) Temperatures in ambient mantle and plumes: constraints from basalts, picrites, and komatiites. *Geochem Geophys Geosyst* 8(2), Q02006. doi:10.1029/2006GC001390
- Hirschmann MM (2000) Mantle solidus: experimental constraints and the effects of peridotite composition. *Geochem Geophys Geosyst* 1:1042. doi:10.1029/2000GC000070

- Hirschmann MM (2006) Water, melting, and the deep Earth H<sub>2</sub>O cycle. *Annu Rev Earth Planet Sci* 34: 629–653
- Hofmann AW (1988) Chemical differentiation of the Earth: the relationship between mantle, continental crust, and oceanic crust. *Earth Planet Sci Lett* 90: 297–314
- Ishii M, Tromp J (2004) Constraining large-scale mantle heterogeneity using mantle and inner-core sensitive normal modes. *Phys Earth Plan Int* 146:113–124
- Karato S (1990) The role of hydrogen in the electrical conductivity of the upper mantle. *Nature* 347:272–273
- Karato S (2011) Water distribution across the mantle transition zone and its implications for global material circulation. *Earth Planet Sci Lett* 301:413–423
- Katz RF, Spiegelman M, Langmuir CH (2003) A new parameterization of hydrous mantle melting. *Geochem Geophys Geosyst* 4:1073. doi:10.1029/2002GC000433
- Lee C-TA (2003) Compositional variation of density and seismic velocities in natural peridotites at STP conditions: implications for seismic imaging of compositional heterogeneities in the upper mantle. *J Geophys Res* 108(B9):2441. doi:10.1029/2003JB002413
- Ligi M, Bonatti E, Caratori Tontini F, Cipriani A, Cocchi L, Schettino A, Bortoluzzi G, Ferrante V, Khalil S, Mitchell NC, Rasul N (2011) Initial burst of oceanic crust accretion in the Red Sea due to edge-driven mantle convection. *Geology* 39(11):1019–1022. doi:10.1130/G32243.1
- Ligi M, Bonatti E, Bortoluzzi G, Cipriani A, Cocchi L, Caratori Tontini F, Carminati E, Ottolini L, Schettino A (2012) Birth of an ocean in the Red Sea: initial pangs. *Geochem Geophys Geosyst* 13(8), Q08009. doi:10.1029/2012GC004155
- McKenzie D, Bickle MJ (1988) The volume and composition of melt generated by extension of the lithosphere. *J Petrol* 29(3):623–679
- McKenzie D, Jackson J, Priestley K (2005) Thermal structure of oceanic and continental lithosphere. *Earth Planet Sci Lett* 233:337–349. doi:10.1016/j.epsl.2005.02.005
- Morgan WJ (1971) Convection plumes in the lower mantle. *Nature* 230:42–43. doi:10.1038/230042a0
- Müller RD, Sdrolias M, Gaina C, Roest WR (2008) Age, spreading rates, and spreading asymmetry of the world's ocean crust. *Geochem Geophys Geosyst* 9(4), Q04006. doi:10.1029/2007GC001743
- Nicolas A (1989) Structures of ophiolites and dynamics of oceanic lithosphere. Kluwer Academic, Dordrecht, 359 pp
- Ohtani E, Toma M, Litasov K, Kubo T, Suzuki A (2001) Stability of dense hydrous magnesium silicate phases and water storage capacity in the transition zone and lower mantle. *Phys Earth Plan Int* 124:105–117
- Phipps Morgan J (1987) Melt migration beneath mid-ocean spreading centers. *Geophys Res Lett* 14(12):1238–1241
- Plank T, Langmuir CH (1992) Effects of the melting regime on the composition of the oceanic crust. *J Geophys Res* 97(B13):19749–19770
- Plomerová J, Kouba D, Babuška V (2002) Mapping the lithosphere–asthenosphere boundary through changes in surface-wave anisotropy. *Tectonophysics* 358: 175–185
- Ringwood A (1975) Pyrolite and the chondritic Earth model. In: Composition and petrology of the Earth's mantle. International Series in the Earth's and Planetary Sciences. McGraw Hill, pp 189–194
- Ritsema J, van Heijst HJ, Woodhouse JH (2004) Global transition zone tomography. *J Geophys Res* 109, B02302. doi:10.1029/2003JB002610
- Robinson PT, Malpas J, Dilek Y, Zhou M (2008) The significance of sheeted dike complexes in ophiolites. *GSA Today* 18(11):4–10. doi:10.1130/GSATG22A.1
- Rudnick RL, Gao S (2003) Composition of the continental crust. In: Rudnick RL (ed) Treatise on geochemistry, vol 3. Elsevier, Amsterdam, pp 1–64
- Schlische RW, Withjack MO, Olsen PE (2002) Relative timing of CAMP, rifting, continental breakup, and basin inversion: tectonic significance. In: Hames W (ed) The central Atlantic magmatic province: insights from fragments of Pangea, vol 136, *Geophys. Monogr. Ser. AGU*, Washington, DC, pp 33–59. doi:10.1029/136GM03
- Schmidt MW, Poli S (1998) Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet Sci Lett* 163: 361–379
- Shearer PM, Earle PS (2004) The global short-period wavefield modelled with a Monte Carlo seismic phonon method. *Geophys J Int* 158:1103–1117. doi:10.1111/j.1365-246X.2004.02378.x
- Shearer PM, Masters TG (1992) Global mapping of topography on the 660-km discontinuity. *Nature* 355:791–796
- Shimizu H, Koyama T, Baba K, Utada H (2010) Revised 1-D mantle electrical conductivity structure beneath the north Pacific. *Geophys J Int* 180:1030–1048. doi:10.1111/j.1365-246X.2009.04466.x
- Shirey SB, Richardson SH (2011) Start of the Wilson cycle at 3 Ga shown by diamonds from subcontinental mantle. *Science* 333:434–436. doi:10.1126/science.1206275
- Stacey FD (2010) Thermodynamics of the Earth. *Rep Prog Phys* 73:22 pp. doi:10.1088/0034-4885/73/4/046801
- Talwani M, Abreu V (2000) Inferences regarding initiation of oceanic crust formation from the U.S. East Coast margin and conjugate South Atlantic margins. In: Mohriak W, Talwani M (eds) Atlantic rifts and continental margins, vol 115, *Geophys. Monogr. Ser. AGU*, Washington, DC, pp 211–233. doi:10.1029/GM115p0211
- Tanimoto T, Anderson DL (1984) Mapping convection in the mantle. *Geophys Res Lett* 11:287–290. doi:10.1029/GL011i004p00287

- Trampert J, Deschamps F, Resovsky J, Yuen D (2004) Probabilistic tomography maps chemical heterogeneities throughout the lower mantle. *Science* 306:853–856. doi:10.1126/science.1101996
- van der Hilst RD, Widiyantoro S, Engdahl ER (1997) Evidence for deep mantle circulation from global tomography. *Nature* 386:578–584. doi:10.1038/386578a0
- Walzer U, Hendel R, Baumgardner J (2004) The effects of a variation of the radial viscosity profile on mantle evolution. *Tectonophysics* 384:55–90
- Weidner DJ, Wang Y (1998) Chemical- and Clapeyron-induced buoyancy at the 660 km discontinuity. *J Geophys Res* 103(B4):7431–7441. doi:10.1029/97JB03511
- White RS, McKenzie D (1989) Magmatism at rift zones: the generation of volcanic continental margins and flood basalts. *J Geophys Res* 94(B6):7685–7729
- White RS, McKenzie D, O’Nions RK (1992) Oceanic crustal thickness from seismic measurements and rare earth element inversions. *J Geophys Res* 97(B13):19683–19715. doi:10.1029/92JB01749
- White RS, Minshull TA, Bickle MJ, Robinson CJ (2001) Melt generation at very slow-spreading oceanic ridges: constraints from geochemical and geophysical data. *J Petrol* 42(6):1171–1196. doi:10.1093/petrology/42.6.1171
- Workman RK, Hart SR (2005) Major and trace element composition of the depleted MORB mantle (DMM). *Earth Planet Sci Lett* 231: 53–72
- Zhong S, Gurnis M (1994) Role of plates and temperature-dependent viscosity in phase change dynamics. *J Geophys Res* 99(B8):15903–15917. doi:10.1029/94JB00545

**Abstract**

Plate kinematics represents a fundamental sub-discipline of plate tectonics. In this chapter, I describe the geometry of plate motions independently from the geodynamic factors (forces, torques, stresses) that drive the movement or changes in the state of motion of a tectonic plate. At this stage, the focus is on modelling, in particular on plate reconstructions, thereby the general description proceeds assuming that kinematic data are already available.

## 2.1 The Continuum Mechanics Representation

Earth's crust and mantle are deformable solids, composed by a large number of closely spaced microscopic mineral grains of arbitrary shape and size. At macroscopic scale, a rigorous quantitative description of the geodynamic evolution of a rock system starts from the introduction of infinitesimal quantities, the *volume elements*  $dV$ , which represent the smallest chemically and physically homogeneous parts in which a rock assemblage can be divided. It is usually assumed that a volume element fills a continuous region of the three-dimensional space, namely a closed subset  $\mathbf{R} \subset \mathbb{R}^3$ , and has regular shape, for example a parallelepiped  $dV = dx dy dz$ . In practice, the computational techniques employed in plate tectonics often require a definition of volume elements having dimensions up to several km, depending on the scale of the problem, yet being small in relation to the total volume of the rock system.

In the continuum mechanics representation of solid Earth systems, any geophysical entity (for example, a subducting slab) is formed by a continuous distribution of small volume elements,  $dV$ , whose locations are described by position vectors  $\mathbf{r}$  in the selected reference frame. In this representation, the *intensive variables* (also known as *bulk properties*) are quantities describing *local* physical properties of the volume elements, for example their temperature, velocity, etc. It is assumed that these quantities vary smoothly across the region  $\mathbf{R}$ , so that they can be represented mathematically by continuous functions of position vectors  $\mathbf{r} \in \mathbf{R}$ . Often the intensive variables are associated with scalar fields (see Appendix 1),  $\phi = \phi(\mathbf{r})$ , having appropriate continuity properties. Typical examples are the local temperature,  $T = T(\mathbf{r})$ , and pressure,  $p = p(\mathbf{r})$ , of rocks. However, not all of the intensive variables can be represented by scalar fields. For instance, the displacement of a point  $\mathbf{r}$  during deformation must be described by a vector quantity,  $\mathbf{u} = \mathbf{u}(\mathbf{r})$ , which varies from point to point in  $\mathbf{R}$ . Therefore, intensive variables

are sometimes associated with vector or even tensor fields (see Appendix 1).

The continuum mechanics representation of Earth systems also includes *extensive variables*. These quantities are *global* physical properties, which depend from the total volume  $V$  of a system through integral expressions involving *density* functions. A classic example is the total mass of a rock body. Let  $dV$  be a volume element centered at position  $\mathbf{r}$  in the region  $\mathbf{R}$ . The approach of continuum mechanics is to consider the mass contained in  $dV$  as the analog of a point mass, so that the classic equations of elementary physics can be easily generalized to the new framework. To this purpose, we can introduce a new intensive quantity, the *density of mass*,  $\rho = \rho(\mathbf{r})$ , such that the infinitesimal mass contained in the volume  $dV$  will be given by:  $dm = \rho(\mathbf{r})dV$ . In this instance, the total mass of a body is an extensive property that can be computed by evaluating the following integral expression:

$$M = \int_{\mathbf{R}} \rho(\mathbf{r}) dV \quad (2.1)$$

Similar expressions can be written for the total electric charge, magnetization, etc. introducing appropriate density functions. If a continuous rock system is subject to an external action-at-a-distance force field, such as a gravity or magnetic field, this force operates on each volume element in  $\mathbf{R}$ . Therefore, we can introduce a *body force density* (force per unit volume),  $\mathbf{f} = \mathbf{f}(\mathbf{r})$ , such that the infinitesimal force exerted on a volume element  $dV$  will be given by:  $d\mathbf{F} = \mathbf{f}(\mathbf{r})dV$ . Using this definition, the total force,  $\mathbf{F}$ , and the torque,  $\mathbf{N}$ , exerted on the whole body are extensive variables given respectively by:

$$\mathbf{F} = \int_{\mathbf{R}} \mathbf{f}(\mathbf{r}) dV \quad (2.2)$$

$$\mathbf{N} = \int_{\mathbf{R}} \mathbf{r} \times \mathbf{f}(\mathbf{r}) dV \quad (2.3)$$

An important kinematic parameter of a point mass distribution is the *center of mass*, which is a position vector representing the location of

the entire system. In elementary mechanics, this vector is obtained by taking the weighted average of the individual position vectors, and using the mass of each particle as a weighting parameter. The continuum mechanics analogue of this quantity is another extensive variable, which can be calculated by substituting the sum appearing in the elementary definition by an integral expression.

Therefore, the center of mass of a continuous distribution is defined as follows:

$$\mathbf{R} = \frac{1}{M} \int_{\mathbf{R}} \rho(\mathbf{r}) \mathbf{r} dV \quad (2.4)$$

where  $M$  is the total mass. The last extensive variable considered here is the *angular momentum* of the system, which measures the rotational component of motion with respect to an arbitrary reference point. This quantity is usually calculated with respect to the origin of the reference frame or, alternatively, with respect to the center of mass depending on the problem under consideration. In the former case, the angular momentum is given by the following integral expression, which is an obvious extension of the elementary definition:

$$\mathbf{L} = \int_{\mathbf{R}} \mathbf{r} \times \rho(\mathbf{r}) \mathbf{v}(\mathbf{r}) dV \quad (2.5)$$

In this expression, the vector field  $\mathbf{v} = \mathbf{v}(\mathbf{r})$  represents the velocity of the mass element at position  $\mathbf{r}$ . In the next section, we shall consider a special form of expression (2.5), which is particularly useful in plate kinematics, where mass distributions represent rigid tectonic plates.

## 2.2 Euler's Theorem and Rigid Rotations

Plate dynamics and kinematics, in short plate tectonics, cannot be described using a unique mathematical apparatus and a single physical theory, because the various interacting subsystems of the solid Earth (plates, slabs, asthenosphere, etc.) conform to different physical laws, depending on the time scale of observation (seconds, years,

thousands or million years). Even when considered at a common temporal scale, these subsystems display distinct mechanical behaviors. For example, the motion of tectonic plates during the geological time (intervals of Myrs) can be described in terms of rigid body's kinematics, whereas the asthenosphere behaves as a fluid at the same temporal scale. However, both can be considered as elastic bodies if we are studying earthquakes and propagation of seismic waves in the solid Earth. In summary, the sole unifying framework of plate tectonics is the continuum mechanics representation illustrated in the previous section, while both the kinematic description of the processes and the geodynamic laws that link forces to kinematics will be different depending on the subsystem and the temporal scale of observation.

Observation suggests that tectonic plates can be considered as *rigid* bodies at first approximation. Consequently, the volume elements that fill a region  $\mathbf{R} \subset \mathbb{R}^3$ , representative of a tectonic plate, are also rigid entities, and the distance between any pair of volume elements in  $\mathbf{R}$  is an invariant. This is equivalent to say that the electromagnetic interaction between adjacent volume elements is so strong that any external force is overcome, so that deformation is negligible. In this instance, an important theorem, due to Leonhard Euler (1775), can be used as a starting point for the mathematical description of plate kinematics. The statement of Euler's theorem is very simple:

### Euler's Theorem

*If a sphere  $S$  is moved about its center,  $O$ , it is always possible to find a diameter,  $D$ , of fixed points.*

*Proof* Let  $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$  be three position vectors pointing to arbitrary points,  $P_1, P_2$ , and  $P_3$  in the original sphere. After an arbitrary change of orientation of the sphere about its center, these points are moved to new locations, say:  $P'_1, P'_2$ , and  $P'_3$ , represented by the position vectors:  $\mathbf{r}'_1, \mathbf{r}'_2, \mathbf{r}'_3$ . Let  $\mathbf{T}$  and  $\mathbf{T}'$  be the  $3 \times 3$  matrices formed with the components of these vectors:

$$\mathbf{T} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}; \mathbf{T}' = \begin{bmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \\ z'_1 & z'_2 & z'_3 \end{bmatrix}$$

Now let us define a new  $3 \times 3$  matrix:

$$\mathbf{A} = \mathbf{T}'\mathbf{T}^{-1} \quad (2.6)$$

The matrix  $\mathbf{A}$  has the property to transform the original matrix  $\mathbf{T}$  into the new matrix  $\mathbf{T}'$ :

$$\mathbf{A}\mathbf{T} = \mathbf{T}' \quad (2.7)$$

This equation implies, in turn, that  $\mathbf{A}$  transforms each vector  $\mathbf{r}_i$  into the corresponding rotated vector  $\mathbf{r}'_i$ :

$$\mathbf{A}\mathbf{r}_i = \mathbf{r}'_i \quad (2.8)$$

In general, for any position vector,  $\mathbf{r}$ , the transformation  $\mathbf{A}$  preserves the distance of the transformed point from the origin, because the sphere is assumed to be rigid:

$$\|\mathbf{A}\mathbf{r}\| = \|\mathbf{r}\| \quad (2.9)$$

Squaring this equation gives:

$$\mathbf{r}^T \mathbf{A}^T \mathbf{A} \mathbf{r} = \mathbf{r}^T \mathbf{r} \quad (2.10)$$

thereby  $\mathbf{A}^T \mathbf{A} = \mathbf{I}$ , where  $\mathbf{I}$  is the  $3 \times 3$  identity matrix, and  $\mathbf{A}$  is orthogonal. Now let us take the determinant of  $\mathbf{A}^T \mathbf{A}$ . It results:

$$\det(\mathbf{A}^T \mathbf{A}) = [\det(\mathbf{A})]^2 = 1 \quad (2.11)$$

Therefore,  $\det(\mathbf{A}) = \pm 1$ . If we consider a null rotation of the sphere from its initial position, then  $\mathbf{A} = \mathbf{I}$  and  $\det(\mathbf{A}) = \det(\mathbf{I}) = +1$ . By continuity, any subsequent infinitesimal rotation or sequence of rotations must give  $\det(\mathbf{A}) = +1$ . Furthermore, by the orthogonality of  $\mathbf{A}$  we have:

$$\mathbf{A}^T \mathbf{A} - \mathbf{A} = (\mathbf{A}^T - \mathbf{I}) \mathbf{A} = \mathbf{I} - \mathbf{A} \quad (2.12)$$

$$\det(\mathbf{A}^T - \mathbf{I}) = \det((\mathbf{A} - \mathbf{I})^T) = \det(\mathbf{A} - \mathbf{I}) \quad (2.13)$$

Therefore, taking the determinant of Eq. (2.12) gives:

$$\begin{aligned} \det((A^T - I)A) &= \det(A^T - I) \det(A) \\ &= \det(A^T - I) = \det(A - I) = \det(I - A) \end{aligned} \quad (2.14)$$

Now we take into account that for any  $3 \times 3$  matrix  $\mathbf{R}$ :  $\det(-\mathbf{R}) = (-1)^3 \mathbf{R} = -\mathbf{R}$ . In the case of Eq. (2.14), it results:

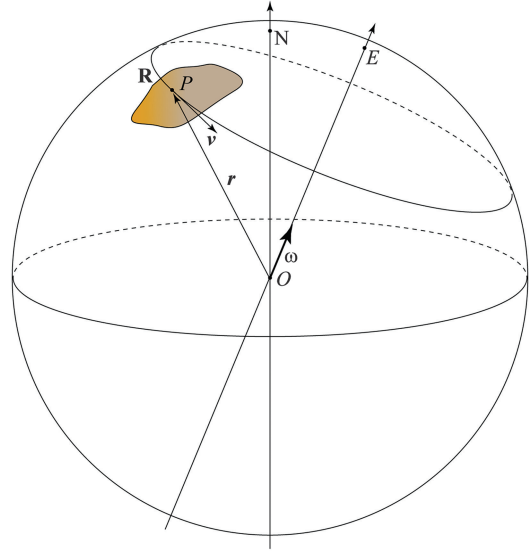
$$\begin{aligned} \det(A - I) &= \det(I - A) = -\det(A - I) \\ \iff \det(A - I) &= 0 \end{aligned} \quad (2.15)$$

Therefore, a vector  $\mathbf{n} \neq \mathbf{0}$  exists such that:

$$(A - I)\mathbf{n} = 0 \quad (2.16)$$

Equation (2.16) is a particular eigenvalue equation, where  $\mathbf{A}$  has eigenvalue  $\lambda = +1$  and  $\mathbf{n}$  is the corresponding eigenvector. This means that  $\mathbf{n}$  is invariant under transformation  $\mathbf{A}$ . Therefore, a diameter aligned with the direction of  $\mathbf{n}$  will remain unchanged after the transformation. This proves Euler's theorem. ■

The importance of Euler's theorem for the mathematical formulation of plate kinematics is not immediately evident, despite almost all books and articles about this subject emphasize its fundamental role. If we assume a spherical Earth, tectonic plates can be considered as rigid *spherical caps*, which are constrained to move about its centre. Their instantaneous motion is always represented by an infinitesimal rotation about an axis, as illustrated in Fig. 2.1, and this statement holds independently from Euler's theorem, despite it could be inferred from it. In these rotations, an arbitrary point  $P$  lying on a tectonic plate,  $\mathbf{R}$ , is moved along a *small circle arc* about the rotation axis with a velocity  $\mathbf{v}$  whose magnitude depends from the distance of  $P$  from the axis. However, the rigidity of  $\mathbf{R}$  ensures that different points will travel the same *angular distance*  $d\alpha$  in a small time interval  $dt$ , so that a unique angular velocity  $\omega = d\alpha/dt$ , which is independent from the point, characterizes the instantaneous



**Fig. 2.1** Geometry of the instantaneous motion of a tectonic plate  $\mathbf{R}$ .  $E$  is the Euler pole,  $N$  is the North Pole.  $P$  is a representative point on  $\mathbf{R}$ , whose instantaneous linear velocity is  $\mathbf{v}$ .  $\omega$  is the Euler vector of  $\mathbf{R}$

rotation. We can easily build a vector, which contains all the information associated with an instantaneous rotation. Such a vector is called an *Euler vector* and has magnitude  $\omega$  and direction coinciding with the direction of the rotation axis (Fig. 2.1). In order to assign a unique versor,  $\mathbf{n}$ , to an Euler vector, we conform to the common practice of assuming that all rotations are counter-clockwise rotations.

In this instance, the linear velocity of a point  $P$  can be calculated by the following formula:

$$\mathbf{v} = \omega \times \mathbf{r} \quad (2.17)$$

where  $\omega = \omega \mathbf{n}$  is the Euler vector of the instantaneous rotation. In general, the motion of a tectonic plate proceeds through a sequence of infinitesimal rotations about continuously changing Euler axes. Thus, in principle, reconstructing its position at a given time in the geologic past would require a backtracking procedure, based upon a complete knowledge of the sequence of instantaneous rotations. However, the standard approach adopted in plate kinematic modelling

follows an opposite pathway, which starts from a specification of the orientations of tectonic plates at some ages in the geologic past, *before* any determination of the sequence of instantaneous rotations that carried them to the present day location. The reason is that in most cases the only information available in advance, through the analysis of magnetic or structural data, is represented by the relative positions of the tectonic plates at some ages in the geologic past. These relative positions link the orientation of a tectonic plate at any time  $t$  directly to its present day position, regardless of the specific trajectory followed by the plate. The importance of Euler's theorem is just to ensure the existence of a unique invariant axis associated with such transformation from the present day position to the orientation at any time  $t$ . This means that we can always find a

rotation axis  $\mathbf{n}$  and a *finite* angular displacement,  $\Omega$ , such that a present day tectonic plate can be moved to the location that it occupied at time  $t$ , even before knowing the details of the complex sequence of instantaneous rotations that link the past position to the present day location.

Euler's theorem implies that rotations can be composed to furnish other rotations. Therefore, two transformation matrices  $\mathbf{A}$  and  $\mathbf{B}$  can be multiplied to give a new rotation matrix  $\mathbf{C} = \mathbf{AB}$  and this operation is not commutative ( $\mathbf{AB} \neq \mathbf{BA}$ ). The set of all transformation matrices associated with the rigid rotations of a sphere forms a group known as the SO(3) group (special orthogonal group in  $\mathbb{R}^3$ ). It can be shown that the orthogonal matrix associated with a finite rotation by an angle  $\Omega$  about an axis represented by the unit vector  $\mathbf{n}$ , is given by:

---

$\mathbf{R}(\mathbf{n}, \Omega)$

$$= \begin{bmatrix} n_x^2(1 - \cos \Omega) + \cos \Omega & n_x n_y(1 - \cos \Omega) - n_z \sin \Omega & n_x n_z(1 - \cos \Omega) + n_y \sin \Omega \\ n_x n_y(1 - \cos \Omega) + n_z \sin \Omega & n_y^2(1 - \cos \Omega) + \cos \Omega & n_y n_z(1 - \cos \Omega) - n_x \sin \Omega \\ n_x n_z(1 - \cos \Omega) - n_y \sin \Omega & n_y n_z(1 - \cos \Omega) + n_x \sin \Omega & n_z^2(1 - \cos \Omega) + \cos \Omega \end{bmatrix} \quad (2.18)$$


---

The unit vector  $\mathbf{n}$  has only two independent components, thereby three independent parameters are sufficient to describe a rigid rotation. The point where a positive rotation axis intersects the Earth's surface is called *Euler pole* (Fig. 2.1), its antipodal is called the *antipole*. An Euler pole, expressed through its geographic coordinates (latitude and longitude), and a rotation angle, are the three parameters generally used in

plate kinematics to indicate a finite rotation. In Sect. 2.7 we shall learn how to use these finite rotations to describe the kinematics of tectonic plates through the geological time.

Now let us consider again the instantaneous kinematics of a tectonic plate  $\mathbf{R}$ . Using the expression (2.17), the continuum mechanics analogue of the kinetic energy,  $K$ , will be given by:

$$\begin{aligned} K &= \frac{1}{2} \int_{\mathbf{R}} \rho(\mathbf{r}) v^2(\mathbf{r}) dV = \frac{1}{2} \int_{\mathbf{R}} \rho(\mathbf{r}) \omega^2 r^2 \sin^2 \theta(\mathbf{r}) dV = \frac{1}{2} \omega^2 \int_{\mathbf{R}} \rho(\mathbf{r}) [r^2 - r^2 \cos^2 \theta(\mathbf{r})] dV = \\ &= \frac{1}{2} \omega^2 \int_{\mathbf{R}} \rho(\mathbf{r}) [r^2 - (\mathbf{n} \cdot \mathbf{r})^2] dV \equiv \frac{1}{2} I(\mathbf{n}) \omega^2 \end{aligned} \quad (2.19)$$


---

where  $\theta(\mathbf{r})$  is the angle between  $\mathbf{n}$  and  $\mathbf{r}$  and the quantity  $I(\mathbf{n})$ , which depends from plate geometry, mass distribution, and the rotation axis, is the *momentum of inertia* about the axis  $\mathbf{n}$ :

$$I(\mathbf{n}) \equiv \int_{\mathbf{R}} \rho(\mathbf{r}) [r^2 - (\mathbf{n} \cdot \mathbf{r})^2] dV \quad (2.20)$$

This quantity is a measure of the inertial resistance that a rigid plate opposes to variations of its angular velocity about a rotation axis. Another quantity that can be expressed in terms of Euler vectors is the angular momentum,  $\mathbf{L}$ . Starting from expression (2.5), we can write:

$$\begin{aligned}\mathbf{L} &= \int_{\mathbf{R}} \rho(\mathbf{r}) \mathbf{r} \times \mathbf{v}(\mathbf{r}) dV \\ &= \int_{\mathbf{R}} \rho(\mathbf{r}) \mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r}) dV \\ &= \int_{\mathbf{R}} \rho(\mathbf{r}) [r^2 \boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \mathbf{r}) \mathbf{r}] dV \quad (2.21)\end{aligned}$$

A more compact expression can be determined introducing the index notation (see Appendix 1),  $x_1 \equiv x$ ,  $x_2 \equiv y$ ,  $x_3 \equiv z$ , and Einstein's summation convention. With this notation, it is easy to prove that the angular momentum has the following simple expression in terms of Euler vectors:

$$L_i = I_{ij} \omega_j \quad (2.22)$$

where the quantities  $I_{ij}$  form a rank 2 symmetric tensor, which is known as *inertial tensor*:

$$I_{ij} = \int_{\mathbf{R}} \rho(\mathbf{r}) (r^2 \delta_{ij} - x_i x_j) dV; i, j = 1, 2, 3 \quad (2.23)$$

In this expression, the quantity  $\delta_{ij}$  represents the *Kronecker delta* ( $\delta_{ij} = 1$  if  $i = j$ , zero otherwise). The components of the inertial tensor depend from the mass distribution and the plate geometry, just like the moments of inertia (Eq. 2.20). Therefore, we expect that a relation exists between these quantities. It is quite evident from (2.23) that the diagonal components of  $\mathbf{I}$  coincide with the moments of inertia about the three coordinate axes:

$$\begin{aligned}I_{ii} \equiv I_i &= \int_{\mathbf{R}} \rho(\mathbf{r}) (r^2 - x_i^2) dV = I(\mathbf{e}_i); \\ i &= 1, 2, 3\end{aligned} \quad (2.24)$$

where  $\mathbf{e}_i$  ( $i = 1, 2, 3$ ) are the base versors of the coordinate system. In general, it is possible to show that the momentum of inertia of a tectonic plate about an arbitrary rotation axis  $\mathbf{n}$  can be expressed as a linear combination of the components of the inertial tensor, thereby this tensor contains all the relevant information for the determination of the moment of inertia about any rotation axis. In fact, using Eqs. (2.20) and (2.21) we see that the component of the angular momentum in the direction of  $\mathbf{n}$  is given by:

$$\mathbf{L} \cdot \mathbf{n} = \int_{\mathbf{R}} \rho(\mathbf{r}) [r^2 \omega - (\mathbf{n} \cdot \mathbf{r})^2 \omega] dV = \omega I(\mathbf{n}) \quad (2.25)$$

Using Eq. (2.22), and taking into account that  $\omega_j = \omega n_j$ , we can also write:

$$\mathbf{L} \cdot \mathbf{n} = L_i n_i = n_i I_{ij} \omega_j = \omega n_i I_{ij} n_j$$

Therefore, a comparison with Eq. (2.25) furnishes:

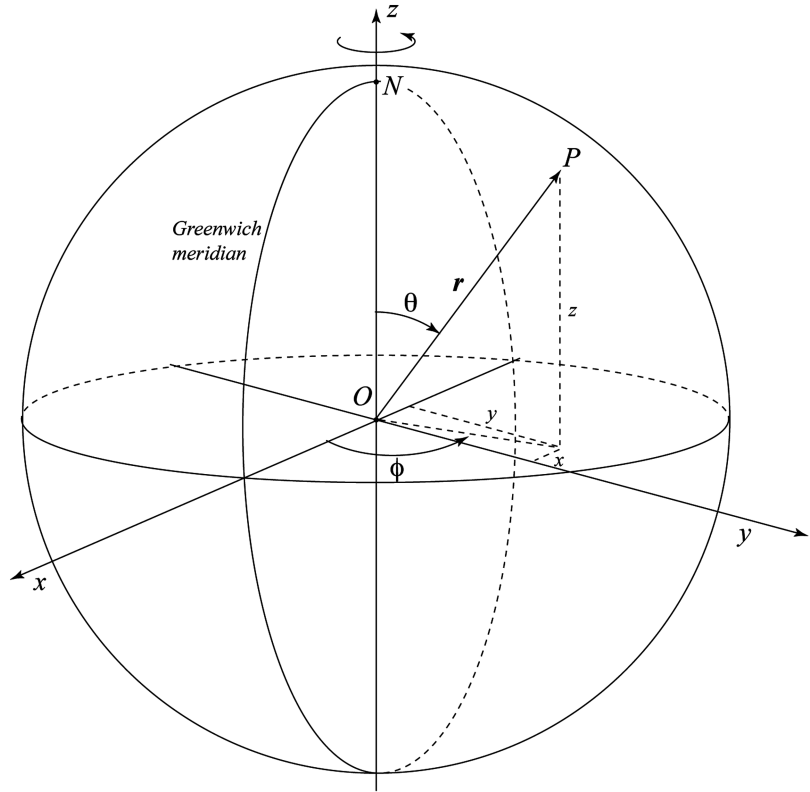
$$I(\mathbf{n}) = n_i I_{ij} n_j \quad (2.26)$$

This expression proves our statement. The previous equations represent the basic framework for the description of the instantaneous kinematics of any rotating rigid plate, independently from the choice of a reference frame. In the next section, we shall consider the specific frames of reference used in plate tectonics.

## 2.3 Reference Frames

Two broad classes of reference frames are used in plate tectonics. *Geocentric reference frames* are *global* frames that are built assuming that the Earth's centre of mass,  $\mathbf{R}$ , coincides with the origin of a Cartesian system of coordinates, so that  $\mathbf{R} = \mathbf{0}$ . The best known of these reference frames is the usual *geographic coordinate system*, in which the  $z$  axis coincides with the Earth's spin axis, and the  $x$  and  $y$  axes are in the Equatorial plane and point, respectively, to the Greenwich

**Fig. 2.2** Cartesian  $(x,y,z)$  and spherical  $(r,\theta,\phi)$  coordinates of a point  $P$  in the geographic reference frame



meridian and  $90^\circ\text{E}$ . Clearly, a point in the city of London, on the Eurasian plate, has constant longitude  $\phi = 0$  in this reference frame (Fig. 2.2).

In plate kinematics, the Earth is assumed to have a spherical shape, so that the Cartesian coordinates  $(x,y,z)$  of a point at distance  $r$  from the Earth's centre are related to the geographic coordinates  $(\theta,\phi)$ , colatitude and longitude, by the following equations:

$$\begin{cases} x = r \sin \theta \cos \phi \\ y = r \sin \theta \sin \phi \\ z = r \cos \theta \end{cases} \quad (2.27)$$

Figure 2.2 illustrates the relation between Cartesian and geographic (spherical) coordinates of a point. Equations 2.27 can be easily inverted to get an expression of the spherical coordinates as a function of the Cartesian components:

$$\begin{cases} \phi = \arctan(y/x) \\ \theta = \arccos(z/r) \\ r = \sqrt{x^2 + y^2 + z^2} \end{cases} \quad (2.28)$$

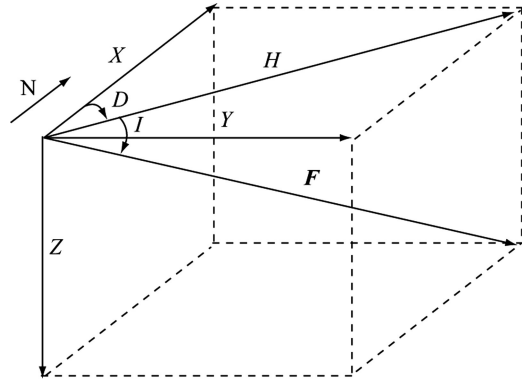
Another useful geocentric reference frame is the *geomagnetic coordinate system* (e.g., Campbell 2003). This frame is built on the basis of the observation that the present day Earth's magnetic field can be approximated as the field generated by a magnetic dipole placed at the Earth's centre, as we shall see in Chap. 4. Such a dipole has not fixed direction, but precedes irregularly about the North Pole according to the so-called *secular variation* of the core field. It is mathematically represented by a magnetic moment vector,  $\mathbf{m}$ , which currently (December 31st 2013) points to a location placed in the southern hemisphere, at about  $(80.24^\circ\text{S}, 107.46^\circ\text{E})$ . This location is called the *geomagnetic South Pole*, and its antipodal point at  $(80.24^\circ\text{N}, 72.54^\circ\text{W})$  is known as the *geomagnetic North Pole*. The axis passing through these two points defines the  $z$ -axis of the geomagnetic reference frame. The  $x$ -axis of this coordinate system is chosen in such a way that the prime meridian passes through the geographic South Pole. Finally, the  $y$ -axis will be also placed in the geomagnetic dipole equator,  $90^\circ$  from the  $x$ -axis.

In a *paleomagnetic reference frame*, the  $z$ -axis always coincides with the apparent directions of the Earth's spin axis as determined by a sequence of *paleomagnetic* poles (Chap. 6). In these frames, the longitude of a point is relative to an arbitrarily selected location on a reference continent. For example, if the central African craton is chosen to be the reference continent, then we could select a reference site in central Africa and assign to this location a fixed longitude coinciding with the present day value. This approach can be found in Besse and Courtillot (1988). Other more complex techniques assign a changing longitude (in the paleomagnetic reference frame) to the reference site according to specific algorithms (e.g., Schettino and Scotese 2005), but in any case the longitude of any other point is referred to this site and not to the Greenwich meridian.

The second broad class of reference frames is represented by *local coordinate systems*, which have the following common features: (a) the origin is an observation point at the Earth's surface (seismic station, magnetic field measurement point, etc.); (b) the  $z$ -axis is aligned with the vertical to the observation point (plumb line), so that the  $xy$  plane is a tangent plane to the Earth's surface. These reference frames are usually employed to represent the geometry of faults, focal mechanisms of earthquakes, and magnetic field measurements, but they can be used to characterize any local vector or tensor quantity of geophysical interest (Cox and Hart 1986). Figure 2.3 illustrates the conventions used in geomagnetism, where the  $z$ -axis is directed downwards, the  $x$ -axis is directed northwards, and the  $y$ -axis is directed eastwards. In this instance, the Earth's core field vector,  $F$ , can be represented by three Cartesian components ( $X, Y, Z$ ) or, alternatively by its *declination*,  $D$ , by an *inclination*,  $I$ , and a magnitude,  $F$ .

From Fig. 2.3, we see that the equations of transformation from ( $F, D, I$ ) to ( $X, Y, Z$ ) are:

$$\begin{cases} X = F \cos I \cos D \\ Y = F \cos I \sin D \\ Z = F \sin I \end{cases} \quad (2.29)$$



**Fig. 2.3** Local Cartesian components of the Earth's main field,  $F = (X, Y, Z)$  and horizontal component,  $H$ . The declination,  $D$ , is the azimuth of  $H$ , while the inclination,  $I$  is the angle between  $F$  and  $H$ , positive downward

The inverse transformation can be easily obtained from these expressions. It follows that:

$$\begin{cases} D = \arctan(Y/X) \\ I = \arcsin(Z/F) \\ F = \sqrt{X^2 + Y^2 + Z^2} \end{cases} \quad (2.30)$$

Finally, from the definition of horizontal component,  $H = \sqrt{X^2 + Y^2}$ , it follows that the inclination can be also expressed as a function of  $Z$  and  $H$ :

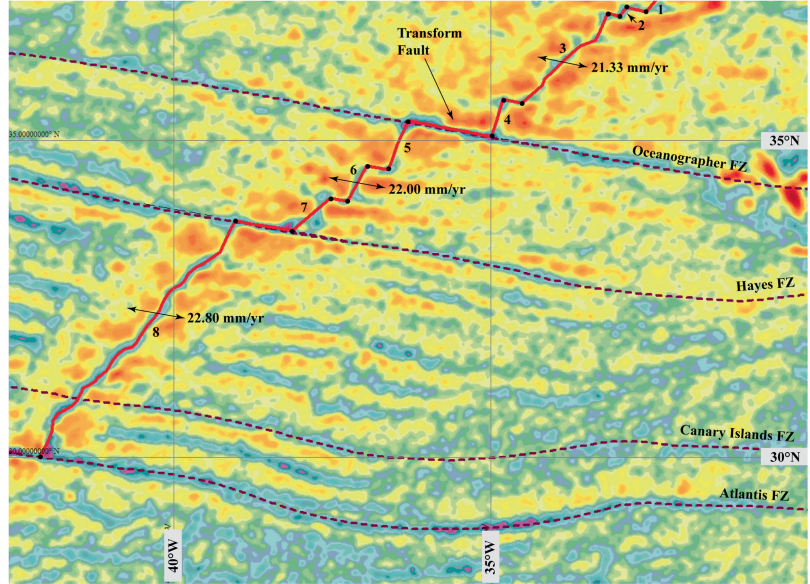
$$I = \arctan(Z/H) \quad (2.31)$$

We emphasize that although these equations refer to the specific case of the geomagnetic field, they can be used to express the components of any other vector quantity in a local coordinate system at the Earth's surface.

## 2.4 Plate Boundaries

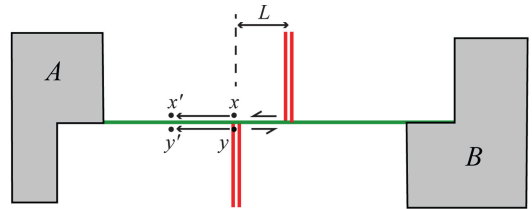
Three fundamental kinds of plate boundaries can be observed in the oceanic domain, which have three counterparts in continental areas. In the oceans, we find mid-ocean ridges, trenches, and strike-slip faults. The continental analogues of these tectonic structures are, respectively, rifts, collision zones, and transcurrent faults. Now we

**Fig. 2.4** Geometry of the Mid-Atlantic ridge in the central Atlantic area (red line). Dashed lines are major fracture zones. Black dots bound transform faults and spreading segments along the ridge. Numbers label the eight segments identified along this tract of the ridge. Double arrows indicate directions of spreading. Also shown is the full spreading rate. The background image, showing free-air gravity anomalies (Sandwell and Smith 1997), enhances the location of the axial zone of the ridge and the track of the fracture zones (FZ)



are going to consider all these faults and systems of faults from the point of view of plate kinematics. Mid-ocean ridges are extensional boundaries in the oceanic domain. We have seen in Chap. 1 that these features are spreading centers, where new oceanic crust is passively accreted as a consequence of divergent motion between two tectonic plates. These boundaries are formed by sequences of ridge segments that are linked together by transform faults, as illustrated in Fig. 2.4.

Transform faults are faults with a pure strike-slip kinematics and a strike that reflects the local direction of instantaneous motion between two plates. Therefore, these faults are always parallel to velocity vectors of relative motion. This relative motion is clearly left-lateral strike-slip in the case of dextral offset of the spreading segments and vice versa. For example, all the transform faults shown in Fig. 2.4 imply left-lateral strike-slip motion. The adjective “transform” that is attributed to these tectonic features arises from the fact that they generate active bathymetric discontinuities, as far as the two plates move apart. Such discontinuities are called fracture zones and represent linear features that apparently pursue the transform faults toward the continental margins (Fig. 2.4). Therefore, the latter seem to be “converted” into a different



**Fig. 2.5** Age discontinuities across fracture zones. Two points,  $x$  and  $y$ , on plate A, having very different ages, are at contact through a transform fault. When the two points are moved to locations  $x'$  and  $y'$ , their difference of age,  $\Delta T$ , is conserved

class of faults, characterized by vertical slip. The explanation of this phenomenon is quite simple.

Let  $v$  and  $L$  be, respectively, the relative velocity along a transform fault and its length, and consider a point  $x$  close to one of the ends of the fault (Fig. 2.5). The age of formation of the crust at  $x$  is  $T(x) = 2L/v$ , because the velocity of accretion on both sides of the ridge is on average  $v/2$ . This point is clearly at contact with another point, say  $y$ , which is close to the ridge and has age  $T(y) = 0$ . Therefore, the difference of age between  $x$  and  $y$  is  $\Delta T = 2L/v$ . At any successive time  $t$ , these two points will be displaced away from the ridge by the same offset, to two close locations  $x'$  and  $y'$ , independently from any change of  $v$  with time. Therefore, they

will remain neighbor points. Thus, the motion of points  $x$  and  $y$  away from the ridge will leave a linear track of age discontinuities,  $\Delta T$ , which crosses the whole oceanic part of the plate and reaches the continental margin, as illustrated in Fig. 2.5.

Furthermore, a specular trace will form on the conjugate plate, reaching the opposite continental margin. Generally, fracture zone tracks are easily identified on bathymetric or gravity anomaly maps (Fig. 2.4), because age discontinuities are always associated with bathymetric gaps. In Chap. 12, we shall prove that the depth to the sea floor increases with the crustal age, so that an age discontinuity always implies a bathymetric gap. Despite the invariance of the age discontinuity,  $\Delta T$ , along a fracture zone track, the difference of depth across the two sides changes with time, because the rate of sea floor subsidence is not a linear function of time, especially during the first  $\sim 100$  Myrs. This implies a lateral discontinuity in the amount of subsidence, so that fracture zones can be assimilated to vertical faults characterized by vertical slip. Therefore, as suggested by their name, transform faults are converted to a different class of faults, fracture zones, which are not associated with horizontal slip and do not represent plate boundaries, but simply are active bathymetric gaps associated with discontinuities in the age of the sea floor. Finally, it is necessary to keep in mind that although transform faults and ridge segments are very different tectonic features, they are part of unique plate boundaries, namely the mid-ocean spreading centers. In other words, they cannot be considered as distinct classes of plate boundaries.

Now let us consider the kinematics along mid-ocean spreading ridges. In principle, these features should be orthogonal to the relative velocity field between two plates. However, the example of Fig. 2.4 shows that the azimuth of the segments composing a mid-ocean ridge is not necessarily  $90^\circ$  from the direction of spreading. This phenomenon is called spreading obliquity, and is quantified measuring the angle between the normal to the ridge trend and the direction of a transform fault. Observation suggests that the spreading obliquity is particularly strong in the

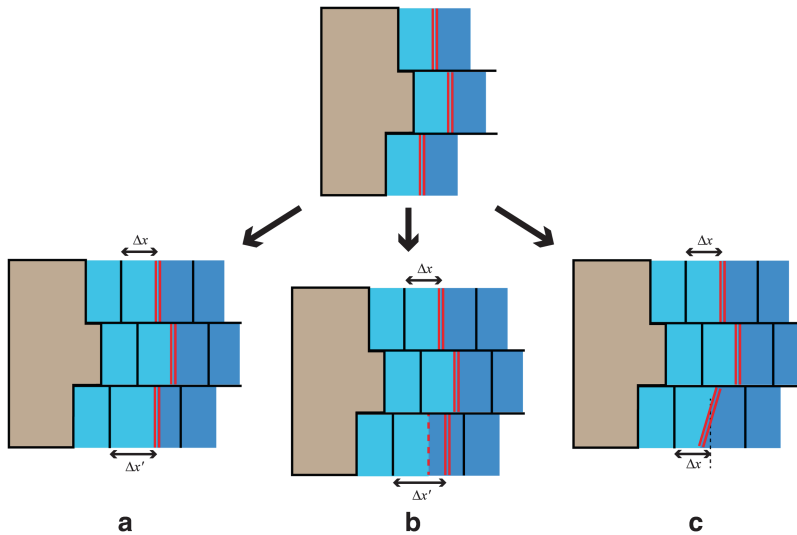
case of slow-spreading ridges (e.g., Southwest Indian Ridge and North Atlantic Ridge), where it could be as high as  $\sim 80^\circ$  (Whittaker et al. 2008). It is always necessary to take into account of this parameter when interpreting marine magnetic data. In Chap. 5, we shall learn how to deal with oblique spreading. In general, plate kinematics studies require an accurate preliminary mapping of the plate boundaries through GIS software, especially in the case of mid-ocean ridges. In this instance, the location and the geometry of the segments forming a spreading center, as well as the trace of transform faults, can be established by close inspection of the axial valley topography and by the analysis of gravity anomalies (Fig. 2.4). However, in most cases a precise definition of the ridge segments will require a successive refinement, based upon the analysis of marine magnetic anomalies, as we shall see in Chap. 5. The overall geometry of mid-ocean ridges is not constant through the geological time. It is subject to changes, even in absence of variations of relative motion, as a consequence of three basic mechanisms: spreading asymmetry, ridge jumps, and ridge segment reorientations. Figure 2.6 illustrates these three possibilities.

Spreading asymmetry occurs when the rate of accretion of new crust is not uniform across the two sides of a spreading segment (Fig. 2.6a). Let  $v$  be the full spreading rate along a ridge segment. This quantity clearly coincides with the local magnitude of the velocity vector of a plate  $A$  with respect to another plate  $B$ .

We can introduce a quantity  $-1 < \alpha < +1$ , such that the widths of the crust accreted to the right and left sides of a spreading segment in a time interval  $\Delta t$  are:

$$\Delta x_R = \frac{1}{2} (1 + \alpha) v \Delta t; \Delta x_L = \frac{1}{2} (1 - \alpha) v \Delta t \quad (2.32)$$

The quantity  $\alpha$  is an expression of the asymmetry of spreading across a mid-ocean ridge segment. In normal conditions ( $\alpha = 0$ ), a spreading segment moves at velocity  $v/2$  with respect to each of the conjugate plates. In the case of spreading asymmetry, the segment will move at a



**Fig. 2.6** Mechanisms responsible for the changing geometry of mid-ocean ridges. After a time interval  $\Delta t$ , a normal ridge segment travels a distance  $\Delta x = \frac{1}{2}v\Delta t$ . In **a** the third segment shows spreading asymmetry, because a larger amount of crust is accreted to the left side. In **b** there is no spreading asymmetry, but after a time interval

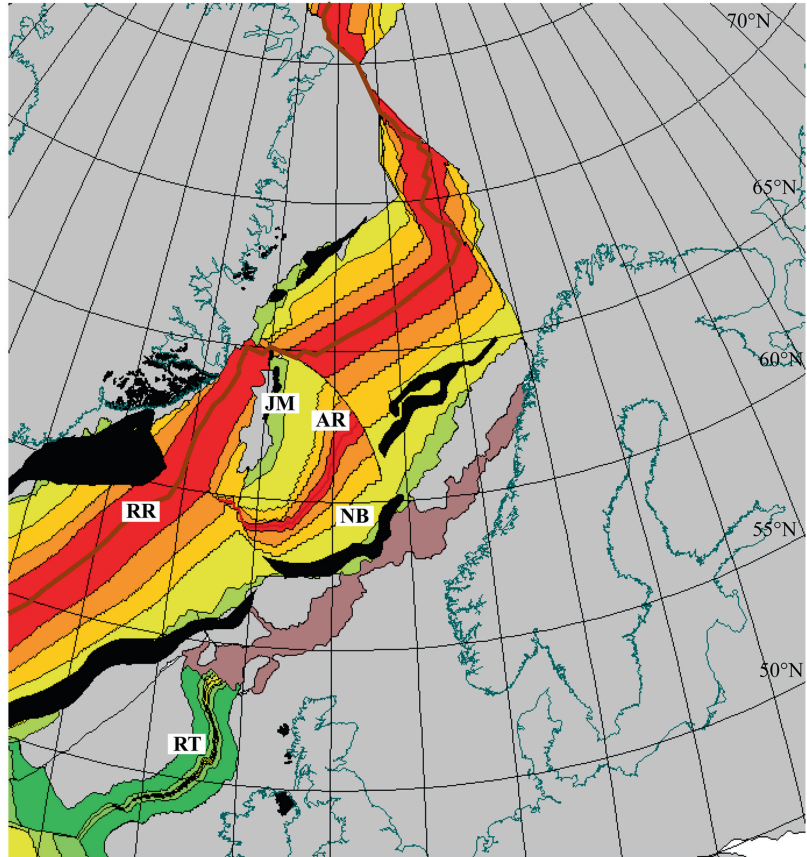
$\Delta t$  the third segment becomes extinct and a new spreading segment forms at another location, transferring a piece of oceanic lithosphere to the conjugate plate. In **c** the orientation of the third segment changes as a consequence of variations of spreading asymmetry along the same segment

velocity  $v' = (1 + \alpha)v/2$  with respect to one plate, and  $v'' = (1 - \alpha)v/2$  with respect to the conjugate plate. Therefore, any variation of asymmetry between neighbor segments of a mid-ocean ridge will modify the geometry of the plate boundary, as illustrated in Fig. 2.6a. Another mechanism, which is responsible for changing the geometry of the mid-ocean ridges, is represented by the ridge jumps (Fig. 2.6b). In this instance, a spreading segment is abandoned and a new center of spreading forms at a certain distance from the original position. The final effect is apparently similar to that associated with spreading asymmetry. However, in this instance the piece of intervening lithosphere between the old and the new segments will be accreted instantaneously to the conjugate plate, determining a complex pattern of the magnetic lineations and possible errors in the interpretation of the sea floor magnetization pattern. This problem will be considered in detail in Chap. 5. Finally, the analysis of marine magnetic anomalies shows that the geometry of the mid-ocean ridges can be modified by reorientations of the spreading segments. This phenomenon is ultimately a spe-

cial form of spreading asymmetry, which may linearly change within the same segment, determining a continuous rotation that accommodates the reorientation. The mechanism was originally proposed by Menard and Atwater (1968) to explain the effect of changes of the Euler pole of relative motion, and consequently of the spreading directions, on the mid-ocean ridge geometry. A series of subsequent studies (e.g., Hey 1977; Hey et al. 1988) proposed a much more complex mechanism for the reorientation of the spreading segments, known as ridge propagation. However, the simple process of segment rotation is a still valid model in absence of Euler pole changes.

Apart from the present day system of mid-ocean ridges, the oceanic domain also includes many extinct plate boundaries that in origin were spreading centers. Some of these extinct boundaries were produced by large-scale ridge jumps, which left wide remnant oceanic basins. Important examples are the proto-Atlantic Basin in the central Atlantic (e.g., Schettino and Turco 2009), the Mascarene Basin in the Indian Ocean (Schlich 1974), the Rockall Trough (Smythe 1989) and the Norway Basin (e.g., Gaina et al. 2009) in

**Fig. 2.7** Reconstruction of the North Atlantic region in the early Burdigalian ( $\sim 20$  Ma). The continental lithosphere is shown in *grey*. Thinned continental areas are shown in *light brown*. Regions in *black* are LIPs. Present day coastlines are shown for reference. Late Cretaceous oceanic crust is shown by *green* colours, Eocene crust is shown by *yellow* to *orange* colours, Oligocene oceanic crust is shown in *red*. *RR* Reykjanes Ridge, *RT* Rockall Trough, *NB* Norway Basin, *JM* Jan Mayen microcontinent, *AR* Aegir Ridge



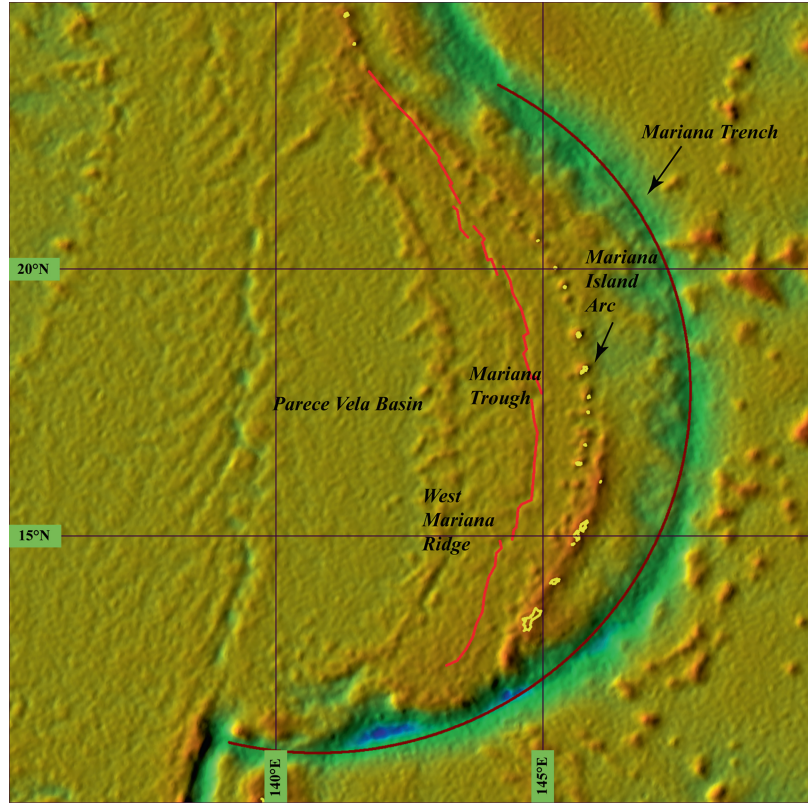
the North Atlantic. Figure 2.7 illustrates a plate reconstruction of the North Atlantic region at  $\sim 20$  Ma, shortly after a westward ridge jump that determined the extinction of the Aegir Ridge and cessation of extension in the Norway Basin. The newly formed spreading segment rifted the Greenland margin, determining the separation of a continental fragment: the Jan Mayen microplate (Jung and Vogt 1997).

Other extinct plate boundaries formed as a consequence of cessation of divergent motion between the conjugate plates, not because of a reorganization of the boundary. In this instance, a direct causal relation with a nearby onset of spreading is missing, although the final result is the same: a remnant oceanic basin and an extinct ridge testifying the former existence of divergent plate motions. Important examples of remnant oceanic basins associated with ridge extinction are the Jurassic Ligurian Basin in the western Mediterranean (Schettino and Turco 2011), the

Labrador Basin (Roest and Srivastava 1989), the South China Sea (Briais et al. 1993), the Somali Basin (e.g., Coffin and Rabinowitz 1987), the Gulf of Mexico (Ross and Scotese 1988), the Amerasian Basin (Rowley and Lottes 1988), and the Tasman Sea (Gaina et al. 1998).

The second kind of oceanic plate boundaries is represented by the trench zones (or subduction zones). These are convergent boundaries, where oceanic lithosphere bends and sinks into the asthenosphere. The structural, stratigraphic, and petrologic features associated with trenches and island arcs have been described extensively in the geologic literature (e.g., Frisch et al. 2011). Here we shall consider only some aspects that are significant for plate kinematics. The geometry of a subduction zone is that of a small circle arc, both if we consider the subducting lithosphere as a flexible-inextensible spherical shell (Frank 1968) or as a body that can be extended or shortened during the passive sinking in the mantle (e.g.,

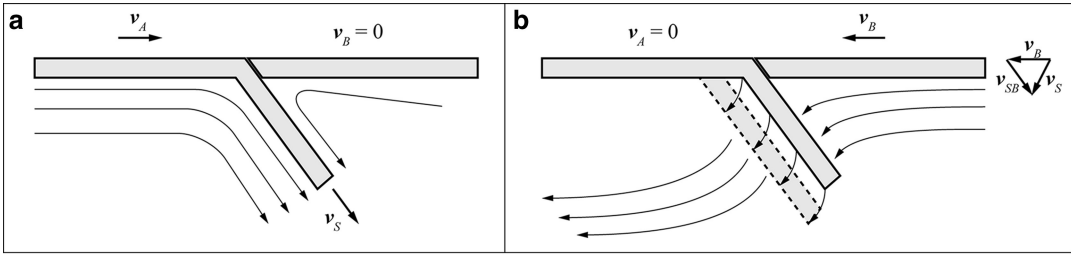
**Fig. 2.8** Geometry of the Mariana subduction zone. The *brown line* represents the best-fit *small circle* to the trench zone, determined using the algorithm of Schettino and Tassi (2012). The background image shows the free-air gravity anomaly field (Sandwell and Smith 1997). The *red line* is the active back-arc spreading center (see text). The West Mariana Ridge is an extinct island arc



Schettino and Tassi 2012). Figure 2.8 shows the example of the Mariana subduction zone in the western Pacific, which forms an almost perfect small circle arc. We have already mentioned that the geometry of tectonic plates can be represented by spherical caps. Therefore, trench zones can be viewed as the surficial hinge lines of bending of spherical caps, associated with the subduction of oceanic lithosphere. Such hinge lines do not have constant curvature, but their geometry generally varies to accommodate changes of the relative velocity field of convergence. For example, Fig. 2.8 shows that the present day Mariana Trench is deforming as a consequence of an extensional process in the back-arc area, accompanied by the formation of new oceanic crust along a back-arc spreading center in the Mariana Trough. This spreading ridge is placed to the East of an extinct island arc, the West Mariana Ridge, which is representative of a previous geometry of subduction. Almost every subduction zone shows evidence of back-arc activity, either currently or sometime in the past. Even in the case of Andean-

like trench zones, where the oceanic lithosphere is subducting beneath a continent, we observe back-arc deformation of the continental margin in the geologic past (e.g., Dalziel 1981; Ramos et al. 2002), to the point that the present day western margin of South America has the shape of a sequence of arcuate hinge lines, which can be represented anyway by small circle arcs.

An important aspect of the subduction kinematics is represented by the velocity relations between subducting plate, hinge zone, and upper plate (Fig. 2.9). Hamilton (2002) criticized what he defined a widespread misconception in the Earth Sciences community, consisting in the view that subducting plates roll over stationary hinge lines and slide down fixed slots. In reality, in a reference frame fixed to an unsubducted plate  $A$ , the hinge line  $H$  always moves toward the oceanic foreland with some velocity  $v_{HA}$ , just as the margin of the overriding plate,  $B$ , and with equal velocity (because  $v_{BH} = 0$ ). If we consider a reference frame fixed to the top of the transition zone,  $O$ , then the hinge line and the overriding



**Fig. 2.9** Cross-sections illustrating the kinematics of a subduction zone. Velocity vectors  $v_A$ ,  $v_B$ , and  $v_S$  are relative to the lower mantle,  $v_{SB}$  is the slab velocity relative to plate  $B$ . Lines in the upper mantle represent streamlines,

everywhere parallel to mantle velocity vectors (Chap. 13). Two end-member dynamic scenarios are shown. Any other dynamic scenario results from superposition of these two basic configurations

plate will move at a velocity given by:

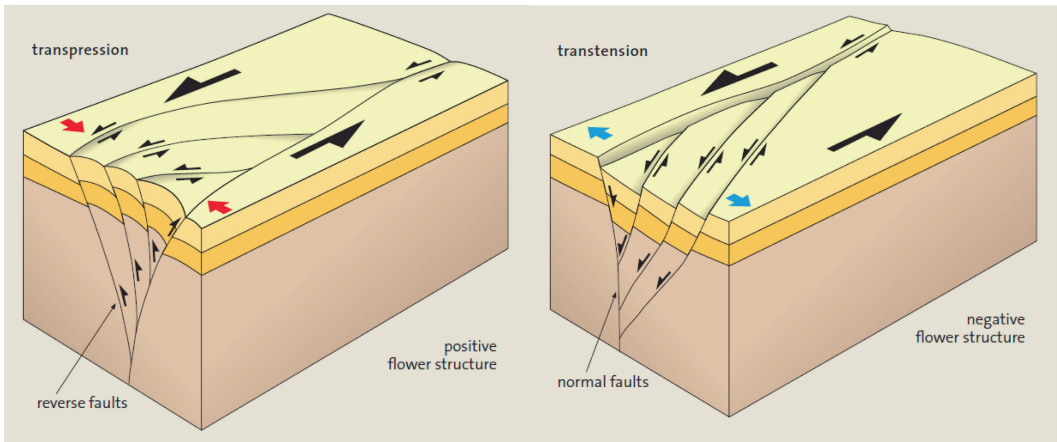
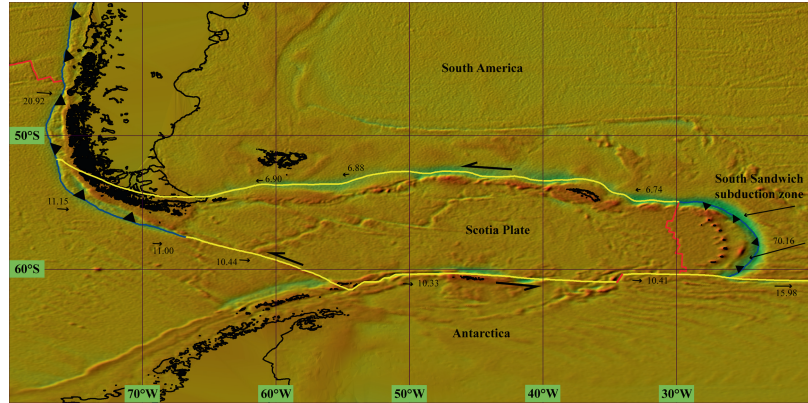
$$v_H = v_B = v_{HA} + v_A \quad (2.33)$$

If the relative hinge line velocity  $v_{HA}$  increases, for any reason, then the overriding plate margin will be subject to back-arc extension, in order to preserve the coupling between margin and hinge line, a phenomenon which is known as *trench retreat*. Conversely, any decrease of  $v_{HA}$  will lead to episodes of back-arc compression and shortening. In general, the deformation of active continental margins and island arcs is an expression of both spatial and temporal changes of the relative velocity field between the convergent plates, as well as of variations in the hinge line velocity  $v_{HA}$ . Regarding the possibility that subducting plates roll over stationary hinge lines, thereby they would slide down fixed slots, this is not a wrong interpretation of the subduction process but one of two end-member geodynamic configurations at a subduction zone, as illustrated in Fig. 2.9a. In fact, when the velocity  $v_B = 0$ , a subducting plate effectively bends and rolls over a fixed hinge line in the mantle reference frame, dragging the surrounding asthenosphere, as we shall prove in Chap. 13. In these dynamic conditions, the pull exerted by the sinking slab is balanced by the resistive viscous drag and subduction tends to proceed at constant velocity. However, the active drag exerted by the corner asthenospheric flow below the overriding plate  $B$  pulls this plate trenchward, so that  $B$  will acquire a small velocity  $v_B$  opposite to

$v_A$ . The velocity  $v_B$  increases progressively until the additional frictional resistance that is generated at the trench balances the trenchward asthenospheric drag. If the starting dynamic configuration is that illustrated in Fig. 2.9b, so that the overriding plate velocity  $v_B \neq 0$  while  $v_A = 0$ , the induced oceanic corner flow below  $A$  drags actively this plate trenchwards, thereby  $A$  will acquire a small velocity opposite to  $v_B$ . Therefore, the final equilibrium configuration will be similar to the previous one. Finally, hypothetical scenarios such that the two converging plates move in the same direction but with different velocity (in the upper mantle reference frame) cannot be stable geodynamic configuration. In fact, in this instance the excess hydrodynamic resistance exerted on the slab should be supplied by additional torques other than the known driving forces of plate tectonics (Chap. 12). Subduction roll-back is the geodynamic process determining the progressive oceanward migration of the trench zone in the upper mantle reference frame. In fact, in a reference frame fixed to the subducting plate  $A$  the hinge line always migrates toward the foreland of  $A$ .

The third kind of oceanic plate boundary is represented by strike-slip faults, which are characterized by pure or prevalent left-lateral or right-lateral motion. Differently from transform faults, which must be considered in the context of mid-ocean spreading centers, these structures are truly independent plate boundaries, which can be linked to mid-ocean ridges, trenches, or other strike-slip faults. In the present day

**Fig. 2.10** Strike-slip plate boundaries in the Scotia Sea (yellow lines). The arrows represent direction and magnitude of the relative velocity field between plates pairs. Numbers are magnitudes of velocity in mm/year

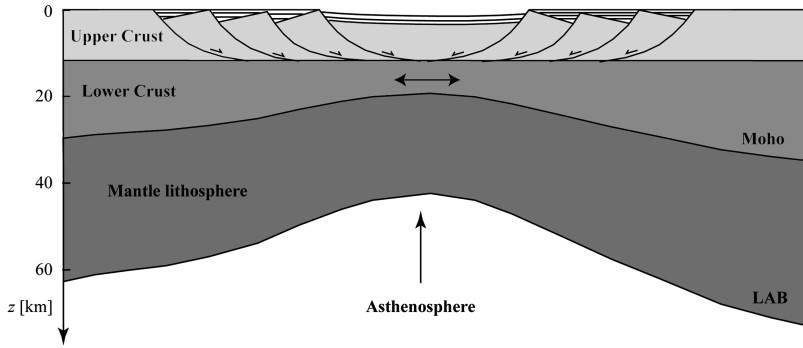


**Fig. 2.11** Flower structures associated with left-lateral transcurrent motion (From Frisch et al. 2011)

oceanic regions, only few examples can be found that belong to this class of boundaries. In the South America – Antarctica – Pacific system, the northern, the southern, and part of the western boundary of the Scotia plate are classic examples of strike-slip plate boundaries (Fig. 2.10). Other important examples are the Owen Fracture Zone in the Arabian Sea (Fournier et al. 2001), the Cayman Trough transform system along the present day North America – Caribbean plate boundary (Pindell et al. 1988), and the McDougall Ridge in the southwest Pacific (Lebrun et al. 2003).

We have mentioned that the continental counterparts of the oceanic plate boundaries are zones of active rifting, collision, or transcurrent motion. It is interesting to note that while the average relative velocity along rifts and orogenic belts is one order of magnitude less than along the

oceanic domain analogues, in the case of transcurrent boundaries the opposite is generally true. For example, the magnitude of strike-slip motion along modern transcurrent faults like the San Andreas Fault in western US (Pacific – North America boundary) and the North Anatolian Fault in Turkey (Anatolia – Eurasia boundary) is respectively of  $\sim 47 \text{ mm yr}^{-1}$  and  $\sim 29 \text{ mm yr}^{-1}$ , while less than  $8 \text{ mm yr}^{-1}$  and  $11 \text{ mm yr}^{-1}$  can be observed, respectively along the Owen Fracture Zone and around the Scotia plate. From the structural point of view, transcurrent boundaries are generally associated with characteristic flower structures at the upper crustal level, whose complexity arises from the necessity of accommodating transpressional or transtensional components within a prevalent strike-slip kinematics. Figure 2.11 illustrates the typical structural setting along these boundaries.



**Fig. 2.12** Uniform stretching in McKenzie's (1978) model of rifting

Important modern examples of transcurrent boundaries are the San Andreas (e.g., Beck 1986) and North Anatolian (Sengör 1979) faults mentioned above, the East Anatolian Fault of southern Turkey (McKenzie 1976), and the Dead Sea Fault Zone in Middle East (e.g., Garfunkel 1981; Butler et al. 1997).

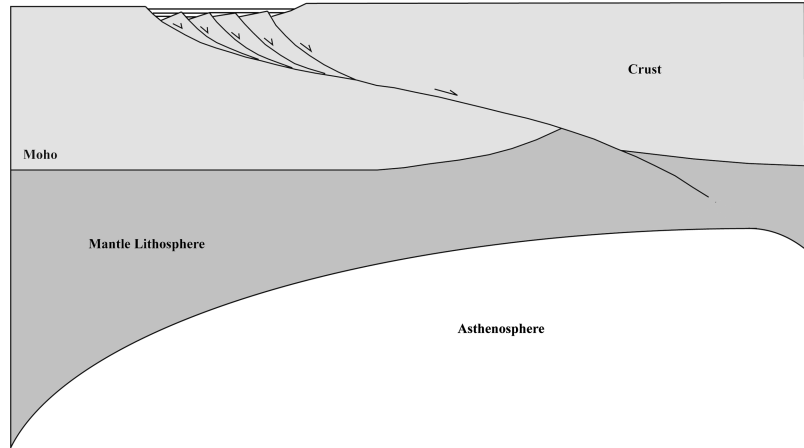
Rift zones are extensional plate boundaries, which usually exhibit complex systems of normal faults, linked together by transfer zones. The basic model of evolution of the basins associated with these fault systems is McKenzie's (1978) model, which assumes that the crustal stretching generated by an extensional force field occurs through a uniform continuous thinning of the ductile lower crust, a process that is known as pure shear. Conversely, extension in the brittle upper crust would be accommodated symmetrically by listric (upward-concave) normal faults, as shown in Fig. 2.12. In the model of McKenzie (1978), rifting events start with rapid stretching of the continental lithosphere, which determines passive upwelling of asthenosphere (Fig. 2.12). This stage is characterized by faulting of the upper crust and tectonic subsidence. Then, on a time scale of 50–100 Myrs conductive cooling of the lithosphere determines an increase of thickness and a phase of slow thermal subsidence that is not accompanied by faulting.

The amount of thinning is measured by the *stretching factor*,  $\beta$ , which is simply given by:  $\beta = H_1/H_2$ , where  $H_1$  and  $H_2$  are respectively the average thicknesses of the lithosphere at the

beginning and at the end of the phase of tectonic subsidence. Although McKenzie's model furnishes a simple and elegant picture of the thermal evolution of rift basins, many conjugate pairs of continental margins show an asymmetric pattern of faulting and the presence of exhumed lower crust, which cannot be explained by the symmetric model. Wernicke (1985) proposed an alternative asymmetric model of rifting that describes accurately many geological features of these zones (Fig. 2.13).

The key difference of Wernicke's model with respect to McKenzie's model is the recognition of low-angle detachment faults at crustal scale and simple shear (that is, localized, non-distributed shear) as the main mechanisms of lithospheric thinning during rifting. In general, several successive studies (e.g., Buck 1991; Brun 1999; Corti et al. 2003) have shown that the models of McKenzie and Wernicke must be considered as descriptions of distinct modes of rifting. Depending from crustal thickness, heat flow, and rate of extension, we can distinguish two basic modes of rifting: (a) a wide rift mode, characterized by high thinning of the crust and the mantle lithosphere over an area larger than  $\sim 100$  km, and (b) a narrow rift mode, in which extension by normal faulting is concentrated in a limited area (less than  $\sim 100$  km wide). Classic modern examples of wide and narrow rifts are respectively the Basin and Range region of western US (Hamilton 1987) and the East African Rift (e.g., Jestin et al. 1994). A transitional mode of

**Fig. 2.13** The simple shear model of rifting of Wernicke (1985)



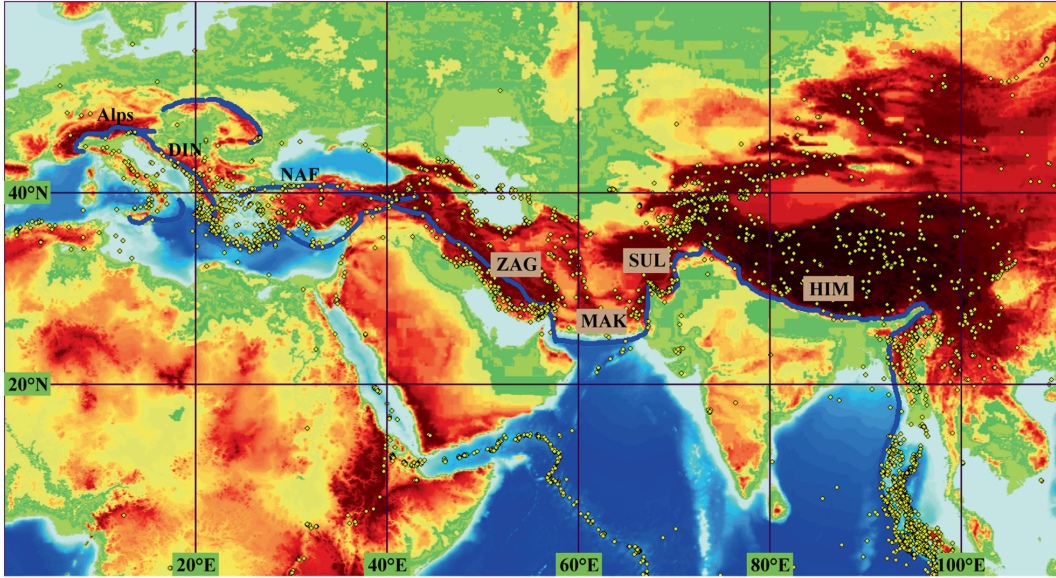
continental extension is represented by the so-called core complexes, in which high-grade metamorphic rocks from the middle to lower crust are exposed at the surface, surrounded and overlain by low-grade metamorphic rocks. This rifting mode requires high extension rates over narrow zones and an extremely weak, low-viscosity lower crust. In this instance, while extension proceeds, a lower crust inflow from the surrounding region keeps the topography uniform.

Orogenic belts represent the last kind of continental plate boundaries. These are collisional structures that should be distinguished from other kinds of mountain belts, for example from orogens associated with accretionary wedges, like the northern Apennine chain of Italy (Treves 1984), or from Andean-type orogens associated with subduction magmatism and accretion of exotic terranes. Orogenic plate boundaries are the product of continental collisions, which follow the closure of intervening oceans in the context of the Wilson cycle (see Sect. 1.3). The tectonic style of these compressive structures is sometimes called the Alpine style of orogeny (e.g., Frisch et al. 2011), essentially because the spectacular Alpine-Himalayan belt, extending from western Europe to China, is the unique example of active orogenic boundary in the modern Earth. The formation of this mountain belt started after the collision of three continental masses, Africa, Arabia, and India, with the southern Eurasian margin during the Eocene

(~50 Ma). This event followed the closure of the neo-Tethys ocean, a wide oceanic domain that existed between Gondwana and Eurasia since the early Mesozoic (e.g., Schettino and Turco 2011). The collisional structures of this orogenic belt are still active. This is confirmed both by the diffuse seismicity (Fig. 2.14) and by space geodetic observations across the mountain ranges (Kreemer et al. 2003). Figure 2.14 shows the chain of convergent and transpressive boundaries composing the Alpine-Himalayan belt. In the next section we shall learn how the set plate boundaries that are active at any given time can be linked together to form a plate tectonic configuration.

## 2.5 Triple Junctions

Both the direct observation of modern plates and plate reconstructions show that plate boundaries are joined together in groups of three, at locations that are called *triple junctions*. The lack of higher order junctions is not casual but depends from their instability. For example, it is easy to show that a four-order junction always splits into two triple junctions (e.g., Cox and Hart 1986). McKenzie and Morgan (1969) showed that there are 16 possibilities to form triple junctions by linking three plate boundaries at a point. If we designate by  $R$ ,  $T$ , and  $F$ , respectively a mid-ocean ridge, a trench, and a strike-slip fault, then



**Fig. 2.14** The Alpine-Himalayan collisional orogen. This belt is composed by several plate boundaries linked together (blue lines with labels). From the West, they are: Alps, Dinarides (*DIN*), the North Anatolian Fault (*NAF*),

the Makran subduction zone (*MAK*), Sulaiman ranges (*SUL*), and Himalayas (*HIM*). Yellow dots are earthquake epicenters in the Harvard CMT catalog

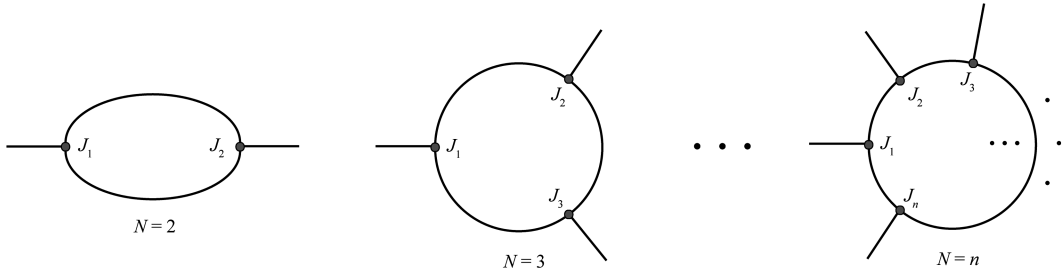
the symbol *RRR* can be used to indicate a triple junction where three spreading ridges meet, *TTR* would indicate a link between two trenches and a ridge, and so on. At any given time, a plate tectonic configuration can be represented by a graph  $G(j,b)$  (in the sense of Computer Science, e.g., see Gould 1988 and Appendix 2) having  $j$  degree-three nodes and  $b$  edges, corresponding respectively to triple junctions and plate boundaries. In this representation, which is useful to investigate the topological properties of the global system of tectonic plates, a plate  $P$  is defined by the ordered cyclic sequence  $\{J_1, J_2, \dots, J_n\}$  of triple junctions that tie its boundaries. It should be noted that in this representation the exact geometry of the plate boundaries is unessential, because what we are describing is a system of relations between tectonic plates, their interactions, not the specific geologic details that implement them. If  $J$  is an arbitrary node in  $G(j,b)$ , then  $J$  is a vertex belonging simultaneously to three adjacent cyclic sequences. Starting from this node, the sequence of triple junctions defining one of the corresponding tectonic plates is obtained applying the following simple traversal algorithm:

#### Algorithm 2.1 (Plate Traversal Algorithm)

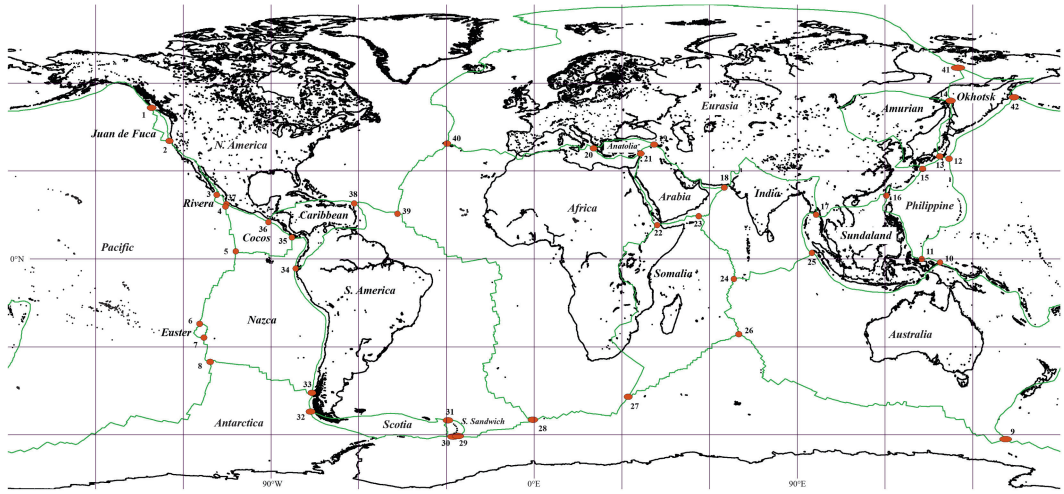
1. Select an arbitrary edge of the starting node  $J$ ;
2. Move to the neighbor node through the selected edge. If this is the starting node  $J$ , then stop;
3. Now you have two alternative (left and right) edges to leave the current node. Select the left edge;
4. Jump to step #2.

In this algorithm, the choice of one of the three plates associated with  $J$  is performed implicitly at step #1 through the selection of a starting edge. The algorithm can be used as a base for the design of more sophisticated computational procedures that investigate the structure of a global plate tectonic configuration.

The theoretical definition of tectonic plates discussed above allows to classify them according to the number of triple junctions that tie the corresponding sequence of plate boundaries. We define the order  $N$  of a tectonic plate as the number of triple junctions that are traversed applying algorithm 2.1, as illustrated in Fig. 2.15. Figure 2.16 shows the present day plate tectonic



**Fig. 2.15** Definition of the *order*  $N$  of a tectonic plate



**Fig. 2.16** Modern plate tectonic configuration and triple junctions. Plate boundaries are a slightly modified version of those proposed by Bird (2003). There are 23 major

plates, 42 triple junctions (red circles), and 63 plate boundaries (green lines)

configuration, assuming a simplified version of the plate boundaries proposed by Bird (2003). This configuration does not include plates and microplates whose relative velocity is negligible with respect to adjacent plates, for example Adria in the central Mediterranean, or that are expression of forearc deformation, such as the Burma platelet in Southeast Asia (e.g., Vigny et al. 2003). In addition, it does not include small triple junction microplates such as the Juan Fernandez (e.g., Anderson-Fontana et al. 1986) and Galapagos (e.g., Lonsdale 1988) microplates in the Pacific. It is not difficult to prove that two simple equations allow to express the number of plates  $p$  as a function of the number of plate boundaries,  $b$ , and triple junctions,  $j$ .

They are:

$$\begin{cases} b = 3(p - 2) \\ j = 2(p - 2) \end{cases} \quad (2.34)$$

In fact, if we start from a hypothetical Earth with only three plates ( $p = 3$ ), then it is quite evident that we have  $b = 3$  and  $j = 2$ . In order to generate a new plate, we must split an existing one. This operation requires cutting two edges of the plate through the insertion of two new triple junctions and a new edge connecting them. Therefore, for each new plate we add two triple junctions, thereby  $j$  is always even. Regarding the number of plate boundaries, although we add only one new edge, the operation of cutting two

**Table 2.1** Order of modern tectonic plates

Plate	N	Plate	N	Plate	N	Plate	N
Pacific	13	Nazca	7	Philippine	5	Scotia	3
N. America	10	Australia	5	Arabia	5	Anatolia	3
Eurasia	9	India	5	Sundaland	5	Amurian	3
Antarctica	9	Somalia	5	Caribbean	4	J. de Fuca	2
S. America	8	Cocos	5	Rivera	3	Easter	2
Africa	7	Okhotsk	5	S. Sandwich	3		

existing boundaries determines further increase of the total number of plate boundaries by two units. Therefore, there are three additional boundaries for each new plate. This proves Eq. (2.34).

The present day configuration illustrated in Fig. 2.16 includes 23 plates. Thus, Eq. (2.34) requires that  $j = 42$  and  $b = 63$ . The order  $N$  of these plates is listed in Table 2.1. The order of a tectonic plate measures the degree of interaction with the global system, because it coincides with the number of neighbor plates. For example, in the modern Earth's configuration the dynamics and kinematics of the Pacific and N. American plates have the largest impact on the global plate system, because they are interacting with 14 of the remaining 21 tectonic plates.

The classification and the kinematics of triple junctions has been the subject of several studies since the 1960s (McKenzie and Morgan 1969; Patriat and Courtillot 1984; Kleinrock and Phipps Morgan 1988). The basic principle describing the instantaneous kinematics of these important tectonic features is represented by the *closure rule*. In general, if  $\omega_{AB}$ ,  $\omega_{BC}$ , and  $\omega_{CA}$  are respectively the Euler vectors of a plate  $A$  with respect to another plate  $B$ , of  $B$  with respect to a third plate  $C$ , and of  $C$  relative to  $A$ , then the closure rule simply states that:

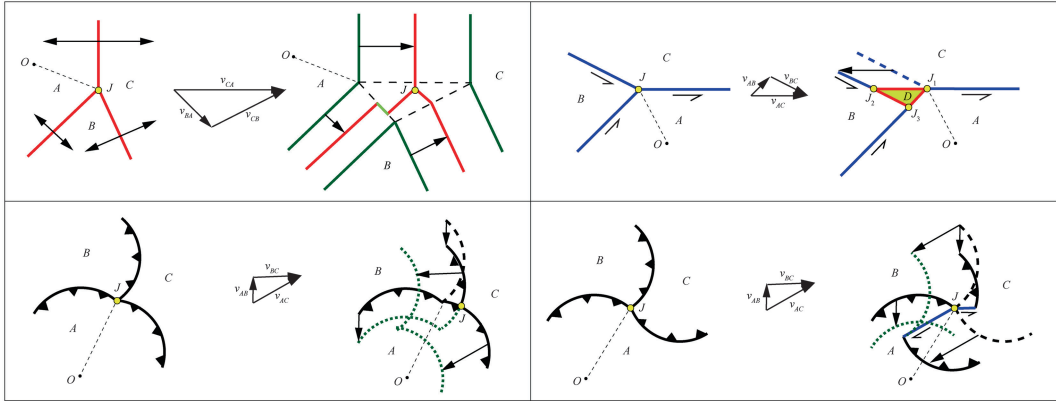
$$\omega_{AB} + \omega_{BC} + \omega_{CA} = \mathbf{0} \quad (2.35)$$

If this three-plates system is connected through a triple junction  $J$ , then this point belongs simultaneously to  $A$ ,  $B$ , and  $C$ . Therefore, applying Eq. (2.17) we have that in this case the closure rule can be expressed in terms of linear velocities at the triple junction:

$$\mathbf{v}_{AB} + \mathbf{v}_{BC} + \mathbf{v}_{CA} = \mathbf{0} \quad (2.36)$$

The velocity triangle associated with Eq. (2.36) can be used to predict the kinematics of triple junctions. The method is illustrated in Fig. 2.17 through four significant examples. It is useful to assume a reference frame fixed to one of the three plates (for example,  $A$ ). Strike-slip boundaries and trenches must be moved according to the magnitude of the relative velocity vectors. However, *trenches are always displaced with the upper (overriding) plate*, thereby they remain at rest when this coincides with the reference plate. An important geological consequence of this behaviour is represented by the development of strike-slip boundaries at triple junctions where a subduction flip occurs (Fig. 2.17 bottom right). This is a general result, which in principle may be observed along any composite flipping convergent boundary between two plates, as illustrated in Fig. 2.18.

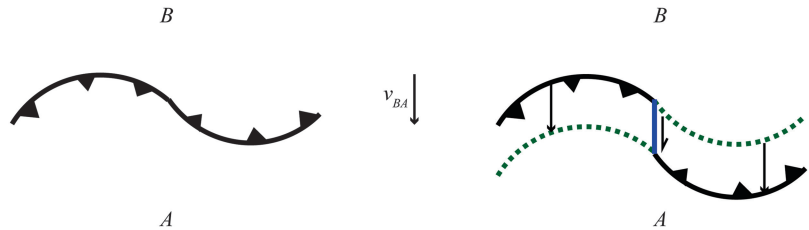
Differently from the other plate boundaries, ridges move at half of the relative velocity  $\mathbf{v}$  between two conjugate plates (Fig. 2.17 top left). In the case of an  $RRR$  junction, an extra space of triangular shape is created during the displacement of the three spreading segments, with edges given by:  $\mathbf{v}_{AB}\Delta t$ ,  $\mathbf{v}_{BC}\Delta t$ , and  $\mathbf{v}_{CA}\Delta t$ . The new triple junction will be placed within this triangle, but the link to the original segments may be somewhat complicated. It may involve either a simple propagation of the spreading segments toward the new location of the triple junction, or the formation of new spreading segments and even of a small microplate, as it is observed in the East Pacific region (Juan Fernandez and Galapagos microplates). The fact that a ridge moves at half velocity with respect to the reference plate clearly implies that any set of points located near a spreading segment at time  $t$  will be displaced



**Fig. 2.17** Evolution of RRR, FFF, and TTT triple junctions. In all panels, a reference frame with origin in  $O$  is fixed to  $A$ . *Top left*: An RRR junction. Arrows are full spreading velocity vectors at the triple junction  $J$ . The new location of  $J$  after a small time interval  $\Delta t$  can be anywhere in the central dashed triangle. The ridges link to this new location of  $J$  by propagation, development of new transforms, or oblique spreading. Dark green lines are points having the same sea floor age (isochrons). *Top right*: An unstable FFF triple junction, which collapses into a new triple junction microplate. In this example, three new RRF junctions are created that substitute the

original junction  $J$ . Dashed lines represent initial locations of plate boundaries. *Bottom*: The two different kinds of TTT junctions. Dotted lines represent subducted points of lower plate margins. The *left panel* shows a stable situation where  $A$  is always upper plate,  $B$  is both upper and lower plate, and  $C$  is always lower plate. The junction  $J$  migrates rightward along the  $A$ – $C$  boundary. The *right panel* illustrates a much more complicated configuration, in which  $A$ ,  $B$ , and  $C$  are all upper and lower plates at the same time. In this instance, the triple junction is not stable, and new strike-slip boundaries develop (blue lines)

**Fig. 2.18** Development of strike-slip faults along flipping convergent boundaries

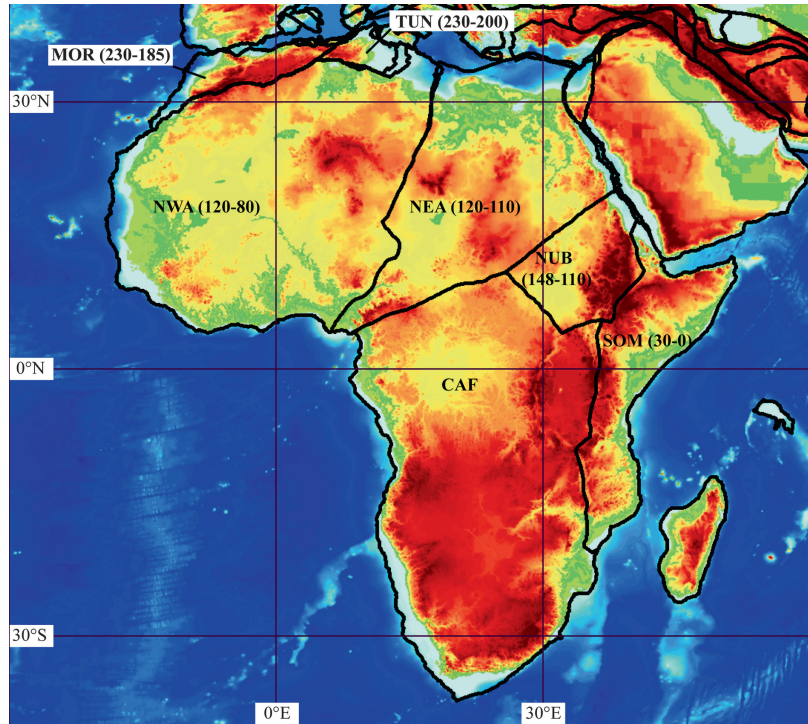


away from the ridge, after a time interval  $\Delta t$ , by a distance  $v\Delta t/2$  (Fig. 2.6). If we link all the displaced points of age  $t$  from one side of a ridge, and combine these segments with points placed along the fracture zones, we obtain a line that represents the geometry of the mid-ocean ridge at a certain time  $t$  in the past. Such a line is called an *isochron*. In Chap. 5, we shall learn the process of construction of *isochron maps*, which describe the pattern of sea-floor spreading through the geological time. For the moment, it is sufficient to note that in the reference frame of a plate  $A$ , an isochron of the conjugate plate  $B$  moves at full velocity  $v$ , just like trenches and strike-slip boundaries, as illustrated in Fig. 2.17.

## 2.6 Tectonic Elements

Computer modelling of plate kinematics has the primary objective of reconstructing sequences of past plate configurations. It requires, at a first step, the specification of the *tectonic elements* that will be included in the reconstructions. On a present day tectonic map, these are defined as *rigid crustal blocks*, bounded by paleo-faults, which have had an independent kinematic history in the geologic past (Ross and Scotese 1988). This quite general definition applies equally well to different scales of modelling (global, regional, or local) and to different structural features, such

**Fig. 2.19** Major Mesozoic and Cenozoic tectonic elements of Africa. *MOR* Morocco, *NWA* northwest Africa, *CAF* central Africa, *NEA* northeast Africa, *NUB* Nubia, *SOM* Somalia, *TUN* Tunisia. Numbers refer to the temporal range of independent motion

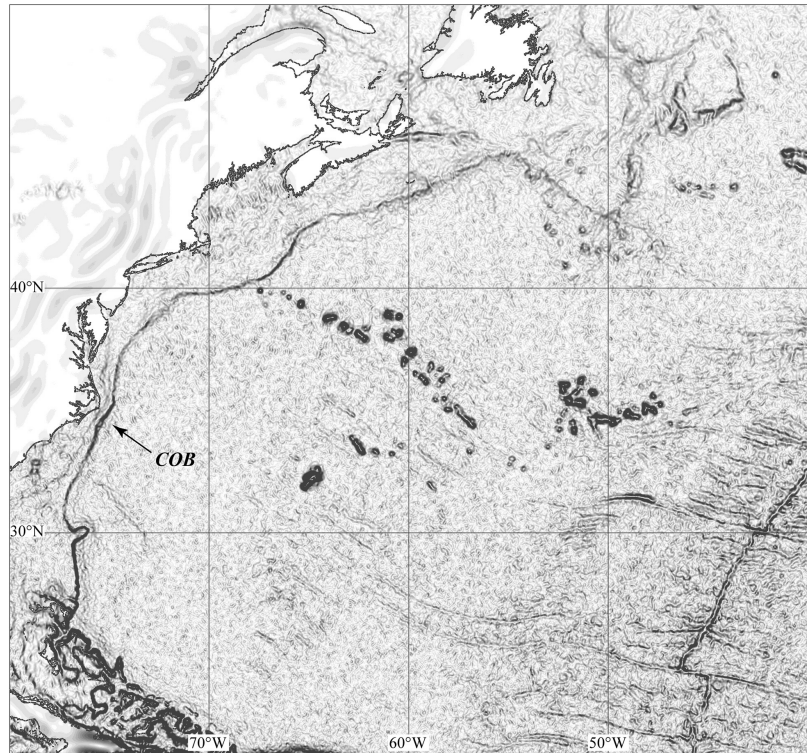


as exotic terranes, thrust sheets, microplates, slivers, continents, blocks of oceanic crust, etc. Usually the boundaries of these tectonic elements are traced using specialized GIS software, such as GlobalMapper™, but some plate tectonic modelling tools allow to draw directly the shape of the crustal blocks as *spherical polygons* (Schettino 1998, 1999a). If we normalize the Earth's radius to unity, then a spherical polygon representing a tectonic element is a cyclic ordered sequence of  $n$  unit vectors,  $P = \{r_1, r_2, \dots, r_n\}$ , placed along its boundaries. The greater is the number of polygon vertices,  $n$ , the higher is the detail through which we graphically represent a crustal block. In general, the definition of the tectonic elements is a digitizing procedure, which could also require employing specialized hardware (digitizers) and thematic base maps, such as gravity or magnetic anomaly maps, topography, bathymetry, geologic maps, structural maps, etc. In any case, the tectonic elements have two time attributes that specify the temporal range of activity along their boundaries. These attributes define the time interval of existence of the crustal blocks as independent kinematic entities. For example,

Mesozoic plate reconstructions involving Africa require the specification of at least six tectonic elements to account for the deformation of this continent. They are: Morocco, Tunisia, northwest Africa, northeast Africa, central Africa, and Nubia (Fig. 2.19).

Geological field studies performed in Morocco suggest that this block moved with respect to northwest Africa only between  $\sim 230$  Ma (late Ladinian) and  $\sim 185$  Ma (Pliensbachian), during the formation of the Atlas Rift (e.g., Laville and Piqué 1991), whereas for the rest of the Mesozoic it remained fixed to northwest Africa (Schettino and Turco 2009). Similarly, northwest Africa moved with respect to central Africa only between  $\sim 120$  Ma (Aptian) and  $\sim 80$  Ma (Campanian), during an episode of extension that produced the Benue Trough, a failed arm of the northward propagating South Atlantic rift (e.g., Fairhead 1988). Therefore, it is always necessary, when defining a set of tectonic elements, to start from reliable field data, having strong geochronologic constraints, and use only faults that were active during the time interval under consideration.

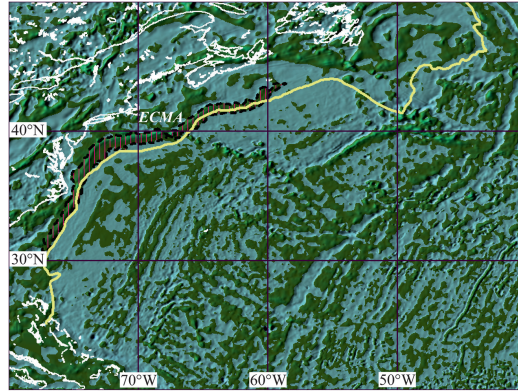
**Fig. 2.20** Eastern North American COB, based on the maximum horizontal gradient of the free-air gravity anomaly field (Sandwell and Smith 1997)



Some tectonic element boundaries are defined on the basis of geophysical constraints. An important example is represented by COBs (see Sect. 1.3). In the first plate reconstructions based on rigorous computational methods, Bullard et al. (1965) matched the conjugate continental margins around the Atlantic on the basis of a fitting algorithm. They applied the method to the 100, 500, 1,000, and 2,000 fm isobaths (1 fathom = 6 ft), then chose the best fitting set of curves (500 fm = 914.4 m) as most representative of the conjugate COBs.

Although this method produced a set of reconstructions that resisted through time and can be considered milestones in the history of plate kinematics modelling, none of the selected isobaths was really representative of the true boundary between continental and oceanic crust. The shape of bathymetry contour lines is strongly affected by the sedimentary cover. For instance, deposits associated with the delta of a great river (such as the Nile or the Ganges) can progressively shift the 1,000 m isobath seawards by hundreds of kilometers in a few

million years. The existence of such processes implies that even though conjugate isobaths may fit well at the beginning of the ocean opening, subsequent sedimentation generally decreases this initial goodness of fit. Therefore, more recent studies have adopted a geophysical definition of the COBs. For example, Schettino and Scotese (2005) assumed that the COBs were associated with the maximum horizontal gradient of the gravity anomaly field. This assumption was based on the observation that marine gravity anomaly maps adequately filter the sediment cover, and that COBs are associated with important lithological discontinuities, which give rise to sharp variations of the gravity anomaly field. In this context, free-air gravity anomaly maps such as those proposed by Sandwell and Smith (1997) represent an invaluable tool for modern plate tectonic modeling. Figure 2.20 shows a map of the horizontal gradient of the gravity anomaly field along the eastern margin of North America. The line of maximum gradient can be easily used to trace the North American COB.



**Fig. 2.21** Pattern of magnetic anomalies along the eastern North American margin and in the western central Atlantic, extracted from the global data set of Korhonen et al. (2007). Positive anomalies are shown in dark green. The location of the ECMA (vertical hatch) marks a zone

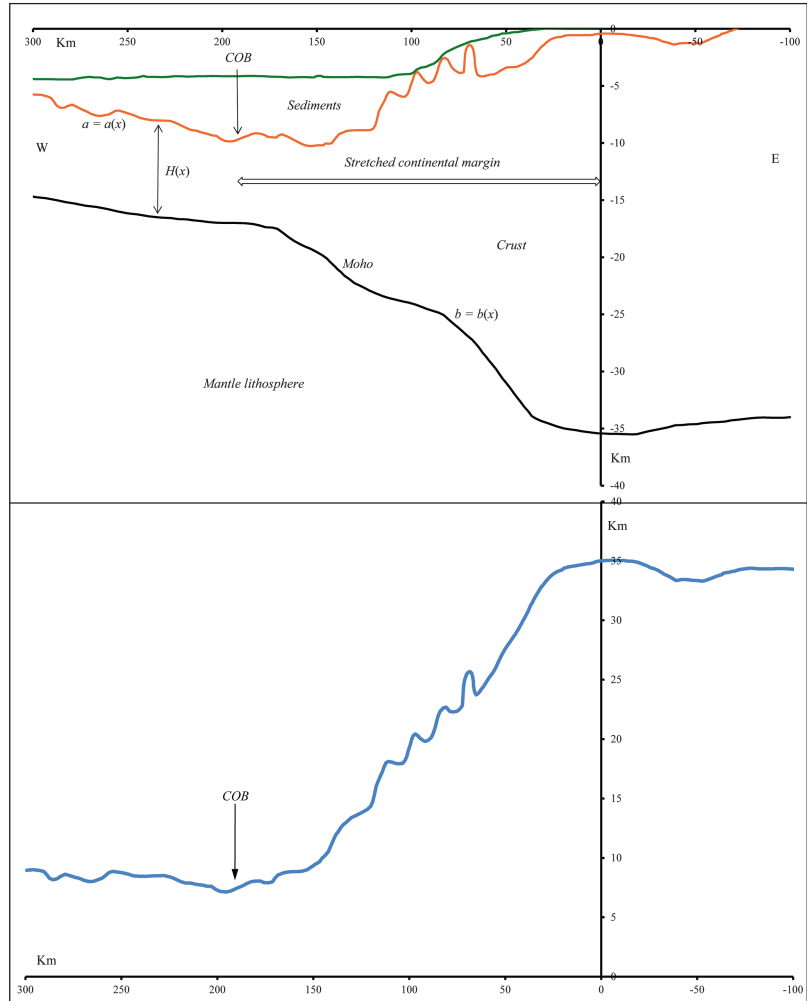
of transitional crust (or anomalous oceanic crust) along the volcanic margin. The yellow line shows the location of the COB according to the gravity data of Fig. 2.20 (maximum horizontal gradient of the gravity anomalies)

Another class of geophysical markers that are useful to identify COBs is represented by the magnetic anomalies that border volcanic OCTs (see Sect. 1.3). A classic example is represented by the East Coast Magnetic Anomaly (ECMA) of eastern North America (e.g., Schettino and Turco 2009). In this instance, a strong linear magnetic anomaly associated with the extrusives and intrusives of the initial magmatic pulse marks the site of transition from the rifting stage to drifting, hence the location of the COB, as illustrated in the example of Fig. 2.21. However, a comparison between the location (and the geometry) of the ECMA and the COB defined on the basis of gravity anomaly data (Fig. 2.20) shows that the coincidence of these features is only approximate, and that differences of up to 70–80 km exist between the two lineaments. Therefore, even when based on a geophysical approach, the definition of COBs remains to some extent qualitative.

A major problem in the definition of both COBs and tectonic boundaries that are placed along rift zones is represented by the considerable thinning that characterizes the passive margins of the corresponding tectonic elements. If we use one of the geophysical techniques described above to define a conjugate pair of COBs, then a reconstruction based on the fit of the margins will be representative of the onset of sea floor spreading, not of the pre-rift configuration. In fact, tec-

tonic elements whose extensional boundaries are defined using potential field data (either gravity or magnetic data) are *stretched* elements, which should be restored to their original size when the objective is to make a pre-rift reconstruction. There are three approaches to the solution of this problem, which clearly does not affect the reconstruction of the spreading history of oceanic basins. All these methods require an estimation of the amount of stretching that occurred during the rifting stage. This is usually expressed in terms of stretching factor  $\beta$  (see Sect. 2.4). A determination of this quantity can be made when a set of crustal profiles along the continental margins, obtained from seismic refraction experiments, is available (e.g., Schettino and Turco 2009). The first step consists into an estimation of the directions of stretching, for example by landward prolongation of the first post-rift directions of sea floor spreading. We shall see that these directions can be easily calculated on the basis of a kinematic model. Then, the seismic cross-sections are projected onto the directions of stretching, to avoid an incorrect determination of the continental margin width. At the next step, the upper and lower boundaries of the stretched crust, the latter coinciding obviously with the Moho, are identified on the cross-sections. Assuming that seismic profiles always start on unstretched crust, then these boundaries are two functions,

**Fig. 2.22** Seismic refraction profile SIS-04, offshore Morocco (Contrucci et al. 2004), projected onto the direction of Triassic-Jurassic rifting (*top*). The *green line* represents the bathymetric surface. The *bottom panel* shows a plot of the crustal thickness (excluding sediments),  $H(x) = a(x) - b(x)$ , along the margin. The stretched continental margin as width  $L \cong 193$  km. If we restore the thickness to  $H_0 = 35$  km, the width of the margin is reduced to  $L_0 \cong 110$  km (pre-rift width). Therefore, the stretching factor  $\beta$  is given by:  $\beta = L/L_0 \cong 1.75$

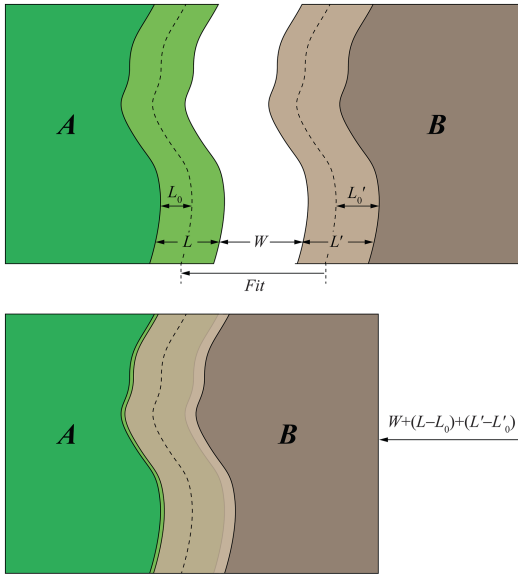


say  $a = a(x)$  and  $b = b(x)$ , of seaward increasing offsets  $x$  along the profile. The example illustrated in Fig. 2.22 shows the seismic profile SIS-04, located offshore Morocco (Contrucci et al. 2004), after projection onto the direction of Triassic-Jurassic rifting (Schettino and Turco 2009). If  $L$  is the size of the stretched margin (for example determined by the COB) and  $H_0$  is the normal unstretched crustal thickness, then the pre-rift restored size,  $L_0$ , and the stretching factor,  $\beta$ , will be given by (Fig. 2.22):

$$L_0 = \frac{1}{H_0} \int_0^L [a(x) - b(x)] dx \equiv \frac{1}{H_0} \int_0^L H(x) dx \quad (2.37)$$

$$\beta = L/L_0 = \frac{LH_0}{\int_0^L H(x) dx} \quad (2.38)$$

As soon as the stretching factor  $\beta$  has been estimated, there are three possibilities for taking into account thinning of passive margins during the syn-rift stage in pre-rift reconstructions. In a first method, the tectonic elements are defined through their present day stretched boundaries, but a pre-rift reconstruction will require an overlap of the stretched margins, as illustrated in Fig. 2.23. This method works well both in the case of passive margins of oceans and for failed



**Fig. 2.23** Overlap of the stretched continental margins in absence of palinspastic restoration. Let  $W$  be the width of the oceanic domain. In a pre-rift reconstruction, plate  $B$  is displaced by  $W + (L - L_0) + (L' - L'_0)$  with respect to the reference plate  $A$ , because the unstretched margins (dashed lines) must be matched. Therefore, the stretched continental margins will overlap by a quantity  $F = (L - L_0) + (L' - L'_0)$

rifts. The amount of overlap is locally defined by the quantity:

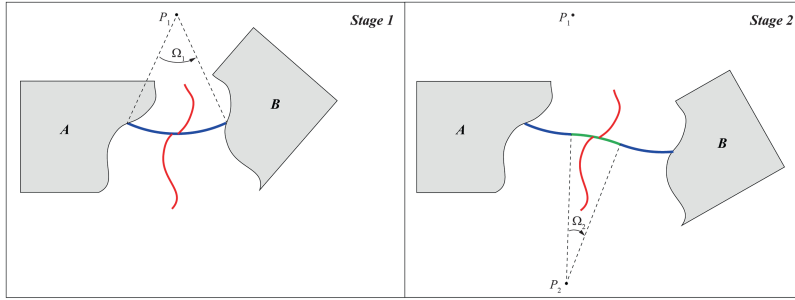
$$\begin{aligned} F &= (L - L_0) + (L' - L'_0) \\ &= L_0 (\beta - 1) + L'_0 (\beta' - 1) \end{aligned} \quad (2.39)$$

where the quantities  $L'$ ,  $L'_0$ , and  $\beta'$  refer to the conjugate plate. A second possibility requires a preliminary restoration of the tectonic elements to their unstretched shape through a process that geologists call *palinspastic restoration* (e.g., Schmid et al. 1996; Schmid and Kissling 2000; Schettino and Turco 2006). In this instance, the boundaries of conjugate tectonic elements will match in pre-rift reconstructions, but a reconstruction at the time of rift-drift transition will show a gap. Finally, a third approach could be releasing the constraint of rigid tectonic elements and allow stretching (or shortening) of the crustal blocks along the directions of relative motion. A modern computer program for

making plate reconstructions should allow usage of this method, which is particularly attractive in the case of small-scale reconstructions of collisional settings and represents a link between the classic approach of rigid plate kinematics, mostly suitable for describing the evolution of oceanic basins, and structural geology. In this approach, the amount of deformation (i.e. *strain*) of the tectonic elements should be specified directly in the kinematic model, and the shape of any crustal block included in the compilation of tectonic elements would be defined by the present day stretched or shortened margins, so that it would not be necessary anymore to perform manually a palinspastic restoration. In this instance, the procedure of restoration would be accomplished automatically by the software with the correct timing.

## 2.7 Plate Circuits and Rotation Models

Now we will consider the problem of representing the motion of a set of tectonic elements through the geological time. The kinematics of a set of tectonic plates can be described by generating a sequence of *plate reconstructions*, eventually combined into a computer animation. In this representation, the temporal range of the reconstructions is always subdivided into a series of *tectonic stages*, time intervals during which the relative motions can be described as rotations about fixed Euler axes at constant angular velocities. The existence of such time intervals, which may span several million years, is probably the most fundamental principle of plate kinematics. During a stage, it is possible to determine a set of plate pairs ( $A, B$ ), sharing common boundaries, such that the motion of  $B$  with respect to  $A$ , considered at rest in the present day position, can be represented as a rotation about a fixed Euler axis at constant angular velocity  $\omega$ . In this instance, the relative motion between any plate pair in the set occurs, within each stage, along flow lines that are small circle arcs about an Euler axis. The experimental evidence of this fundamental principle relies on the analysis of



**Fig. 2.24** Geometry of fracture zones. In a reference frame fixed to  $A$ , plate  $B$  rotates counterclockwise about pole  $P_1$  by an angle  $\Omega_1$  (stage 1). The resulting fracture zone has the shape of a small circle arc about  $P_1$  (blue

line). The subsequent clockwise rotation about pole  $P_2$  during stage 2 separates the fracture zone in two arms and inserts a new small circle about  $P_2$  in the central zone (green line)

marine magnetic anomalies and the observation that the trace of oceanic fracture zones can be invariably decomposed into chains of small circle arcs. However, we shall prove that it has a more general validity. Furthermore, this principle puts strong constraints on the equilibrium of the driving forces of plate tectonics, thus representing a conceptual linkage between plate kinematics and geodynamics.

The sketch of Fig. 2.24 illustrates the process through which a fracture zone assumes the distinctive shape of a chain of small circles. On the basis of the fundamental principle stated above, if  $(T_0, T_2, \dots, T_n)$  are stage boundaries,  $T_0 < T_1 < \dots < T_n$ , then there exist  $n$  stage poles  $P_1, P_2, \dots, P_n$  such that the relative displacement of a plate  $B$  during the  $k$ -th stage can be represented by a finite rotation matrix  $S_{BA}(T_{k-1}, T_k)$  about the axis  $\mathbf{n}_k$  associated with the  $k$ -th Euler pole  $P_k$  (Eq. 2.18). This rotation, which moves  $B$  from the location at time  $T_{k-1}$  to the position occupied at time  $T_k$ , is called a *stage rotation*. It can be represented either by matrix components (Eq. 2.18) or as a triplet  $(\lambda_k, \phi_k, \Omega_k)$ , where  $(\lambda_k, \phi_k)$  are the geographic coordinates of a stage pole  $P_k$  and  $\Omega_k$  is the finite angle of rotation during that stage (Fig. 2.24). Clearly, if a set of forward transformations  $S_{BA}(T_{k-1}, T_k)$  has been established, we can also move back through time and determine the relative position of  $B$  at time  $T_{k-1}$  given the position at time  $T_k$ . This operation simply requires application of the inverse transformation:

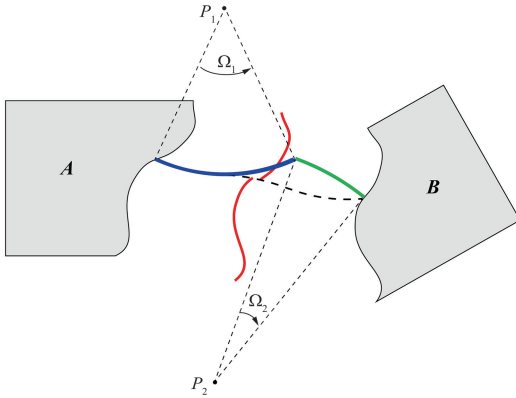
$$S_{BA}(T_k, T_{k-1}) = S_{BA}^{-1}(T_{k-1}, T_k) \quad (2.40)$$

If  $T_{k-1} \leq T \leq T_k$  is an intermediate time during the  $k$ -th stage, then the relative position of  $B$  at time  $T$  can be determined by taking as rotation angle the reduced angle  $\Omega(T)$ :

$$\Omega(T) = \frac{T - T_{k-1}}{T_k - T_{k-1}} \Omega_k \quad (2.41)$$

Now let us consider the central point of a transform fault that belongs to a present day or extinct mid-ocean ridge. Let  $\mathbf{r}_0$  be the position vector of this point. We are looking for an algorithm that simulates the geometry of the fracture zone passing through  $\mathbf{r}_0$ . The task should be accomplished calculating iteratively the location of this point, relative to each of the conjugate plates  $A$  and  $B$ , at any time  $T \in [T_0, T_n]$ , where  $T_0$  and  $T_n$  are respectively the time of onset of sea floor spreading and the present day (or the time of extinction). In this algorithm, it is necessary to take into account that for any stage the point moves by half of the full stage angle  $\Omega_k$  on each of the conjugate plates. Therefore, the point must be rotated using *reduced* backward stage pole matrices,  $S_{BA}^*(T_k, T_{k-1})$  and  $S_{AB}^*(T_k, T_{k-1})$ , having the same stage pole locations of  $S_{BA}(T_k, T_{k-1})$  and  $S_{AB}(T_k, T_{k-1})$ , respectively, but halved stage angles. If  $T_{k-1} \leq T \leq T_k$ , then the location of  $\mathbf{r}_0$  at time  $T$  on plate  $A$ ,  $\mathbf{r}_A(T)$ , will be given by:

$$\begin{aligned} \mathbf{r}_A(T) = & S_{BA}^*(T_k, T) \dots S_{BA}^*(T_{n-1}, T_{n-2}) \\ & \times S_{BA}^*(T_n, T_{n-1}) \mathbf{r}_0 \end{aligned} \quad (2.42)$$



**Fig. 2.25** Construction of flow lines. The sequence of stage rotations is the same of Fig. 2.24, and the *dashed line* shows the corresponding fracture zone

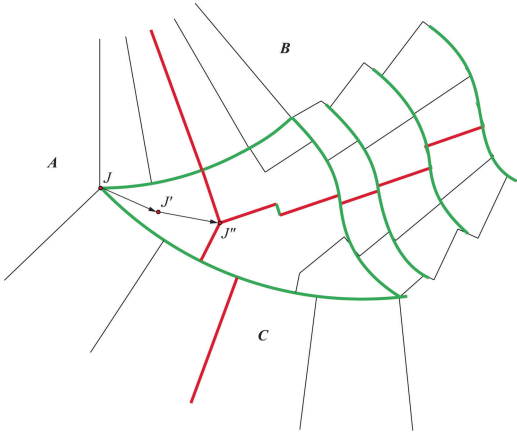
A similar formula allows to calculate the location of  $\mathbf{r}_0$  at time  $T$  on plate  $B$ ,  $\mathbf{r}_B(T)$ . This algorithm can be used to test the compatibility of existing kinematic models with real fracture zones trend. In fact, the chain of rotations included in Eq. (2.42) implies that even small errors on the single stage rotations are enhanced after few matrix multiplications. The algorithm, which should reproduce the geometry of any fracture zone, was applied for the first time by Shaw (1987) in a study on the South Atlantic plate motions. More recently, Schettino and Turco (2009) used this method to give further evidence that an independent Moroccan plate existed in the central Atlantic during the Oligocene and early Miocene. Equation 2.42 allows to predict the geometry of fracture zones given a plate motions model, that is, given a sequence of stage rotations. These lines must not be confused with the *flow lines* of relative motions, which display the path of a representative point on a plate  $B$  relative to a reference plate  $A$ . Figure 2.25 illustrates the process of constructions of flow lines and the difference with fracture zones. Although flow lines can be traced for both oceanic basins and zones of convergence, the latter tectonic context historically represents the principal field of application of this kind of kinematic representation (Dewey et al. 1989; Schettino and Turco 2011). Generally, the algorithm for generating flow lines is simple in the case of oceanic basins and follows

the approach used for the modelling of fracture zones (Eq. 2.42). The method is illustrated in Fig. 2.25. In this example, a point that is currently placed along the COB of one of the two plates, say  $B$ , is moved backward through time to the locations occupied at any time  $T \in [T_0, T_n]$  with respect to the conjugate reference plate  $A$ .

In this instance, however, the stage rotations are performed using the full stage angles  $\Omega_k$  and not, as we saw in the case of fracture zones, the halved angles. If  $\mathbf{r}_0$  is the position vector of a starting location along the continental margin of  $B$ , then the location at time  $T$ ,  $T_{k-1} \leq T \leq T_k$ , is a vector  $\mathbf{r}(T)$  given by:

$$\mathbf{r}(T) = \mathbf{S}_{BA}(T_k, T) \dots \mathbf{S}_{BA}(T_{n-1}, T_{n-2}) \times \mathbf{S}_{BA}(T_n, T_{n-1}) \mathbf{r}_0 \quad (2.43)$$

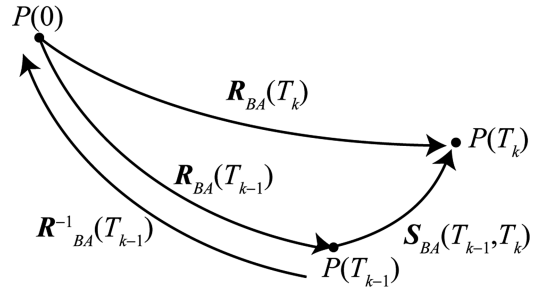
This method can be used for some, but *not* for *all*, pairs of oceanic plates. Furthermore, it is not generally applicable to the case of convergent settings. The reason is that stage rotations exist for some plate pairs sharing a common boundary, but not for *any* pair of plates, not even when they share a boundary. This theorem can be easily proved considering the simple case of a three-plates system. Let us consider an *RRR* triple junction like that of Fig. 2.17. If the relative motion of  $B$  with respect to  $A$  is described by a stage rotation, then by definition the Euler pole  $P_{BA}$  is at rest in the reference frame of  $A$ . Similarly, if the motion of  $C$  with respect to  $A$  can be described by a stage rotation, then the location of the stage pole  $P_{CA}$  is also at rest in the reference frame of  $A$ . However,  $P_{BA}$  will not be at rest with respect to  $C$ , just like  $P_{CA}$  will not be at rest with respect to  $B$ . Therefore, the Euler vector  $\boldsymbol{\omega}_{BC} = \boldsymbol{\omega}_{BA} + \boldsymbol{\omega}_{AC}$  cannot be an invariant neither in the frame of  $B$  nor in that of  $C$ . This implies that a stage pole does not exist for the plate pair  $(B, C)$ , even though both  $B$  and  $C$  move by stage rotations with respect to  $A$ . For this reason, we shall use the term “conjugate plates” only in the case of plates sharing a common boundary (not necessarily a spreading ridge) and whose relative motion can be described by a sequence of stage rotations. When the divergent relative motion between two



**Fig. 2.26** Fracture zones pattern and triple junction migration path for a system of three divergent plates. Plates *B* and *C* move about fixed Euler axes at constant rate with respect to *A*. The relative motion between *B* and *C* cannot be represented by a stage rotation

plates does not occur about a fixed rotation axis, the fracture zones assume the complicate shape shown in Fig. 2.26 between *B* and *C*.

Let us consider now the problem of determining stage rotations for a pair of conjugate plates *A* and *B*. In the case of oceanic basins, a subdivision of the opening history in tectonic stages is performed on the basis of the geometry of fracture zones and an analysis of marine magnetic anomalies. This procedure will be explained in detail in Chap. 5. It allows to determine both the stage boundaries ( $T_0, T_2, \dots, T_n$ ) and the relative position of *B* with respect to *A* (considered at rest in the present day position) at each time  $T_k$  ( $k = 1, 2, \dots, n$ ). Regarding the relative position at time  $T_0$ , it can be determined through a fitting algorithm applied to the conjugate COBs, as we shall see in the next section. All these relative positions are specified through *finite reconstruction* matrices  $R_{BA}(T_k)$ , whose expression is given by (2.18). The existence of these transformations is ensured by Euler's theorem (see Sect. 2.2). Generally, the components of the versor  $\mathbf{n}(T_k)$ , associated with the rotation axis at time  $T_k$ , are expressed in terms of geographic coordinates ( $\lambda(T_k), \phi(T_k)$ ) of a *finite reconstruction pole*  $P_k$ . In this instance, the triplet  $(\lambda(T_k), \phi(T_k), \Omega(T_k))$ ,  $\Omega(T_k)$  being the rotation angle, specifies all the necessary param-



**Fig. 2.27** Relationship between stage rotations and finite reconstructions. To move a point *P* from the location at time  $T_{k-1}$  to that at time  $T_k$ , it is possible to go first to the present day through an inverse finite reconstruction, then to time  $T_k$  through a direct finite transformation

eters to perform the transformation of a tectonic element from the present day location to the position at time  $T_k$  relative to the conjugate plate. Once the finite reconstruction matrices associated with each stage boundary have been determined, it is easy to calculate the corresponding stage rotations by the following formula, whose graphical proof is shown in Fig. 2.27:

$$S_{BA}(T_{k-1}, T_k) = R_{BA}(T_k) R_{BA}^{-1}(T_{k-1});$$

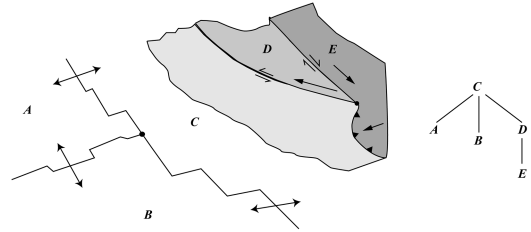
$$k = 1, 2, \dots, n \quad (2.44)$$

When considering finite reconstructions, it is always necessary to keep in mind that they do not represent real plate motions, but the combined result of many instantaneous or stage rotations. Therefore, the small circle arcs associated with a finite reconstruction pole are never expressions of existing geological structures, contrarily to the case of stage and instantaneous Euler poles. This is a key point for the correct interpretation of kinematic data, which has been discussed in depth in a seminal paper by John Dewey (1975).

Now we are ready to consider the process of construction of a *rotation model*, which represents the primary data structure that is used in plate tectonic modelling to store the relevant kinematic information. In fact, this file contains the information that is needed to reconstruct both the plate motions and the tectonic history of a region during a selected time interval. Rotation models are tables, generally stored in ASCII

files, used by dedicated algorithms during the process of construction of some hierarchical data structures (or *trees* in the sense of Computer Science, see Appendix 2) that describe the relative positions of a set of tectonic plates at any given time. A *rotation tree* can be considered as a data structure that specifies the multilevel tectonic hierarchy of a plate system at an assigned time  $T$  (Ross and Scotese 1988). It is often referred to as a *plate circuit*, although this term is also used when relative velocities are specified rather than relative positions, usually in studies on current plate motions. The nodes of these data structures are tectonic elements, while an edge between any pair of nodes indicates that their relative motion can be described by a sequence of stage rotations. Therefore, given a stage  $S$ , the edges of a plate circuit  $C$  for time  $T \in S$  define a set of *conjugate plate boundaries* in a system of interacting tectonic elements during the stage  $S$ , not the whole set of active plate boundaries, although all the existing plates at time  $T$  are represented in  $C$ . Thus, if  $e$  and  $p$  are respectively the size (that is, the number of edges) and the order (number of plates) of  $C$ , then by (2.34) it always results:  $e < 3(p-2)$ . It is also important to note that the topology of plate circuits is not constant through time, but changes as a consequence of major plate boundary reorganizations. In general, the definition of a plate circuit topology for each tectonic stage is based on the geological or geophysical evidence and the identification of a set of conjugate boundaries, such that the resulting graph is a connected tree (that is, for any two nodes  $u$  and  $v$  there exists a unique path from  $u$  to  $v$ ). The topology is specified implicitly during the compilation of a rotation model, while the duty of the reconstruction algorithms is to build a rotation tree for any assigned reconstruction time  $T$ .

In the example of Fig. 2.28, we assume that the relative motion between the plate pairs  $(A,C)$ ,  $(B,C)$ ,  $(C,D)$ , and  $(D,E)$  is represented by rotations at constant angular velocity during a time interval  $S \equiv [T', T'']$ . Therefore,  $S$  is assumed to be a tectonic stage. This assumption most likely relies on the geometry of fracture zones in the oceanic area and on geological field



**Fig. 2.28** Sketch map illustrating the construction of plate circuits. *Left*: A system of five plates. Finite rotations of  $A$  with respect to  $C$ ,  $B$  to  $C$ ,  $D$  to  $C$ , and  $E$  to  $D$  are known. *Right*: The corresponding plate circuit

evidence regarding the tectonic activity along the transcurrent faults that separate  $D$  from  $C$  and  $E$ . In this instance, four finite reconstruction matrices must be defined for the conjugate boundaries, which allow to determine four independent stage rotations,  $S_{AC}(T', T'')$ ,  $S_{BC}(T', T'')$ ,  $S_{DC}(T', T'')$ , and  $S_{ED}(T', T'')$  through Eq. (2.44). If  $R_{AC}(T)$ ,  $R_{BC}(T)$ ,  $R_{DC}(T)$ , and  $R_{ED}(T)$  are the finite reconstruction matrices at any time  $T \in [T', T'']$ , then any other relative position at time  $T$  can be calculated by combining these basic rotations.

For example, it is possible to determine the relative position of  $A$  with respect to  $B$  and that of  $E$  with respect to  $C$  at time  $T$ :

$$\begin{aligned} R_{AB}(T) &= R_{CB}(T)R_{AC}(T); R_{EC}(T) \\ &= R_{DC}(T)R_{ED}(T) \end{aligned}$$

In general, the tree structures associated with plate circuits are arranged so that the greater is the degree of a node in a plate circuit  $C$ , that is, the number of edges incident with the node, the higher will be its level in the hierarchical structure. Therefore, the neighborhoods of nodes in  $C$  will increase in size when we move toward higher levels in the data structure.

Plate circuits are built by reconstruction algorithms starting from a rotation model. This table specifies, for any stage boundary, the finite reconstruction pole and rotation angle of each identified pair of conjugate plates. A sample fragment of these data structures is shown in Fig. 2.29. In these tables, the tectonic elements

**Fig. 2.29** A fragment of rotation model

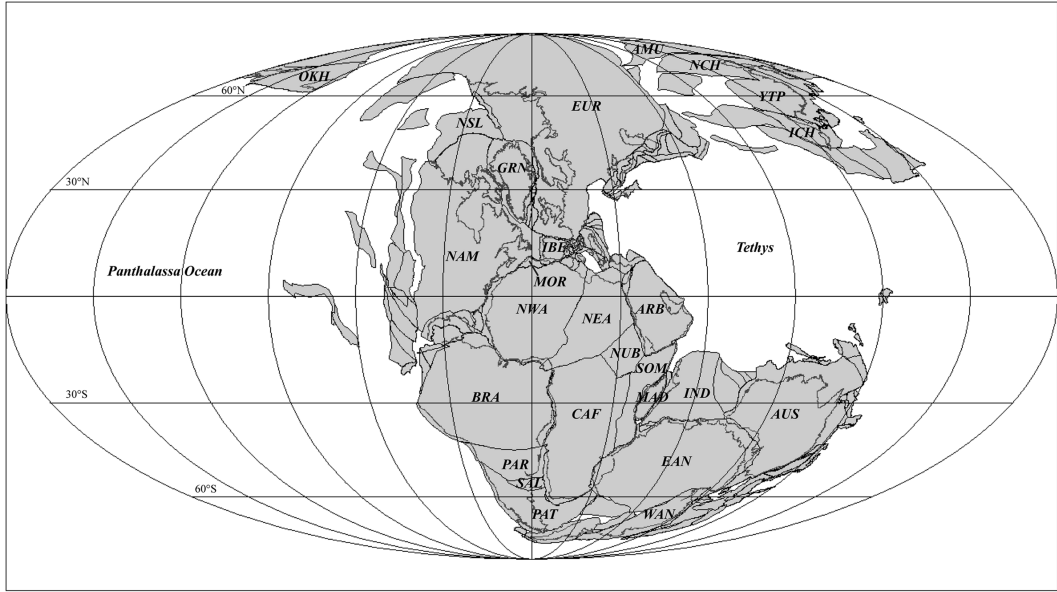
Plate Identifier	Age	Euler Pole Latitude	Euler Pole Longitude	Rotation Angle	Reference Plate	References
503	0.0	0.00	0.00	0.00	715	!Arabia/Northeast Africa
503	1.7	32.75	22.64	-1.89	715	!LePichon & Gaulier 1988
503	13.0	33.15	22.59	-5.46	715	!LePichon & Gaulier 1988
503	30.0	32.13	22.58	-6.36	715	!LePichon & Gaulier 1988
504	0.0	0.00	0.00	0.00	301	!Anatolia-Aegea/Eurasia
504	3.2	32.73	32.03	-5.50	301	!LePichon et al., 1995
		0.00	0.00	0.00	715	!Sinai-Northeast Africa
				0.00	715	!Le Pichon & Gaulier 1988
					715	!Le Pichon & Gaulier 1988

are coded through plate identifiers. In order to determine which of the two plates in a conjugate pair must be considered as the reference plate, we shall conform to the principle that high-degree nodes always appear at higher levels in the hierarchical structure. Regarding the definition of the stage boundaries, it is necessary to distinguish between the large first-order plates, in a context of global tectonics, and the case of small tectonic elements associated with intra-plate deformation or collisional settings. A key observation is that the changes of relative motion between large plates during the Mesozoic and the Cenozoic, hence presumably also during earlier time intervals, seem to have occurred synchronously on a global scale, thereby the major stage boundaries coincide. As an example, the classic plate motions model of Müller et al. (1997) is based on 15 synchronous stages from the early Jurassic to the present. This implies that the motions of the major tectonic plates cannot be determined *exclusively* by processes occurring in the mantle, including the subduction of slabs, and independently from each other. Therefore, at any time Earth's tectonic plates must be considered as a system of interacting bodies. Conversely, stage boundaries associated with changes of stage poles between small plates and other tectonic elements must be established on a geological basis and are not necessarily synchronous with major events of reorganization of the plate boundaries. In the next section, we shall discuss some important details of the procedures followed in plate kinematics for the construction of plate motions models.

## 2.8 Plate Tectonic Reconstructions

Usually plate motions models include a reconstruction of the initial configuration, preceding the development of plate boundaries. Figure 2.30 illustrates an example of fit of Pangaea, the large supercontinent that existed before the opening of the Atlantic ocean.

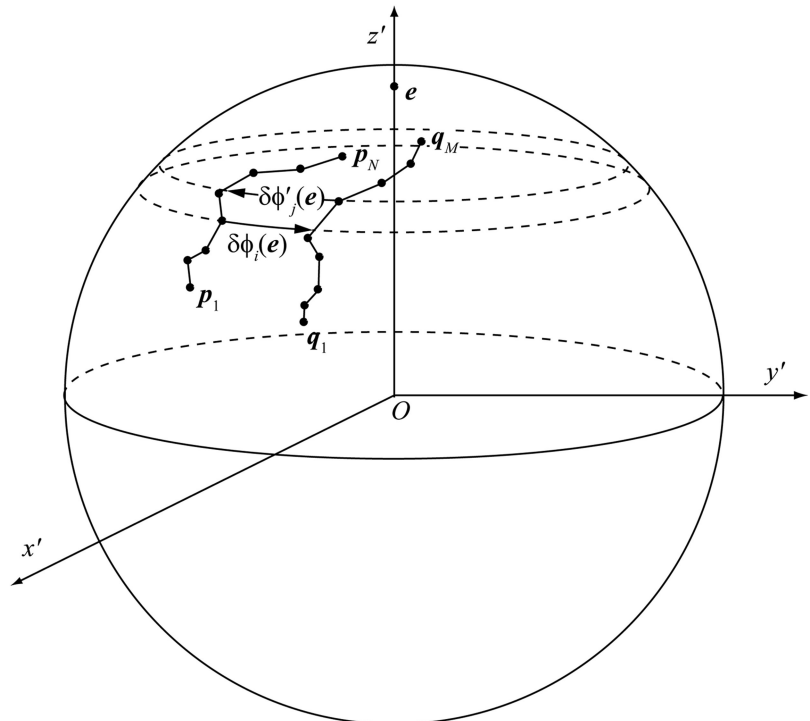
In the previous sections, we have learnt that there are two kinds of initial fits: pre-rift fits, which show the configuration of the continental masses preceding the development of plate boundaries, and post-rift fits, which match the stretched continental margins and show the configuration at the onset of sea floor spreading. In both cases, the match between the conjugate COBs is performed through a geometrical fitting procedure. The algorithm used by Bullard et al. (1965) in their reconstructions of Pangaea was the first rigorous method for fitting continental margins. Here we shall discuss an improved version of this algorithm, which was proposed by Schettino and Turco (2009). Let us assume that the COBs to be fitted are represented by two series of unit vectors, respectively  $(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N)$  and  $(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_M)$ , which have been preliminarily rotated to a geographic reference frame where a test Euler pole  $\mathbf{e}$ , with coordinates  $(\lambda_e, \phi_e)$ , has been moved to the North Pole (Fig. 2.31). A transformation of the standard geographic coordinates to this new reference frame is obtained by rotating each position vector  $\mathbf{p}_i$  and  $\mathbf{q}_j$  about an Equatorial pole placed at  $(0^\circ, \phi_e + 90^\circ)$  by



**Fig. 2.30** A pre-rift fit of Pangaea, based on Schettino and Scotese (2005) and Schettino and Turco (2009). The present day coastlines are shown for reference. The major tectonic elements in this reconstruction are: central Africa (CAF), northwest Africa (NWA), northeast Africa (NEA), Nubia (NUB), Somalia (SOM), Madagascar (MAD), Morocco (MOR), Arabia (ARB), India (IND), E. Antarctica

(EAN), Australia (AUS), W. Antarctica (WAN), Brazilian Craton (BRA), Paraná (PAR), Salado Block (SAL), Patagonia (PAT), N. America (NAM), Greenland (GRN), North Slope (NSL), Okhotsk (OKH), Eurasia (EUR), Iberia (IBE), Amurian Plate (AMU), N. China (NCH), Yang Tze Platform (YTP), and Indochina (ICH)

**Fig. 2.31** Geometrical fit of conjugate COBs. In a reference frame  $(x', y', z')$  where a test Euler pole,  $e$ , has been moved to the North Pole, the fitting procedure consists to find the best rotation about the North Pole, by an angle  $\Omega$  that minimizes the squared sum of longitude misfits  $\delta\phi_i$  ( $i \leq N$ ) and  $\delta\phi'_j$  ( $j \leq M$ )



an angle  $-\theta_e = -(\pi/2 - \lambda_e)$ . For each point  $\mathbf{p}_i$  on the first line, which can be carried onto the second line at position  $\mathbf{p}'_i$  by rotation about the North Pole, let  $\delta\phi_i(\mathbf{e})$  be the longitude difference between the two locations. Similarly, for each point  $\mathbf{q}_j$  on the second line, which can be carried back onto the first one at position  $\mathbf{q}'_j$  by rotation about the North Pole, let  $\delta\phi'_j(\mathbf{e})$  be the longitude difference. In general, only  $n \leq N$  points of the first line can be projected onto the second line, and only  $m \leq M$  points of the second line can be projected back onto the first line. If we rotate the western line by an angle  $\Omega$  about the North Pole, then misfit between a rotated vertex and its projection along the eastern line is given by  $\delta\phi_i(\mathbf{e}) - \phi_0$ . Similarly, if we rotate the eastern line by angle  $-\Omega$  about the North Pole, we obtain individual misfits  $\delta\phi'_j(\mathbf{e}) - \Omega$ . The total mean-square misfit will be given by:

$$\chi^2(\mathbf{e}) = \frac{N}{n^2} \sum_{i=1}^n (\delta\phi_i(\mathbf{e}) - \Omega)^2 + \frac{M}{m^2} \sum_{j=1}^m (\delta\phi'_j(\mathbf{e}) - \Omega)^2 \quad (2.45)$$

This formula shows some differences with respect to the one used by Bullard et al. (1965). In fact, the original formula of these authors assumed that the same number of points was projected between the two lines. This assumption is adequate only when the two COBs may match perfectly, that is, when each line can be fit against the whole conjugate line and not against a subset of the input data. For example, we could have missing information from one of the two conjugate COBs. In this instance, we must search for a best fit of one line against a *subset* of the second line, not necessarily a whole geometrical fit. Eq. (2.45) takes into account of the possibility that one the two lines is not complete. In these conditions, the best fit Euler pole searching algorithm also tries to maximize the percentage of matched segments from each line, that is, the number of projected points, because we could find wrong Euler poles that furnish very good fits of small segments of the two lines. This problem is solved in Eq. (2.45) by multiplying the squared misfit

of each line respectively by  $N/n$  and  $M/m$ . This expression reaches a minimum when the rotation angle  $\Omega = \Omega(\mathbf{e})$  is given by:

$$\Omega(\mathbf{e}) = \frac{\frac{N}{n^2} \sum_{i=1}^n \delta\phi_i(\mathbf{e}) + \frac{M}{m^2} \sum_{j=1}^m \delta\phi'_j(\mathbf{e})}{\frac{N}{n} + \frac{M}{m}} \quad (2.46)$$

The fitting procedure is a searching algorithm of the Euler pole  $\mathbf{e}$  which minimizes the misfit  $\chi^2$  in expression (2.45). The search is generally based on trial Euler poles that are chosen over a coarse grid of locations (for example, a  $1 \times 1^\circ$  global grid). For each trial pole  $\mathbf{e}$ , the angle  $\Omega$  that minimizes  $\chi^2$  is determined through Eq. (2.46). A first approximate location of the Euler pole is obtained by selecting the trial pole that gives the minimum value of  $\Omega$  over the global grid. Now a new scan is performed over a neighbor of this point using a reduced grid spacing, for example  $0.1^\circ$ , so that a new more precise location of the Euler pole and a new angle of rotation are determined. The algorithm stops when the desired resolution is reached.

Now let us consider the procedure for reconstructing the position of a tectonic element at time  $T$  in the geologic past, starting from a corresponding plate circuit  $C(T)$ . This reconstruction algorithm has the following simple structure:

#### Algorithm 2.2 (Reconstruction Algorithm)

Input: a node  $n \in C(T)$ ;

Output: A total reconstruction matrix  $\mathbf{R}_n(T)$ ;

```
{
  1.  $\mathbf{R}_n(T) \leftarrow \mathbf{I}$ ;  $c \leftarrow n$ ;
  2.  $p \leftarrow \text{Parent}(c)$ ;
  3.  $p = 0 \Rightarrow$  jump #7;
  4.  $\mathbf{R}_n(T) \leftarrow \mathbf{R}_{cp}(T)\mathbf{R}_n(T)$ ;
  5.  $c \leftarrow p$ ;
  6. Jump #2;
  7.  $\mathbf{R}_n(T) \leftarrow \mathbf{R}_c(T)\mathbf{R}_n(T)$ ;
}
```

A *total reconstruction matrix*,  $\mathbf{R}_n(T)$ , is a matrix that moves a tectonic element  $n$  from its present day location, in the geographic reference

frame, to the position that this block had at time  $T$  in a *paleotectonic reference frame*. An important example of this class of frames is represented by the paleomagnetic frames mentioned in Sect. 2.3, but one could wish to refer the reconstructions to a set of hot spots (e.g., Müller et al. 1993) or even use a present day continent, for example N. America or Eurasia, as a reference frame for plate reconstructions. The existence of total reconstruction matrices is again a consequence of Euler's theorem. The corresponding Euler pole is called a *total reconstruction pole*. At step #1 of Algorithm 2.1, the total reconstruction matrix is initialized by the identity matrix  $\mathbf{I}$ , and the current node,  $c$ , is set to be the starting node. At step #2, a variable  $p$  is assigned the parent of the current node in the tree structure. At the next step, if the current node  $c$  coincides with the root of the tree, so that  $p = 0$ , then the iteration stops and the final reconstruction matrix is updated by adding the transformation of the root node with respect to the paleotectonic reference frame,  $\mathbf{R}_c$ . At step #4, the current rotation matrix is updated by adding the relative rotation of the current node with respect to its parent. Then, at the next step, we move upwards to the next higher level by assigning the current node its parent and the sequence is restarted. On exit, this algorithm furnishes the total reconstruction matrix of  $n$  at time  $T$  in the variable  $\mathbf{R}_n(T)$ .

As an example, the application of this algorithm to the circuit of Fig. 2.28 would give the following total reconstruction matrices:

$$\begin{cases} \mathbf{R}_A(T) = \mathbf{R}_C(T)\mathbf{R}_{AC}(T) \\ \mathbf{R}_B(T) = \mathbf{R}_C(T)\mathbf{R}_{BC}(T) \\ \mathbf{R}_D(T) = \mathbf{R}_C(T)\mathbf{R}_{DC}(T) \\ \mathbf{R}_E(T) = \mathbf{R}_C(T)\mathbf{R}_{DC}(T)\mathbf{R}_{ED}(T) \end{cases}$$

To calculate the set of finite reconstruction matrices  $\mathbf{R}_{ij}(T)$  associated with a plate circuit at a given intermediate time  $T$ , algorithm 2.2 uses the components of these transformation matrices at stage boundaries. If  $T_{k-1} \leq T \leq T_k$ , then the corresponding finite reconstruction of plate  $i$  relative to plate  $j$  is given by:

$$\mathbf{R}_{ij}(T) = \mathbf{S}_{ij}(T_{k-1}, T) \mathbf{R}_{ij}(T_{k-1}) \quad (2.47)$$

where the rotation  $\mathbf{S}_{ij}$  is calculated using the reduced angle (Eq. 2.41). Now we can address the problem of complementing the kinematic representation of a set of tectonic plates through velocity and acceleration fields. We know that the linear velocity  $\mathbf{v}(\mathbf{r})$  at the location represented by a position vector  $\mathbf{r}$  can be calculated easily starting from an Euler vector  $\boldsymbol{\omega}$  (Eq. 2.17). Therefore, the problem of representing velocity fields can be reduced to the problem of determining the instantaneous axis of relative rotation between two plates sharing a boundary at time  $T$ , independently from the eventuality that these are conjugate plates or not. Furthermore, it is occasionally necessary to determine *absolute* velocity fields in the selected paleotectonic reference frame. Clearly, in the case of relative velocity fields between conjugate plates the calculation should be simplified by the fact that the relative motions are rotations about fixed axes at constant angular velocities. However, even in this eventuality it is necessary to take into account that the rotation axis of a stage pole is fixed with respect to a plate that is considered at rest in the present day geographic frame. Therefore, the axis must be rotated according to the total reconstruction matrix of this plate at time  $T$  before it can be used for calculating velocity vectors. Let  $\mathbf{n}_{ij}(0)$  be the unit vector of the rotation axis associated with a stage rotation  $\mathbf{S}_{ij}(T_{k-1}, T_k)$ . If  $T_{k-1} \leq T \leq T_k$ , and  $\mathbf{R}_j(T)$  is the total reconstruction matrix of the reference plate at time  $T$ , then the orientation of this axis at time  $T$  will be given by:

$$\mathbf{n}_{ij}(T) = \mathbf{R}_j(T)\mathbf{n}_{ij}(0) \quad (2.48)$$

At this point, to form a complete Euler vector we still need to assign an angular velocity  $\omega$  at time  $T$ . This task can be easily accomplished, because during a stage the relative angular velocity between two plates is assumed to be approximately constant, thereby we can always determine this quantity starting from the stage angle  $\Omega_k$  and the temporal boundaries  $T_{k-1}$  and  $T_k$ . It results:

$$\omega = \frac{\Omega_k}{T_k - T_{k-1}} \quad (2.49)$$

In the general case of two plates whose relative motion occurs about a continuously changing Euler pole, it is not possible to determine the instantaneous rotation axis and angular velocity starting from a stage rotation. Therefore, a more general, although complicate, approach to the problem becomes necessary. We know that all the plates are represented in a plate circuit at time  $T$ . For each node  $i$ , a finite reconstruction matrix  $\mathbf{R}_{ij}(T)$  exists, which allows to calculate the position of the corresponding plate with respect to the parent plate  $j$ . We also know that the edges of a plate circuit attest the existence of specific kinematic relations between plate pairs, consisting in the fact that the relative motions are stage rotations, thereby they link pairs of *conjugate plates*. Therefore, for *any* plate  $i$  not coinciding with the root of the tree, it is possible to determine a stage rotation matrix  $\mathbf{S}_{ij}$  and the relative velocity field  $\mathbf{v}_{ij}$  with respect to the parent plate  $j$  in the hierarchical representation using the following simple expression, which combines Eq. (2.17) with Eqs. (2.48) and (2.49):

$$\mathbf{v}_{ij}(\mathbf{r}, T) = \frac{\mathbf{R}_j(T) \mathbf{n}_{ij}(0) \Omega_k}{T_k - T_{k-1}} \quad (2.50)$$

for any  $T_{k-1} \leq T \leq T_k$ . In the case of the root continent,  $r$ , the corresponding finite reconstruction matrix coincides with a total reconstruction matrix,  $\mathbf{R}_r(T)$ , that represents the transformation of  $r$  with respect to the paleotectonic reference frame. Therefore, if  $(i, j, k, \dots, r)$  is a path in the tree structure from node  $i$  to the root, then we can always determine the *absolute velocity field* of a plate in the selected reference frame by composition of velocity vectors:

$$\begin{aligned} \mathbf{v}_i(\mathbf{r}, T) &= \mathbf{v}_{ij}(\mathbf{r}, T) + \mathbf{v}_{jk}(\mathbf{r}, T) \\ &+ \dots + \mathbf{v}_r(\mathbf{r}, T) \end{aligned} \quad (2.51)$$

where the absolute velocity field of the root node,  $\mathbf{v}_r$ , is determined calculating stage rotations with respect to the paleotectonic reference frame.

Now we are ready to solve the problem of determining relative velocity fields between non-conjugate plate pairs. In fact, if  $i$  and  $j$  are *any*

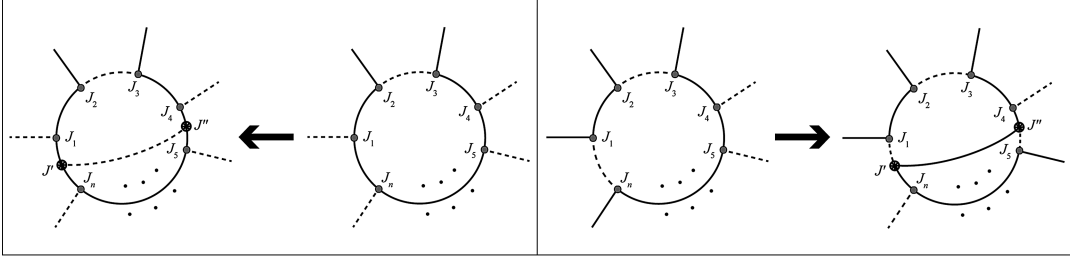
two plates, it is always possible to calculate their relative velocity by subtracting their absolute velocities:

$$\mathbf{v}_{ij}(\mathbf{r}, T) = \mathbf{v}_i(\mathbf{r}, T) - \mathbf{v}_j(\mathbf{r}, T) \quad (2.52)$$

Another important kinematic variable is the relative or absolute acceleration of a tectonic plate. Curiously, this kind of vector fields have been seldom considered in plate kinematics studies, in spite of their importance for the geodynamic assessment of the models. Probably this is a consequence of the fact that accelerations have been traditionally considered as point events that only occur at stage boundaries, consistently with the current description of the driving mechanism of plate tectonics. However, recent research has shown that phases of accelerated motion have existed in the geologic past, possibly associated with the action of mantle plumes (Cande and Stegman 2011). We shall consider in detail the geodynamics of accelerated states in Chap. 13. In order to obtain an acceleration field, it is necessary to consider two successive times,  $T$  and  $T'$ , close enough, and calculate a velocity field for each of them. For example, Schettino and Scotese (2002) used a time interval of 1 Myr for determining the acceleration across stage boundaries in the Mediterranean region during the Mesozoic. The acceleration is calculated simply by dividing the velocity variation by the size of the time interval:

$$\mathbf{a}_{ij}(\mathbf{r}, T) = \frac{\mathbf{v}_{ij}(\mathbf{r}, T') - \mathbf{v}_{ij}(\mathbf{r}, T)}{T' - T} \quad (2.53)$$

It is possible that in the previous discussion, about the velocity field between non-conjugate plate pairs, some readers wondered about the effective number of situations characterized by this kind of relative motion. We shall satisfy the curiosity of these readers by proving an interesting topological theorem, which will help to clarify some key features of plate tectonic configurations. We say that a plate boundary is a *conjugate boundary* if it separates a pair of conjugate plates. In this instance, any geological structure associated with strike-slip motion, for example a transform fault, will be aligned with



**Fig. 2.32** Nucleation of a new plate by splitting of an existing  $n$ -th order plate. Free plate boundaries are shown as dashed lines. *Left*: The new plate boundary splits two conjugate boundaries through the insertion of two new triple junctions,  $J'$  and  $J''$ . In this case, the new boundary is a free boundary and there is no need for a global

reorganization. *Right*: If one of the two boundaries that are cut is a free boundary, then a large scale reorganization, involving several conversions between free and conjugate boundaries, is necessary. In this instance, a new conjugate boundary separates the parent plate from the newly formed tectonic element

a small circle arc about a stage pole. When a boundary separates two plates whose relative motion occurs about a continuously changing Euler pole, we say that this is a *free boundary*. In this case, strike-slip faults, in particular transform faults, and fracture zones have a quite complicate pattern, as illustrated in Fig. 2.26. If  $C$  is a plate circuit containing  $p$  nodes, then its size (i.e., the number of edges) is given by  $e = p - 1$ . The following theorem proves that this number does *not* coincide with the total number of conjugate pairs in a plate tectonic configuration, that is, with the total number of conjugate boundaries, but is always lower.

#### Topological Theorem (for Plate Tectonic Configurations)

If  $G(j, b)$  is a global plate configuration, then the number of free and conjugate boundaries are given, respectively, by:

$$f = \frac{1}{3}b = e - 1 = p - 2 \quad (2.54)$$

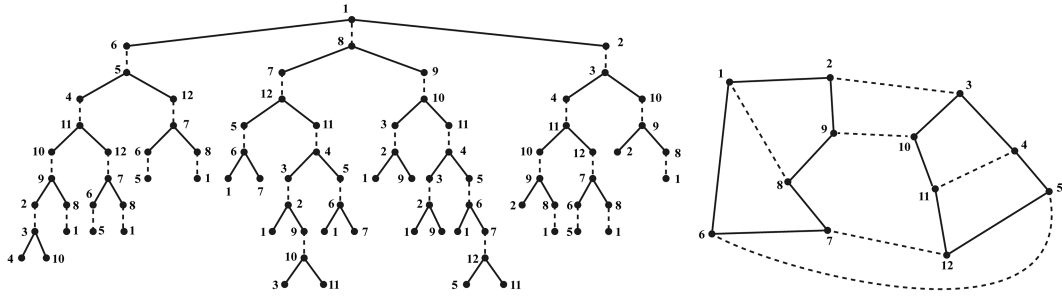
$$c = \frac{2}{3}b = j = 2f \quad (2.55)$$

*Proof* In this proof, we always assume that in *normal conditions* a system of tectonic plates tries to maximize the number of conjugate boundaries during any episode of reorganization, because this is clearly a minimum energy configuration. In a three-plates system, it results by (2.34) that  $b = 3$  and  $j = 2$ . We have already proved

that in this configuration only two of the three boundaries can be conjugate boundaries, thereby we would have  $f = 1$  and  $c = 2$ , in agreement with Eqs. (2.54) and (2.55). Now let us assume that the theorem holds for a system with  $p$  plates. We want to prove that in this case it also holds if one of these plates splits, thus adding a new tectonic element to the system.

Figure 2.32 shows two possible mechanisms for generating a new additional tectonic plate from an existing one. Clearly, in order to create a new boundary that splits an existing plate, two of its boundaries must be broken by insertion of triple junctions. If the edges that are split are conjugate boundaries, two extra conjugate boundaries and one additional free boundary are created and there is no need to change the tectonic style of the remaining plate boundaries. In this instance,  $f$  increases by one, while  $c$  increases by two, thereby Eqs. (2.54) and (2.55) remain valid and the theorem is proved. The new plate boundary separating the two parts of the original plate is always a free boundary when this kind of plate nucleation occurs. A much more complicated situation follows if at least one of the two boundaries that are split is a free boundary. In this instance, the proof relies on the fact that for any pair of triple junctions in  $G$ , there exist at least three *alternate paths* that link the two nodes. An alternate path is a path formed by an alternate sequence of conjugate and free edges.

The example of Fig. 2.33 shows the tree structure that can be formed with the set of all alternate

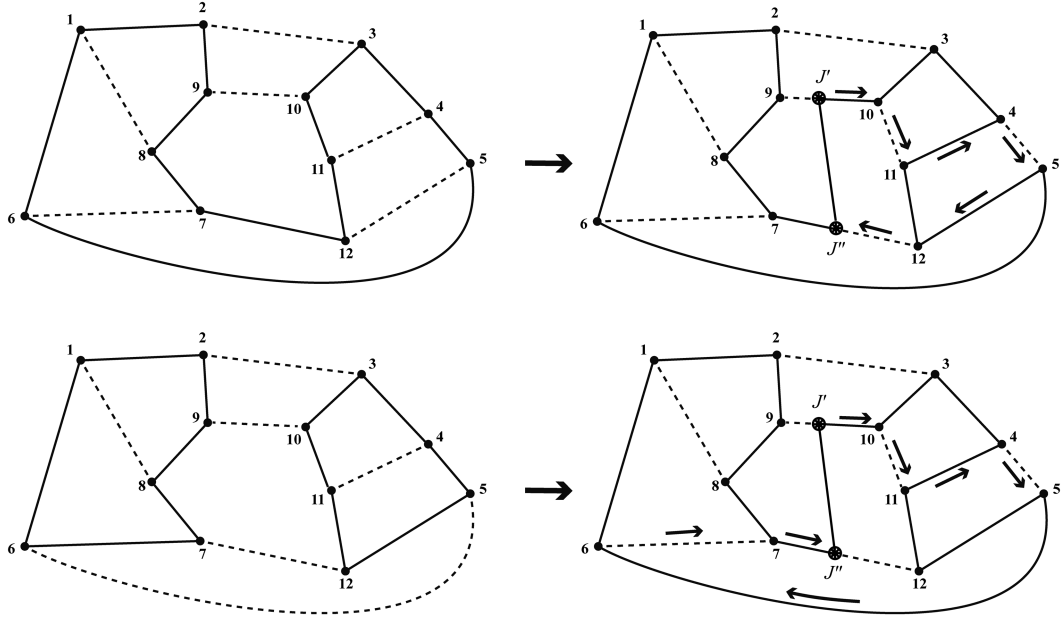


**Fig. 2.33** An example of global plate configuration, and the corresponding tree of alternate paths starting from node  $J_1$ . Dashed and solid lines are respectively free and conjugate plate boundaries

paths starting from a node. The tree is built setting the root node as the start node,  $s$ , and generating all the possible alternate paths originating in  $s$  by a modified depth-first search or breadth-first search algorithm (see Appendix 2). For example, for the graph of Fig. 2.33 we could start from node  $J_1$ , then reach nodes  $J_2$ ,  $J_6$ , and  $J_8$ . From node  $J_2$ , we must proceed to node  $J_3$ , while from  $J_8$  we can proceed either to node  $J_7$ , or node  $J_9$ , etc. When one of the edges that are split in the process of nucleation of a new tectonic plate is a free boundary, the new intermediate boundary separating the original plate cannot be a free boundary. More precisely, it *could* be a free boundary, but this would not be the minimum energy solution. In fact, Fig. 2.32 shows that after insertion of the new triple junctions  $J'$  and  $J''$ , only one of the two arms of a free boundary needs to be converted to a conjugate boundary after splitting (the boundary between  $J'$  and  $J_n$  in the example of Fig. 2.32), whereas any other solution would imply a local reorganization involving more conversions. However, even in this case the conversion of part of the former free boundary induces a series of concatenate transitions from free to conjugate boundaries and vice versa that may propagate outside the perimeter of the splitting plate, determining a total rearrangement of the plate configuration. For each boundary conversion, another boundary incident with the same triple junction must be subject to the inverse of that transformation. At the same time, a stage boundary is created, which could either close an existing stage or start a new one after a time interval of complex relative motion.

The simplest way to terminate the sequence of boundary conversions and establish a new equilibrium is to reach the opposite triple junction,  $J''$ , along the shortest alternate path. This path will include an even number of edges when only one of the splitting boundaries is a free boundary, whereas an odd number of edges are converted when two free boundaries are split. The examples of Fig. 2.34 illustrate these two possibilities.

It should be noted that the new boundary linking  $J'$  and  $J''$ , and the alternate path linking these nodes form a closed loop in  $G$ , whose outgoing edges are unaffected by the boundary conversions. In the case of an even number of converted edges, for each conjugate boundary that is transformed into a free boundary there is a free boundary that is converted into a conjugate boundary. Therefore, after the creation of a new free boundary and two new conjugate boundaries the total number of edges does not change. This implies that also in this case  $f$  and  $c$  increase respectively by one and by two units, thereby Eqs. (2.54) and (2.55) remain valid and the theorem is proved. When the number of converted boundaries is odd, after the creation of two new free boundaries (see Fig. 2.34) and one additional conjugate boundary the sequence of conversions allows an extra conversion of the last free boundary into a conjugate edge, thereby during the conversions  $f$  decreases by one and  $c$  increases by one. Therefore, also in this case  $f$  and  $c$  increase respectively by one and by two units, so that Eqs. (2.54) and (2.55) conserve their validity after the formation of the new plate. This proves the topological theorem. ■



**Fig. 2.34** Alternate paths of converted boundaries during the splitting of a tectonic plate. *Top*: The new edge links a free boundary to a conjugate boundary. *Bottom*: Two free boundaries are split

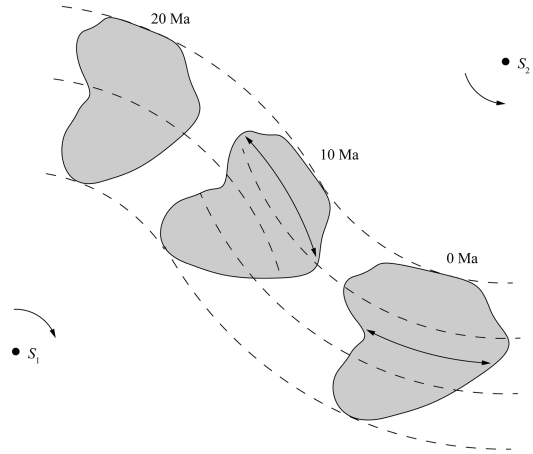
In Sect. 2.7 we have described a method of construction of flow lines of relative motion, which can be used only when the two plates form a conjugate pair. Furthermore, we have mentioned that it is not generally applicable to the case of convergent boundaries. Now we want to describe a technique to produce flow lines in the general case of two plates separated by a free boundary for at least some time intervals. The method is based on the general technique for determining relative velocity fields discussed above. Let  $\mathbf{r}_0$  be the present day position vector of a point belonging to a plate  $B$ . We want to generate a path on the globe, formed by the relative positions of this point with respect to another plate  $A$  at any time  $T$  in the geologic past. The task is easily accomplished calculating the relative velocity field  $\mathbf{v}_{BA}(T)$  for a sequence of closed times  $T', T' + \delta T, T' + 2\delta T, \dots$  in an interval  $[T', T'']$ . In fact, starting from the position at the initial time, which is  $\mathbf{r}(T') = \mathbf{R}_{BA}(T')\mathbf{r}_0$ , we can calculate iteratively the position at any successive time by the following formula:

$$\mathbf{r}(T + \delta T) = \mathbf{r}(T) + \mathbf{v}_{BA}(T)\delta T \quad (2.56)$$

The kinematic methods described so far can be applied equally well to the reconstruction of the tectonic evolution of oceanic basins and to continental tectonics. In the former case, the finite reconstruction matrices associated with the rotation model are determined on the basis of an analysis of fracture zones and the pattern of marine magnetic anomalies, as it will be explained in Chap. 5. In the case of reconstructions that involve deformation of continental crust in the geologic past, for example during rifting or collisions, or when the finite rotations must describe relative motions between continental blocks, these techniques are inapplicable. In this context, the determination of the reconstruction parameters (latitude and longitude of the Euler poles and rotation angles) mostly relies on geologic data that lack of information on measurement uncertainty, thereby it is not generally possible to determine confidence ellipses for the resulting reconstruction parameters. Nevertheless, it is possible to generate semi-quantitative reconstructions that correctly predict the style and timing of tectonic deformation, thus giving a theoretical framework to the geological observation at regional or local scale.

The asymmetric deformation of a tectonic element of continental crust during its motion can be described by modifying the format of the tables associated with rotation models (Fig. 2.29). Such a modification should consist into the introduction of an additional field, which would allow to establish the amount of shortening or extension, with respect to the present day shape, that a block experienced during its motion. This method cannot be used to describe a symmetric extension between two plates during a rifting phase, because the additional attribute refers to a single tectonic element which will be rotated with respect to a parent plate, not to a plate pair. For example, the method can be used in the modelling of back arc basins, or in the description of complex systems, such as the assemblage of microplates, slivers, and orogenic structures that characterizes the Alpine-Himalayan belt. In general, the additional field would describe the deformation of a continental block during a stage rotation through a stretching (or shortening) factor. For example, a value of 0.9 at 10 Ma would mean that in a reconstruction at 10 Ma we must shorten the tectonic element by 10 % with respect to the present size, in the direction determined by the stage matrix between 10 Ma and the present. Similarly, if the deformation parameter is 0.8 at 20 Ma, this value implies two different phases of extension, the first one between 20 and 10 Ma and a second one between 10 Ma and the present. Of course, these two phases of deformation could develop along different directions, determined by diverse stage poles, so that in general the present day shape of a tectonic element would result from the superposition of several tectonic events. An example illustrating the combination of rotations and phases of extension is shown in Fig. 2.35.

Describing the kinematics of deformable bodies requires special techniques and an additional computational effort for generating plate reconstructions of the tectonic evolution of a region. For example, it is not generally possible to determine the initial shape and location of a tectonic element through a single finite rotation, accompanied by deformation of the block, because usually the present day geometry results from the superposition of different phases of extension



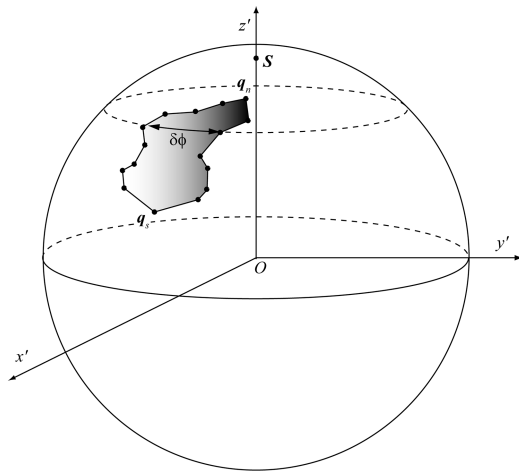
**Fig. 2.35** Deformation of a continental block during its motion. In this example, a tectonic element rotates clockwise about the stage pole  $S_1$  between 20 and 10 Ma, and counterclockwise about  $S_2$  between 10 Ma and the present. During the first stage, it is stretched in the same directions of the flow lines about  $S_1$ . Then, a second phase of deformation is superimposed on the first one, and the block is stretched again, this time in the direction of the flow lines about  $S_2$

or shortening, which develop along distinct axes of deformation. Now we are going to describe, in a simplified form, the procedure of stretching or shortening of a tectonic element along an assigned direction, which cannot be defined as a trivial algorithm.

The procedure of block deformation starts with its rotation to a reference frame where the stage pole coincides with the North Pole (Fig. 2.36). Then, the points having minimum and maximum latitudes, in the set of  $N$  vertices associated with the spherical polygon representation, are determined. Let  $q_s$  and  $q_n$  be respectively these two points (Fig. 2.36), and assume that the sequence of vertices ( $q_1, q_2, \dots, q_N$ ) is a clockwise sequence. The points  $q_s$  and  $q_n$  can be used to divide the perimeter of the tectonic element in two halves: an eastern half that includes vertices from  $q_s$  to  $q_n$ , and a western sequence, which includes vertices from  $q_n$  to  $q_s$ . For each point in the western sequence,  $q_i$ , let  $\delta\phi_i$  be the longitudinal distance from the western boundary. If  $\beta$  is the deformation factor, then  $q_i$  is moved along its parallel to a new location,  $q'_i$ , in such a way that  $\delta\phi'_i = \beta\delta\phi_i$ . Finally, the resulting

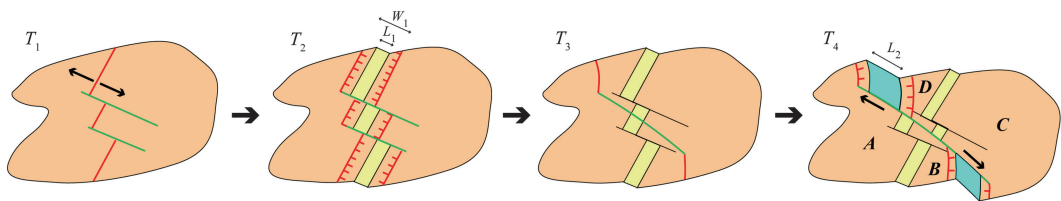
polygon is moved back to the original reference frame before applying the rigid transformation listed in the rotation model.

Another problem that often must be solved, in the context of continental tectonics, is associated with the impossibility to determine a priori, independently from the specific kinematic history, a list of finite reconstruction poles to be included in a rotation model. The reason is that the typical



**Fig. 2.36** Stretching and shortening continental blocks by a factor  $\beta$ . In a reference frame  $(x', y', z')$  where the stage pole,  $S$ , has been moved to the North Pole, the northernmost and southernmost vertices,  $q_n$  and  $q_s$ , of a plate polygon are used to divide the block perimeter in two halves. Then, each point along the eastern half is moved along its parallel of latitude to stretch or shorten by factor  $\beta$  the corresponding *small circle* arc  $\delta\phi$  that separates it from the western boundary

geological data are generally represented by partially incoherent geologic structures (faults, fold axes, etc.), which result from the superposition of two or more phases of deformation, as illustrated in Fig. 2.37. Even assuming that it is possible to separate the original data in coherent subsets, and to identify the timing of the deformation phases through a precise dating of the geologic structures, the oldest tectonic structures cannot be used to determine finite reconstruction parameters, because it is likely that their strike has been affected by the more recent phases of deformation. In these conditions, the typical approach is to reconstruct the tectonic history of a region starting from the most recent phase of deformation and going back through time. If we can identify the most recent set of geologic structures, for example between some time  $T_k$  and the present (Fig. 2.37), then it is possible to determine the parameters of a stage transformation  $S(0, T_k)$ , which clearly coincides with the finite reconstruction matrix at time  $T_k$ :  $S(0, T_k) = R(T_k)$ . At this point, all the structures that are older than  $T_k$ , and that have been affected by the most recent phase of deformation, are rotated using the matrix  $R(T_k)$ , in order to remove the “overprint” of this phase. After this operation, these structures become coherent with other data that had not been affected by the recent deformation. The resulting data set can be used, at the next step, to determine a second stage pole,  $S(T_k, T_{k-1})$ , which in turn allows to calculate the finite reconstruction matrix at time  $T_{k-1}$ :  $R(T_{k-1}) = S(T_k, T_{k-1})R(T_k)$ , and



**Fig. 2.37** Superposition of two phases of deformation of a tectonic element. During phase 1, between  $T_1$  and  $T_2$ , a rift forms with extension axes having direction WNW–ESE. Note that the resulting offset  $L$  between the two separating blocks is always less than the width  $W$  of the stretched zone. This phase is followed by a second episode of extension between  $T_3$  and  $T_4$ , having NW–SE direction, which modifies the strike of some of

the original transfer zones and rift axes. At the end of this phase the original block has been divided in four distinct tectonic elements (A, B, C, and D). To determine a finite reconstruction pole and angle of rotation for phase 1, it is necessary to remove the effects of the second phase of deformation, by reconstructing the shape of the tectonic element at time  $T_3$

**Table 2.2** The NUVEL-1A velocity model

Plate	$\omega_x$	$\omega_y$	$\omega_z$	$\omega$	$\lambda_e$	$\phi_e$
Africa	0.002401	-0.007939	0.013891	0.9270	59.160	-73.174
Antarctica	0.000689	-0.006541	0.013676	0.8695	64.315	-83.984
Arabia	0.008195	-0.005361	0.016730	1.1107	59.658	-33.193
Australia	0.009349	0.000284	0.016253	1.0744	60.080	+1.742
Caribbean	0.001332	-0.008225	0.011550	0.8160	54.195	-80.802
Cocos	-0.008915	-0.026445	0.020895	1.9975	36.823	-108.629
Eurasia	0.000529	-0.007235	0.013123	0.8591	61.066	-85.819
India	0.008181	-0.004800	0.016760	1.1034	60.494	-30.403
North America	0.001768	-0.008439	0.009817	0.7486	48.709	-78.167
Nazca	-0.000022	-0.013417	0.019579	1.3599	55.578	-90.096
South America	0.000472	-0.006355	0.009100	0.6365	54.999	-85.752

$\omega = (\omega_x^2 + \omega_y^2 + \omega_z^2)^{1/2}$  is the angular velocity in deg/Myr;

$\omega_x$ ,  $\omega_y$ , and  $\omega_z$  are expressed in rad/Myr;

$\lambda_e$  and  $\phi_e$  are, respectively, the latitude and longitude of the Euler pole with respect to the Pacific

so on. Thus, in general, the finite reconstruction poles associated with the kinematics of a set of continental plates can be calculated only *after* the stage transformations have been determined by concatenation of stage matrices.

## 2.9 Current Plate Motions

We are going to conclude this chapter, dedicated to plate kinematics, with a description of the techniques used for the determination of the modern plate motions. The first models of current plate kinematics were based on a combination of heterogeneous data, represented by seismic slip vectors, averaged spreading rates, and transform fault azimuths (Chase 1978; Minster and Jordan 1978; DeMets et al. 1990). Each of these models specified a set of  $n-1$  Euler vectors,  $\omega_i$ ,  $n$  being the number of modern plates, relative to a reference plate, for example the Pacific plate. The models were obtained through least squares procedures that minimized the quantity:

$$\chi^2 = \sum_{i=1}^{n-1} \sum_{j=1}^{n_i} \sum_{k=1}^{N_{ij}} \left\{ \frac{[\mathbf{v}(\mathbf{r}_k) - (\omega_i - \omega_j) \times \mathbf{r}_k] \cdot \mathbf{n}_k}{\sigma_k} \right\}^2 + \sum_{i=1}^{n-1} \sum_{j=1}^{m_i} \sum_{k=1}^{M_{ij}} \frac{1}{\sigma_k^2} \left[ \mathbf{s}(\mathbf{r}_k) - \frac{(\omega_i - \omega_j) \times \mathbf{r}_k}{|(\omega_i - \omega_j) \times \mathbf{r}_k|} \right]^2 \quad (2.57)$$

This formula allows to calculate the squared sum of misfits between predicted and observed spreading rates and azimuths of relative velocities. Each plate is assumed to have  $n_i$  boundaries that are spreading ridges, and for each of these boundaries, there are  $N_{ij}$  spreading rate data at locations represented by position vectors  $\mathbf{r}_k$ . Let  $\omega_i - \omega_j$  be the predicted relative angular velocity of the  $i$ -th plate with respect to an adjacent plate separated by a spreading ridge. By (2.17), the predicted linear velocity between the two plates at a location  $\mathbf{r}_k$  is given by:  $(\omega_i - \omega_j) \times \mathbf{r}_k$ . If  $\mathbf{n}_k$  and  $\mathbf{v}(\mathbf{r}_k)$  are respectively a versor normal to the ridge axis and the observed average spreading velocity at  $\mathbf{r}_k$ , then the weighted misfit between observed and predicted spreading rates is given by the scalar difference between the projections of  $\mathbf{v}(\mathbf{r}_k)$  and  $(\omega_i - \omega_j) \times \mathbf{r}_k$  onto the axis of  $\mathbf{n}_k$ , divided by the standard error  $\sigma_k$  attributed to  $\mathbf{v}(\mathbf{r}_k)$ . Similarly, it is assumed that the  $i$ -th plate has  $m_i$  generic boundaries, each having  $M_{ij}$  directional observations. Let  $\mathbf{s}(\mathbf{r}_k)$  be a unit versor representing one of these observations. The predicted direction is clearly given by the versor of the theoretical linear velocity  $(\omega_i - \omega_j) \times \mathbf{r}_k$ . Therefore, the weighted misfits of azimuth data can be defined as the magnitudes of the vector differences between predicted and observed direction versors, divided by the standard error  $\sigma_k$ .

Table 2.2 lists the Euler vectors of NUVEL-1A (DeMets et al. 1994), one of the most widely

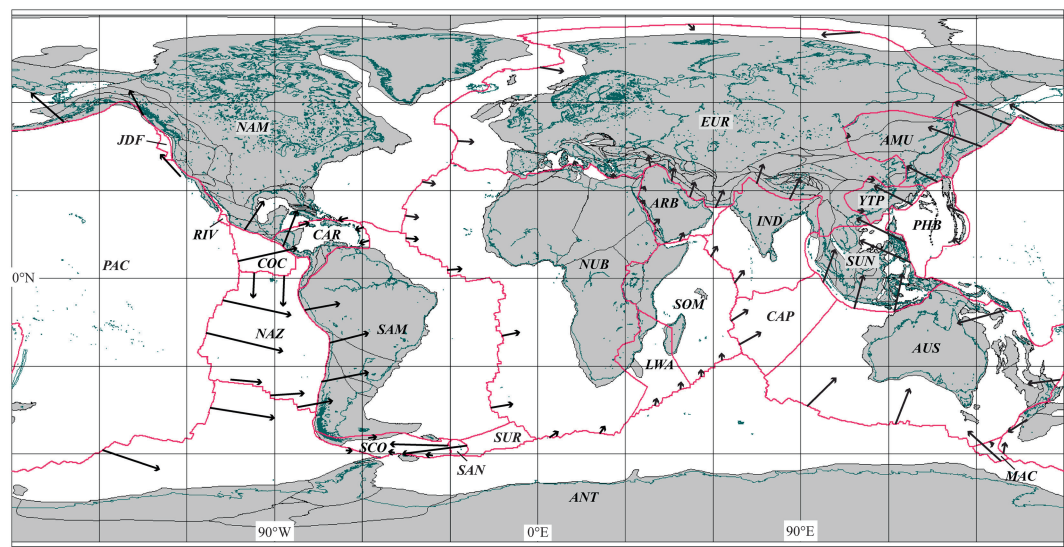
accepted kinematic models for the modern plates, which represents a refinement of a model published in 1990, known as NUVEL-1 (Northwestern University VElocity model ver. 1, DeMets et al. 1990). In this model, the average spreading rates used in the least squares fitting procedure are determined through the analysis of marine magnetic anomalies spanning the last  $\sim 3.2$  Myrs. The model includes 12 large plates, and the components of the Euler vectors are expressed in a reference frame fixed to the Pacific plate. The difference between the two versions is in the geomagnetic polarity time scales used to analyse the marine magnetic anomalies during the determination of the spreading rates, so that the angular velocities of NUVEL-1A are 95.62 % of the corresponding velocities listed in NUVEL-1.

One of the main problems of the classic models is represented by the very different time intervals associated with the input data. The spreading rates along the world's mid-ocean ridges, which are estimated through the analysis of marine magnetic anomalies, represent averages over the last  $\sim 3.2$  Myrs. These averages strongly depend from the choice of a geomagnetic polarity time scale. Conversely, earthquake slip vectors average directions of relative motions over much shorter time intervals (decades to centuries). Another problem is represented by the relatively small number of plates that are considered in these models, which limits their capability to represent the internal deformation of some continents. Therefore, there is not much surprise in seeing that inconsistencies often emerge when the linear velocities predicted on the basis of the Euler vectors are compared to velocities estimated from Global Positioning System (GPS) techniques and other geodetic methods. In fact, the latter data are consistent averages performed over a few decades, which are not necessarily representative of the long-term geological processes. Finally, the most serious issue of NUVEL-1A and its predecessors is probably the failure to satisfy the closure rule (Eq. 2.35) along some three-plate circuits. In particular, NUVEL-1A does not satisfy Eq. 2.36 at the Galapagos triple junction (Pacific-Cocos-Nazca circuit) and at the Bouvet

triple junction (Africa-South America-Antarctic circuit) at the desired level of confidence.

A major improvement to NUVEL-1A, which tries to overcome the difficulties mentioned above, has been proposed in recent times by DeMets et al. (2010). The new model, which has been called MORVEL (Mid-Ocean Ridge VElocity), extends the data set to the 25 plates shown in Fig. 2.38. With respect to the system of 23 plates shown in Fig. 2.16, this model decomposes the eastern part of Africa in two sub-plates (Lwandle and Somalia), separates two sub-plates (Capricorn, and Macquarie) from Australia, introduces the Yang-Tze plate in eastern Asia and Sur in the South Atlantic, but incorporates the Anatolian block in Eurasia, Easter in Nazca, and the Okhotsk plate in N. America. Using Eq. 2.34, we see that this model includes 46 triple junctions and 69 plate boundaries, 23 of which must be free boundaries. Differently from its predecessors, MORVEL is based on few earthquake slip directions. In this model, about 75 % of the input data are sea floor spreading rates and strikes of transform faults. The very limited usage of earthquake slip vectors ( $\sim 2$  % of the total data set) has minimized the possibility of biased estimates of relative velocity directions along the world's subduction zones, which are usually caused by forearc deformation (e.g., Jarrard 1986; McCaffrey 1992). Finally, it has been avoided a mix between long-term geological data and geodetic velocities in the estimation of Euler vectors, the usage of GPS data having been limited to the determination of the motion of six small plates, for which no other data were available. The 24 Euler vectors of MORVEL, relative to the Pacific plate, are listed in Table 2.3, while the resulting linear velocity fields between adjacent plates are shown in Fig. 2.38.

The kinematic models described so far furnish the Euler vectors of the major modern tectonic plates relative to the Pacific. The components of these vectors are expressed in the geographic reference frame (where London, Eurasia, has a fixed longitude). By vector summation, we can calculate the Euler vector of relative motion between any pair of plates, assess closure conditions



**Fig. 2.38** The plate velocity model MORVEL (DeMets et al. 2010). The *length of the arrows* indicates 20 times the relative displacement of a plate with respect to an adjacent plate. Plate labels are listed in Table 2.3

**Table 2.3** The MORVEL velocity model

Plate	Sym	$\lambda_e$	$\phi_e$	$\omega$	$\omega_x$	$\omega_y$	$\omega_z$
Amurian	AMU	65.9	−82.7	0.929	0.000841	−0.006567	0.014801
Antarctica	ANT	65.9	−78.5	0.887	0.001260	−0.006194	0.014132
Arabia	ARB	60.0	−33.2	1.159	0.008463	−0.005538	0.017518
Australia	AUS	60.1	6.3	1.079	0.009331	0.001030	0.016325
Caribbean	CAR	55.8	−77.5	0.905	0.001922	−0.008668	0.013064
Cocos	COC	42.2	−112.8	1.676	−0.008397	−0.019977	0.019649
Capricorn	CAP	62.3	−10.1	1.139	0.009098	−0.001621	0.017601
Eurasia	EUR	61.3	−78.9	0.856	0.001381	−0.007040	0.013105
India	IND	61.4	−31.2	1.141	0.008154	−0.004938	0.017484
Juan de Fuca	JDF	−0.6	37.8	0.625	0.008619	0.006685	−0.000114
Lwandle	LWA	60.0	−66.9	0.932	0.003191	−0.007481	0.014087
Macquarie	MAC	59.2	−8.0	1.686	0.014921	−0.002097	0.025276
North America	NAM	48.9	−71.7	0.750	0.002702	−0.008170	0.009864
Nubia	NUB	58.7	−66.6	0.935	0.003367	−0.007781	0.013944
Nazca	NAZ	55.9	−87.8	1.311	0.000492	−0.012819	0.018947
Philippine	PHB	−4.6	−41.9	0.890	0.011524	−0.010340	−0.001246
Rivera	RIV	25.7	−104.8	4.966	−0.019950	−0.075508	0.037587
South America	SAM	56.0	−77.0	0.653	0.001434	−0.006210	0.009449
Scotia	SCO	57.8	−78.0	0.755	0.001460	−0.006868	0.011150
Somalia	SOM	59.3	−73.5	0.980	0.002480	−0.008373	0.014707
Sur	SUR	55.7	−75.8	0.636	0.001534	−0.006064	0.009170
Sundaland	SUN	59.8	−78.0	0.973	0.001776	−0.008356	0.014677
South Sandwich	SAN	−3.8	−42.4	1.444	0.018570	−0.016957	−0.001670
Yang Tze Platform	YTP	65.5	−82.4	0.968	0.000927	−0.006945	0.015374

across plate circuits, test triple junction velocity triangles, etc. What we cannot do with these models, is to represent the *absolute* velocities of the tectonic plates with respect to a reference frame fixed to the deep mantle, for example fixed to the top of the transition zone as in Fig. 2.9. However, this is a necessary step if we want to consider the kinematics of tectonic plates in relation to the asthenospheric flows, and give a complete geodynamic formulation of plate tectonics. Such approach represents one the fundamental tasks of this book, thereby now we shall illustrate an approximate method to determine the Euler vectors in a reference frame fixed to the deep mantle.

The method was proposed 40 years ago by Solomon and Sleep (1974) and applies equally well to the modern plates and to a paleotectonic context (Solomon et al. 1977). These authors started from the assumption that the total torque  $N$  exerted on the lithosphere (Eq. 2.3) is zero, and that the asthenosphere is dragged passively by the overlying lithosphere. The first assumption is compatible with the fact that, apart from the case of space geodesy studies, we always represent plate motions through the *geological time*, not the physical time, even when we study the present day plate motions. When we consider physical processes that occur at the time scale of the last 2–3 Myrs, it is always necessary to neglect any motion related to the Earth's rotation, including variations in eccentricity of the orbit, axial tilt, and precession. In Chap. 6, we shall discuss the evidence that the total angular momentum  $L$  (Eq. 2.5) of the lithosphere is constant over time intervals of several Myrs, which implies that in equilibrium conditions  $N = \mathbf{0}$  at the time scale of the geological processes. We shall prove that also the second assumption is correct in conditions of geodynamic equilibrium, but *not* during episodes of plate acceleration, such as the northward acceleration of India during the Cretaceous to Eocene time interval (Cande and Stegman 2011).

The method of Solomon and Sleep is based on a balance of the torques exerted on the whole lithosphere. The torques applied on individual plates are associated with the viscous resistive drag force that the asthenosphere exerts on the

base of the overlying lithosphere, and with plate boundary forces, such as the gravitational forces exerted by slabs. However, it is not necessary to include symmetric features such as mid-ocean ridges in the torque balance, because in this instance the corresponding torques cancel out. Therefore, the two fundamental torques that must be included in the torque balance equation are those associated with drag forces and those arising from downward pull forces exerted by slabs. Let us assume that the passive drag applied at the base of the lithosphere follows a simple viscous law, so that it depends linearly from the velocity of the lithosphere relative to the base of the fluid asthenosphere. It is also reasonable to assume that the slab pull force exerted along an active margin does not depend from the plate velocity. Let  $\omega_i$  be the Euler vector of  $i$ -th plate relative to the top transition zone, and  $\mathbf{v}_i(\mathbf{r}) = \omega_i \times \mathbf{r}$  the corresponding linear velocity field at each point  $\mathbf{r}$  along its surface. The simplest law describing the resistive drag force per unit area (or *traction*) at the base of the lithosphere,  $\mathbf{T}_i = \mathbf{T}_i(\mathbf{r})$ , is the following one:

$$\mathbf{T}_i(\mathbf{r}) = -D_i \mathbf{v}_i(\mathbf{r}) = -D_i \omega_i \times \mathbf{r} \quad (2.58)$$

In this expression  $D_i$  is a drag coefficient which may depend from position. To obtain the total torque exerted on the  $i$ -th plate we must integrate the local torque per unit area,  $\mathbf{r} \times \mathbf{T}_i(\mathbf{r})$ , over the surface  $S_i$  of the plate:

$$N_i = - \int_{S_i} [D_i \mathbf{r} \times (\omega_i \times \mathbf{r})] dS \quad (2.59)$$

From this expression, it is easy to calculate the total torque exerted on the lithosphere by the underlying asthenosphere:

$$N = \sum_i N_i = - \sum_i D_i \int_{S_i} [\mathbf{r} \times (\omega_i \times \mathbf{r})] dS \quad (2.60)$$

where for simplicity we have assumed that  $D_i$  is constant along the surface of a plate. If we expand

the vector triple product in (2.60), this expression can be rewritten as follows:

$$N = -\sum_i D_i \int_{S_i} (\mathbf{r} \cdot \mathbf{r}) \boldsymbol{\omega}_i dS + \sum_i D_i \int_{S_i} (\mathbf{r} \cdot \boldsymbol{\omega}_i) \mathbf{r} dS_i \quad (2.61)$$

If we assume that the Earth's radius is normalized to unity, then  $\mathbf{r} \cdot \mathbf{r} = 1$ , so that:

$$N = -\sum_i D_i S_i \boldsymbol{\omega}_i + \sum_i D_i \int_{S_i} (\mathbf{r} \cdot \boldsymbol{\omega}_i) \mathbf{r} dS_i \quad (2.62)$$

Expression (2.62) can be further simplified introducing a new tensor quantity, which is strictly related to the inertial tensor of a tectonic plate (Eq. 2.23). Using the index notation (see Appendix 1), the components of this quantity are given by:

$$\mathcal{Q}_{jk}^i \equiv \int_{S_i} (\delta_{jk} - x_j x_k) dS = A_i \delta_{jk} - \int_{S_i} x_j x_k dS; \quad j, k = 1, 2, 3 \quad (2.63)$$

where  $A_i$  is the area of the  $i$ -th plate. Using this new tensor quantity, which depends only from the plate geometry, Expression (2.62) can be rewritten as follows:

$$N = -\sum_i D_i \mathcal{Q}^i \boldsymbol{\omega}_i \quad (2.64)$$

If this were the only torque exerted on the lithosphere, the torque balance equation would be written:  $N = \mathbf{0}$ , that is:

$$\sum_i D_i \mathcal{Q}^i \boldsymbol{\omega}_i = \mathbf{0} \quad (2.65)$$

Let  $\boldsymbol{\omega}_r$  be the Euler vector of a reference plate, for example the Pacific plate, with respect to the top transition zone. Knowing the Euler vector of any other plate with respect to the reference

plate,  $\boldsymbol{\omega}_{ir}$ , it is possible to determine its absolute Euler vector,  $\boldsymbol{\omega}_i$ , by adding the absolute angular velocity of the reference plate:  $\boldsymbol{\omega}_i = \boldsymbol{\omega}_{ir} + \boldsymbol{\omega}_r$ . Therefore, Eq. (2.65) can be viewed as a linear system of three equations with respect to the three unknown components of  $\boldsymbol{\omega}_r$ :

$$\left( \sum_i D_i \mathcal{Q}^i \right) \boldsymbol{\omega}_r = -\sum_i D_i \mathcal{Q}^i \boldsymbol{\omega}_{ir} \quad (2.66)$$

The total  $\mathcal{Q}$  tensor for the whole lithosphere can be obtained simply by summation of the tensors  $\mathcal{Q}^i$  associated with each plate. It results:

$$\mathcal{Q} = \sum_i \mathcal{Q}^i = \frac{8\pi}{3} \mathbf{I} \quad (2.67)$$

where  $\mathbf{I}$  is the identity matrix. Further simplification of Eq. (2.66) follows if we assume that the drag coefficients  $D_i$  coincide for all plates:  $D_i = D$ . In this instance, using (2.67) we obtain an immediate solution for  $\boldsymbol{\omega}_r$  in terms of the relative Euler vectors of a velocity model:

$$\boldsymbol{\omega}_r = -\frac{3}{8\pi} \sum_i \mathcal{Q}^i \boldsymbol{\omega}_{ir} \quad (2.68)$$

This solution corresponds to a condition of *no-net-rotation* (NRR) for the whole lithosphere ( $\mathbf{L} = \mathbf{0}$ ). In fact, for  $D_i = D$  Eq. (2.65) can be rewritten as follows:

$$\sum_i \mathcal{Q}^i \boldsymbol{\omega}_i \equiv \mathcal{Q} \boldsymbol{\Omega} = \frac{8\pi}{3} \boldsymbol{\Omega} = \mathbf{0} \Rightarrow \boldsymbol{\Omega} = \mathbf{0} \quad (2.69)$$

where  $\boldsymbol{\Omega}$  can be considered as the net rotation of the whole lithosphere. It should be noted that the solution (2.68) only holds in the unlikely event that the unique torques exerted on the lithosphere come from asthenospheric drag, *and* that the drag coefficient  $D$  can be considered constant over the entire lithosphere. Of course, none of these two strong conditions is likely to be verified. Slab pull forces are essential components of the global torque balance, and the drag coefficient along the irregular continental LAB cannot be

equal to the drag coefficient of the oceanic areas. Notwithstanding these issues, the NNR condition (2.68) has been widely used to build “absolute” plate motions models (e.g., Argus and Gordon 1991), and represents the basis for the definition of a geocentric reference frame. This is the International Terrestrial Reference Frame (ITRF), which is particularly important for the representation of kinematic data obtained from geodetic techniques, but it is also linked to an inertial frame tied to stellar objects, the Celestial Reference Frame. This NNR reference frame is periodically updated by the International Earth Rotation and Reference Systems Service (IERS). It is realized through the acquisition of time series of mean station positions at weekly or daily sampling from a global network of observation sites equipped with various space geodesy systems: very long baseline interferometry (VLBI), satellite laser ranging (SLR), Global Positioning System (GPS), and Doppler Orbitography Radio-positioning Integrated by Satellite (DORIS) (Altamimi et al. 2002). Then, an assignment of precise coordinates and linear velocities at reference epochs is made. These data are used, in conjunction with Eq. (2.17), to estimate statistically the angular velocities of each plate having an observation site. Finally, a best fit alignment with the current plates velocity model NNR-NUVEL-1A is performed, in order to satisfy the condition (2.68) (Altamimi et al. 2003).

We can determine the components of the tensors  $\mathbf{Q}^i$  using a computational method proposed by Schettino (1999b). Table 2.4 lists the six independent components of these tensors for the set of MORVEL plates shown in Fig. 2.38. This data set can be used to determine the Euler vector of the reference plate through Eq. (2.68). The instantaneous Euler pole of the Pacific plate, determined on the basis of the relative Euler vectors of Table 2.3 and the  $\mathbf{Q}$  tensor components of Table 2.4, is located at 63.5°S, 114.4°E, and its angular velocity is  $\omega = 0.65^\circ/\text{Myr}$ . The NNR version of MORVEL is listed in Table 2.5, while the corresponding velocity fields are shown in Fig. 2.39.

An estimation of the errors associated with the computation of the tensors  $\mathbf{Q}^i$  can be performed

as follows. First, it is possible to show that the area of each plate,  $A_i$ , can be calculated decomposing the corresponding spherical polygon into a set of spherical triangles, then using the well-known Girard’s formula for calculating the area of each triangle (Schettino 1999b).

From (2.63), we see that these quantities are related to the diagonal components of  $\mathbf{Q}^i$  by the following expression:

$$Tr(\mathbf{Q}^i) = \sum_j \int_{S_i} (1 - x_j^2) dS = 2A_i \quad (2.70)$$

Therefore, an estimate of the errors associated with the diagonal components of  $\mathbf{Q}^i$ , which are listed in the last column of Table 2.4, can be obtained by evaluating the expression:

$$\varepsilon_i = \frac{Tr(\mathbf{Q}^i) - 2A_i}{2A_i} \quad (2.71)$$

It is important to note that the velocity fields of the NNR version of MORVEL do not really represent velocities relative to the deep mantle. In fact, the equations associated with the NNR condition (2.68) do not consider the contribution of slab pull forces to the total torque balance, and are based upon the implausible assumption that the drag coefficient is uniform across the Earth’s LAB. However, the method described above can be considered as a good starting point for the study of the absolute plate motions. For example, we can improve the model introducing in the torque balance equation the torques associated with the pull exerted by subducting slabs.

Slab pull is a downward-directed force that a sinking slab exerts on the unsubducted lithosphere along a trench line (Forsyth and Uyeda 1975). If  $\mathbf{T}_i$  is the small circle representative of a trench line, then this force is everywhere normal to  $\mathbf{T}_i$ . Therefore, if  $d\mathbf{l}$  is an infinitesimal vector element tangent to  $\mathbf{T}_i$ , then the torque exerted on the unsubducted lithosphere is given by:

$$\mathbf{N}_i = C_i \int_{\mathbf{T}_i} \mathbf{r} \times (d\mathbf{l} \times \mathbf{r}) \quad (2.72)$$

**Table 2.4**  $Q$  tensors and areas of the 25 MORVEL plates

Plate	A	$Q_{11}$	$Q_{22}$	$Q_{33}$	$Q_{12}$	$Q_{13}$	$Q_{23}$	Diag. Err.
AMU	0.130659	0.108248	0.089481	0.063589	0.028732	0.036320	-0.051295	0.0000 %
ANT	1.434290	1.328262	1.176115	0.364247	-0.050791	0.052812	0.080667	0.0015 %
ARB	0.120824	0.074248	0.066810	0.100589	-0.048782	-0.029553	-0.031041	-0.0004 %
AUS	0.935403	0.602384	0.568115	0.700304	0.230373	-0.218401	0.241845	-0.0002 %
CAR	0.103729	0.094940	0.014300	0.098213	0.024762	-0.006107	0.020566	-0.0024 %
COC	0.072230	0.071072	0.003020	0.070372	-0.005543	0.001064	0.010141	0.0028 %
CAP	0.203647	0.196537	0.022175	0.188580	-0.021636	0.007182	0.045603	-0.0005 %
EUR	1.218422	1.017712	0.913393	0.505738	-0.041466	-0.222433	-0.315605	0.0000 %
IND	0.30636	0.286350	0.042306	0.284051	-0.057048	-0.013096	-0.060493	-0.0021 %
JDF	0.006315	0.005162	0.004356	0.003111	-0.001501	0.001916	0.002491	-0.0079 %
LWA	0.117115	0.063149	0.081116	0.089959	-0.043343	0.036053	0.029664	-0.0026 %
MAC	0.007890	0.006131	0.007510	0.002139	0.000812	-0.003172	0.001465	0.0000 %
NAM	1.440479	1.282025	1.008008	0.590974	0.079145	0.026680	0.378356	0.0017 %
NUB	1.440653	0.372568	1.301217	1.207515	-0.051346	-0.005428	0.044223	-0.0002 %
NAZ	0.403564	0.391445	0.070630	0.345043	-0.014536	-0.003992	-0.115869	-0.0012 %
PAC	2.681816	1.204054	2.045135	2.114430	-0.400314	0.062295	-0.057354	-0.0002 %
PHB	0.144484	0.081761	0.078620	0.128588	0.062670	0.029123	-0.029347	0.0003 %
RIV	0.002486	0.002289	0.000489	0.002193	-0.000625	0.000239	0.000763	-0.0201 %
SAM	1.023883	0.624948	0.586878	0.835938	0.344415	0.181243	-0.174029	-0.0001 %
SCO	0.042001	0.036816	0.034549	0.012637	0.005706	0.012013	-0.014486	0.0000 %
SOM	0.354795	0.221032	0.153739	0.334814	-0.154901	0.024755	0.035861	-0.0007 %
SUR	0.027055	0.018681	0.026496	0.008933	0.001954	0.012245	-0.002957	0.0000 %
SUN	0.281465	0.232911	0.054798	0.275220	0.093052	0.004760	-0.016178	-0.0002 %
SAN	0.004543	0.003525	0.004269	0.001292	0.000527	0.001817	-0.000940	0.0000 %
YTP	0.062249	0.051303	0.024035	0.049159	0.019688	0.011653	-0.022080	-0.0008 %
Earth	12.566357	8.377553	8.377560	8.377628	0.000004	-0.000012	-0.000029	0.0001 %
Av.%.error	-0.0001 %	-0.0003 %	-0.0002 %	0.0006 %	0.0000 %	0.0000 %	0.0000 %	

Units are in steradians

where  $C_i$  is a constant that is assumed to be independent from the subduction velocity, and the line integral is calculated following a counterclockwise path.

If we expand the triple vector product in (2.72), we obtain the following simple expression for the torque:

$$N_i = C_i \int_{\mathbf{T}_i} d\mathbf{l} = C_i (\mathbf{r}_f - \mathbf{r}_i) \quad (2.73)$$

In this expression,  $\mathbf{r}_i$  and  $\mathbf{r}_f$  are, respectively, the position vectors of the start and end points of the trench line  $\mathbf{T}_i$ . Therefore, we see that the torque exerted on a subducting plate by the attached slab only depends from the width of the subduction zone, not by its curvature. If we

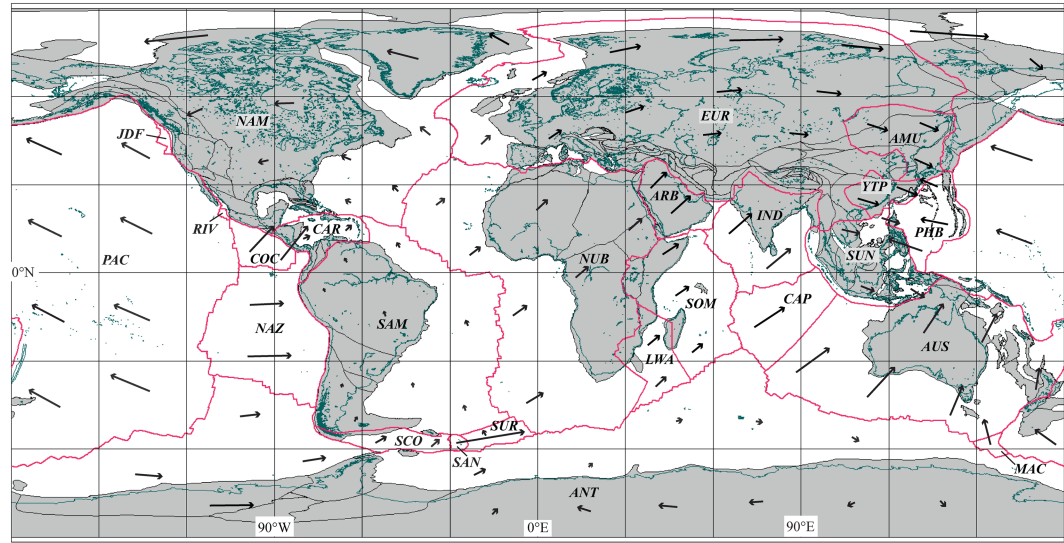
introduce the torques (2.73) in the total torque balance equation, we obtain a more realistic equation, which potentially can be solved to determine the absolute Euler vector of the reference plate:

$$\left( \sum_i D_i \mathbf{Q}^i \right) \boldsymbol{\omega}_r = \sum_j C_j (\mathbf{r}_f^j - \mathbf{r}_i^j) - \sum_i D_i \mathbf{Q}^i \boldsymbol{\omega}_{ir} \quad (2.74)$$

This is a system of three equations in the unknown components of  $\boldsymbol{\omega}_r$ , which can be solved if the drag coefficients  $D_i$  and the constants  $C_j$  are known. In this instance, the lithosphere always has a non-zero angular momentum, even when  $D_i=D$  for all plates.

**Table 2.5** The NNR—MORVEL velocity model

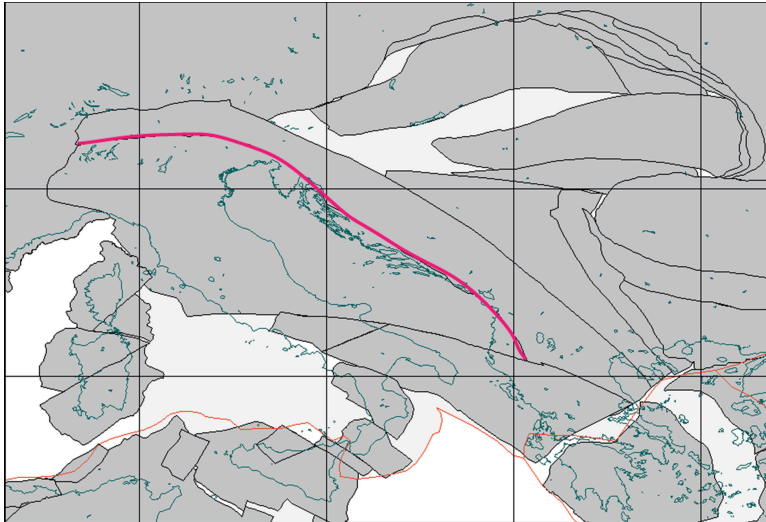
Plate	$\lambda_e$	$\phi_e$	$\omega$	$\omega_1$	$\omega_2$	$\omega_3$
AMU	63.4	237.5	0.298	−0.001248	−0.001962	0.004645
ANT	65.7	242.4	0.250	−0.000829	−0.001589	0.003976
ARB	48.8	351.7	0.561	0.006374	−0.000933	0.007363
AUS	33.9	37.9	0.634	0.007241	0.005636	0.006170
CAR	35.6	267.6	0.286	−0.000168	−0.004062	0.002908
COC	27.0	235.7	1.197	−0.010487	−0.015371	0.009493
CAP	44.3	23.1	0.610	0.007008	0.002985	0.007445
EUR	49.3	253.8	0.223	−0.000708	−0.002435	0.002949
IND	50.4	356.9	0.545	0.006064	−0.000333	0.007329
JDF	−38.2	60.0	0.951	0.006529	0.011291	−0.010270
LWA	51.9	291.0	0.286	0.001101	−0.002876	0.003932
MAC	49.2	371.1	1.145	0.012831	0.002508	0.015120
NAM	−4.6	279.7	0.208	0.000612	−0.003564	−0.000292
NUB	47.9	291.9	0.293	0.001277	−0.003175	0.003788
NAZ	46.4	259.0	0.695	−0.001597	−0.008213	0.008791
PAC	−63.5	114.4	0.650	−0.002090	0.004605	−0.010156
PHB	−45.9	328.7	0.909	0.009435	−0.005735	−0.011401
RIV	20.3	252.7	4.535	−0.022040	−0.070903	0.027431
SAM	−22.2	247.8	0.107	−0.000656	−0.001604	−0.000707
SCO	23.0	254.5	0.146	−0.000630	−0.002263	0.000995
SOM	50.2	275.9	0.339	0.000391	−0.003767	0.004551
SUR	−32.3	249.2	0.106	−0.000555	−0.001459	−0.000986
SUN	50.2	265.2	0.337	−0.000313	−0.003750	0.004522
SAN	−29.9	323.2	1.361	0.016481	−0.012351	−0.011826
YTP	63.4	243.6	0.334	−0.001163	−0.002339	0.005218



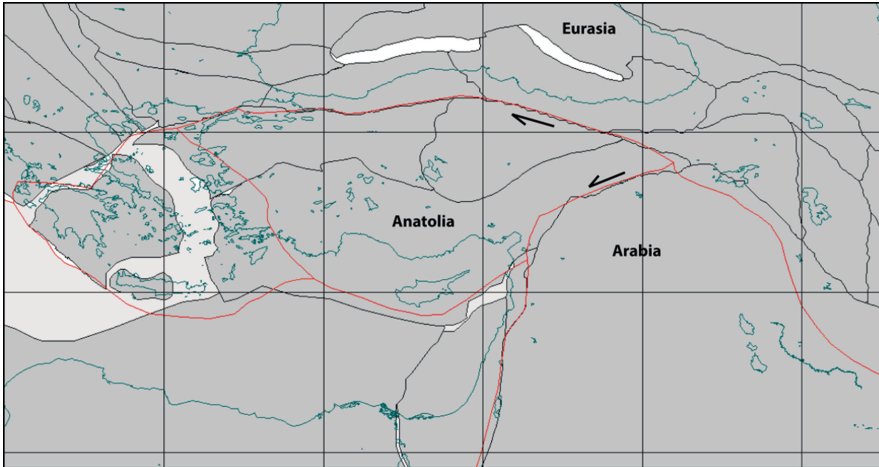
**Fig. 2.39** The “absolute” plate velocity model NNR—MORVEL (DeMets et al. 2010). The *length of the arrows* indicates 20 times the relative displacement of a plate with respect to an adjacent plate

## Problems

1. An Eulerian reference frame is a geocentric reference frame obtained rotating an Euler pole to the North pole. Determine the equation of transformation from geographic to Eulerian latitude;
2. The number of edges in a circuit with  $p$  plates is always  $e = p - 1$ . Explain why;
3. Show that the relation  $e = p - 1$  is compatible with (2.34),  $e$  and  $p$  being the number of edges and the number of plates in a circuit;
4. Given the three-plates system formed by the Pacific, North American, and Juan de Fuca plates, determine the relative velocity vector of Juan de Fuca with respect to N. America at  $(46.5^\circ\text{N}, 125.8^\circ\text{W})$  using the data in Table 2.5;
5. Assuming that the spreading asymmetry is zero along the Juan de Fuca Ridge, how long time is required for a point at  $(46.9^\circ\text{N}, 129.4^\circ\text{W})$  along the ridge to enter the Cascadia Trench and what is the predicted location of ridge subduction?
6. Subduction of the Capricorn plate beneath Sundaland along the Sumatra Trench is highly oblique. Strike-slip motion along the Sumatran Fault, which is parallel to the trench in the forearc region, determines partitioning of such oblique subduction into a trench-normal component and a trench-parallel component. Determine the slip rate and the sense of shear along the Sumatran Fault at  $(2.5^\circ\text{S}, 101.5^\circ\text{E})$ ;
7. The Periadriatic Line in northern Italy and Croatia is a wide E–W and NW–SE structure that accommodated strike-slip motion between Africa and Europe in the geologic past (see figure). What would be the style of this fault at  $(46.4^\circ\text{N}, 11^\circ\text{E})$  and  $(42.8^\circ\text{N}, 17.8^\circ\text{E})$  if it were a present day plate boundary?



8. Anatolia is a small microplate between Arabia and Europe in the eastern Mediterranean, whose N and SE boundaries are transcurrent faults (see figure). Starting from the relative velocity of Arabia with respect to Europe, calculate the westward escape velocity of this microplate along its strike-slip boundaries;



9. Determine the evolution of the Pacific-N. America-J. de Fuca triple junction in the Pacific reference frame, and describe the geological setting around the region where the corresponding plate boundaries meet;
10. Determine the time interval of stability of the triple junction between Nazca, Antarctica, and S. America, and the subsequent migration path;

## References

- Altamimi Z, Sillard P, Boucher C (2002) ITRF2000: a new release of the international terrestrial reference frame for earth science applications. *J Geophys Res* 107(B10):2214. doi:10.1029/2001JB000561
- Altamimi Z, Sillard P, Boucher C (2003) The impact of a no-net-rotation condition on ITRF2000. *Geophys Res Lett* 30(2):1064. doi:10.1029/2002GL016279
- Anderson-Fontana S, Engeln JF, Lundgren P, Larson RL, Stein S (1986) Tectonics and evolution of the Juan Fernandez microplate at the Pacific-Nazca-Antarctic triple junction. *J Geophys Res* 91(B2):2005–2018. doi:10.1029/JB091iB02p02005
- Argus DF, Gordon RG (1991) No-net-rotation model of current plate velocities incorporating plate motion model NUVEL-1. *Geophys Res Lett* 18(11):2039–2042
- Beck ME Jr (1986) Model for late Mesozoic-early Tertiary tectonics of coastal California and western Mexico and speculations on the origin of the San Andreas Fault. *Tectonics* 5(1):49–64. doi:10.1029/TC005i001p00049
- Besse J, Courtillot V (1988) Paleogeographic maps of the continents bordering the Indian Ocean since the early Jurassic. *J Geophys Res* 93(B10):11,791–11,808
- Bird P (2003) An updated digital model of plate boundaries. *Geochem Geophys Geosyst* 4(3):1027. doi:10.1029/2001GC000252
- Briaies A, Patriat P, Tapponnier P (1993) Updated interpretation of magnetic anomalies and seafloor spreading stages in the South China Sea: implications for the Tertiary tectonics of Southeast Asia. *J Geophys Res* 98(B4):6299–6328
- Brun J-P (1999) Narrow rifts versus wide rifts: inferences for the mechanics of rifting from laboratory experiments. *Philos Trans R Soc Lond* 357:695–712
- Buck WR (1991) Modes of continental lithospheric extension. *J Geophys Res* 96(B12):20161–20178. doi:10.1029/91JB01485
- Bullard EC, Everett JE, Smith AG (1965) The fit of the continents around the Atlantic: a symposium on continental drift. *Philos Trans R Soc Lond A* 258(1088): 41–51
- Butler RWH, Spencer S, Griffiths HM (1997) Transcurrent fault activity on the Dead Sea Transform in Lebanon and its implications for plate tectonics and seismic hazard. *J Geol Soc* 154:757–760. doi:10.1144/gsjgs.154.5.0757
- Campbell WH (2003) Introduction to geomagnetic fields, 2nd edn. Cambridge University Press, Cambridge, UK, 337 pp
- Cande SC, Stegman DR (2011) Indian and African plate motions driven by the push force of the Réunion plume head. *Nature* 475:47–52. doi:10.1038/nature10174
- Chase CG (1978) Plate kinematics: the Americas, East Africa, and the rest of the world. *Earth Planet Sci Lett* 37:355–368
- Coffin MF, Rabinowitz PD (1987) Reconstruction of Madagascar and Africa: evidence from the Davie Fracture Zone and Western Somali Basin. *J Geophys Res* 92(B9):9385–9406. doi:10.1029/JB092iB09p09385
- Contrucci I, Klingelhöfer F, Perrot J, Bartolome R, Gutscher MA, Sahabi M, Malod J, Rehault J-P (2004) The crustal structure of the NW-Moroccan continental

- margin from wide-angle and reflection seismic data. *Geophys J Int* 159:117–128
- Corti G, Bonini M, Conticelli S, Innocenti F, Manetti P, Sokoutis D (2003) Analogue modelling of continental extension: a review focused on the relations between the patterns of deformation and the presence of magma. *Earth Sci Rev* 63:169–247. doi:10.1016/S0012-8252(03)00035-7
- Cox A, Hart RB (1986) *Plate tectonics: how it works*. Blackwell Scientific Publications, Palo Alto, 392 pp
- Dalziel IWD (1981) Back-arc extension in the southern Andes: a review and critical reappraisal. *Philos Trans R Soc Lond A* 300(1454):319–335
- DeMets C, Gordon RG, Argus DF, Stein S (1990) Current plate motions. *Geophys J Int* 101:425–478
- DeMets C, Gordon RG, Argus DF, Stein S (1994) Effect of recent revisions to the geomagnetic reversal timescale. *Geophys Res Lett* 21:2191–2194
- DeMets C, Gordon RG, Argus DF (2010) Geologically current plate motions. *Geophys J Int* 181:1–80. doi:10.1111/j.1365-246X.2009.04491.x
- Dewey JF (1975) Finite plate implications: some implications for the evolution of rock masses at plate margins. *Am J Sci* 275-A:260–284
- Dewey JF, Helman ML, Turco E, Hutton DHW, Knott SD (1989) Kinematics of the Western Mediterranean. In: Coward MP, Dietrich D, Park RG (eds) *Alpine tectonics*. *Geol Soc Spec Publ* 45, London, pp 265–283
- Fairhead JD (1988) Mesozoic plate tectonic reconstructions of the central South Atlantic Ocean: the role of the West and Central African rift system. *Tectonophysics* 155(1–4):181–191. doi:10.1016/0040-1951(88)90265-X
- Forsyth D, Uyeda S (1975) On the relative importance of the driving forces of plate motion. *Geophys J Int* 43(1):163–200. doi:10.1111/j.1365-246X.1975.tb00631.x
- Fournier M, Patriat P, Leroy S (2001) Reappraisal of the Arabia–India–Somalia triple junction kinematics. *Earth Planet Sci Lett* 189:103–114
- Frank FC (1968) Curvature of island arcs. *Nature* 220:363
- Frisch W, Meschede M, Blakey R (2011) *Plate tectonics*. Springer, Berlin, 212 pp
- Gaina C, Müller DR, Royer J-Y, Stock J, Hardebeck J, Symonds P (1998) The tectonic history of the Tasman Sea: a puzzle with 13 pieces. *J Geophys Res* 103(B6):12413–12433. doi:10.1029/98JB00386
- Gaina C, Gernigon L, Ball P (2009) Palaeocene–recent plate boundaries in the NE Atlantic and the formation of the Jan Mayen microcontinent. *J Geol Soc Lond* 166:601–616. doi:10.1144/0016-76492008-112
- Garfunkel Z (1981) Internal structure of the Dead Sea leaky transform (rift) in relation to plate kinematics. *Tectonophysics* 80(1–4):81–108. doi:10.1016/0040-1951(81)90143-8
- Gould R (1988) *Graph theory*. Benjamin, Menlo Park, 332 pp
- Hamilton WB (1987) Crustal extension in the basin and Range Province, southwestern United States. In: Coward MP, Dewey JF, Hancock PL (eds) *Continental extensional tectonics*, *GSA Spec. Publ.*, 28 pp 155–176
- Hamilton WB (2002) The closed upper–mantle circulation of plate tectonics. In: Stein S, Freymueller JT (eds) *Plate boundary zones*, vol 30, *Geodynamics series*. AGU, Washington, DC, pp 359–410. doi:10.1029/GD030p0359
- Hey RN (1977) A new class of pseudofaults and their bearing on plate tectonics: a propagating rift model. *Earth Planet Sci Lett* 37:321–325
- Hey RN, Menard HW, Atwater TM, Caress DW (1988) Changes in direction of seafloor spreading revisited. *J Geophys Res* 93(B4):2803–2811
- Jarrard RD (1986) Relations among subduction parameters. *Rev Geophys* 24:217–284
- Jestin F, Huchon P, Gaulier JM (1994) The Somalia plate and the East African rift system: present-day kinematics. *Geophys J Int* 116(3):637–654. doi:10.1111/j.1365-246X.1994.tb03286.x
- Jung W-Y, Vogt PR (1997) A gravity and magnetic anomaly study of the extinct Aegir Ridge, Norwegian Sea. *J Geophys Res* 102(B3):5065–5089
- Kleinrock MC, Phipps Morgan J (1988) Triple junction reorganization. *J Geophys Res* 93(B4):2981–2996. doi:10.1029/JB093iB04p02981
- Korhonen JV et al (2007) Magnetic anomaly map of the world (and associated DVD), Scale: 1:50,000,000, 1st edn. Commission for the Geological Map of the World, Paris
- Kreemer C, Holt WE, Haines AJ (2003) An integrated global model of present-day plate motions and plate boundary deformation. *Geophys J Int* 154: 8–34
- Laville E, Piqué A (1991) La Distension crustale atlantique et atlasique au Maroc au debut du Mesozoïque; le rejeu des structures hercyniennes. *Bull Soc Géol Fr* 162(6):1161–1171
- Lebrun J-F, Lamarche G, Collet J-Y (2003) Subduction initiation at a strike-slip plate boundary: the Cenozoic Pacific–Australian plate boundary, south of New Zealand. *J Geophys Res* 108:2453. doi:10.1029/2002JB002041
- Lonsdale P (1988) Structural pattern of the Galapagos microplate and evolution of the Galapagos triple junctions. *J Geophys Res* 93(B11):13551–13574. doi:10.1029/JB093iB11p13551
- McCaffrey R (1992) Oblique plate convergence, slip vectors, and forearc deformation. *J Geophys Res* 97:8905–8915
- McKenzie D (1976) The East Anatolian Fault: a major structure in Eastern Turkey. *Earth Planet Sci Lett* 29(1):189–193. doi:10.1016/0012-821X(76)90038-8
- McKenzie D (1978) Some remarks on the development of sedimentary basins. *Earth Planet Sci Lett* 40(1):25–32. doi:10.1016/0012-821X(78)90071-7
- McKenzie D, Morgan WJ (1969) Evolution of triple junctions. *Nature* 224:125–133
- Menard HW, Atwater TM (1968) Changes in direction of sea floor spreading. *Nature* 219:463–467

- Minster JB, Jordan TH (1978) Present-day plate motions. *J Geophys Res* 83:5331–5354
- Müller RD, Royer J-Y, Lawver LA (1993) Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* 21:275–278
- Müller RD, Roest WR, Royer J-Y, Gahagan LM, Sclater JG (1997) Digital isochrons of the world's ocean floor. *J Geophys Res* 102(B2):3211–3214
- Patriat P, Courtillot V (1984) On the stability of triple junctions and its relation to episodicity in spreading. *Tectonics* 3(3):317–332. doi:10.1029/TC003i003p00317
- Pindell JL, Cande SC, Pitman WC III, Rowley DB, Dewey JF, Labrecque J, Haxby W (1988) A plate-kinematic framework for models of Caribbean evolution. *Tectonophysics* 155:121–138
- Ramos VA, Cristallini EO, Pérez DJ (2002) The Pampean flat-slab of the Central Andes. *J S Am Earth Sci* 15(1):59–78. doi:10.1016/S0895-9811(02)00006-8
- Roest WR, Srivastava SP (1989) Sea-floor spreading in the Labrador Sea: a new reconstruction. *Geology* 17:1000–1003
- Ross MI, Scotese CR (1988) A hierarchical tectonic model of the Gulf of Mexico and Caribbean region. *Tectonophysics* 155(1–4):139–168
- Rowley DB, Lottes AL (1988) Plate-kinematic reconstructions of the North Atlantic and Arctic: late Jurassic to present. *Tectonophysics* 155(1–4):73–120
- Sandwell DT, Smith WHF (1997) Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *J Geophys Res* 102:10039–10054
- Schettino A (1998) Computer aided paleogeographic reconstructions. *Comput Geosci* 24(3):259–267
- Schettino A (1999a) Polygon intersections in spherical topology: application to plate tectonics. *Comput Geosci* 25(1):61–69
- Schettino A (1999b) Computational methods for calculating geometric parameters of tectonic plates. *Comput Geosci* 25(8):897–907
- Schettino A, Scotese CR (2002) Global kinematic constraints to the tectonic history of the Mediterranean region and surrounding areas during the Jurassic and Cretaceous. In: Rosenbaum G, Lister GS (eds) *Reconstruction of the evolution of the Alpine-Himalayan orogen*. *J Virtual Explor* 7:147–166
- Schettino A, Scotese CR (2005) Apparent polar wander paths for the major continents (200 Ma – present day): a paleomagnetic reference frame for global plate tectonic reconstructions. *Geophys J Int* 163(2):727–759
- Schettino A, Tassi L (2012) Trench curvature and deformation of the subducting lithosphere. *Geophys J Int* 188(1):18–34. doi:10.1111/j.1365-246X.2011.05262.x
- Schettino A, Turco E (2006) Plate kinematics of the Western Mediterranean region during the Oligocene and early Miocene. *Geophys J Int* 166(3):1398–1423
- Schettino A, Turco E (2009) Breakup of Pangaea and plate kinematics of the central Atlantic and Atlas regions. *Geophys J Int* 110:1078–1097
- Schettino A, Turco E (2011) Tectonic history of the western Tethys since the late Triassic. *GSA Bull* 123(1/2):89–105. doi:10.1130/B30064.1
- Schlich R (1974) Sea floor spreading history and deep sea drilling results in the Madagascar and Mascarene Basins, Western Indian Ocean. In: Simpson ESW, Schlich R, et al. (eds) *Initial reports of the deep sea drilling project, 25, US Govt. Printing Office, Washington*, pp 663–678
- Schmid SM, Kissling E (2000) The arc of the western Alps in the light of geophysical data on deep crustal structure. *Tectonics* 19(1):62–85
- Schmid SM, Pfiffner OA, Froitzheim N, Schönborn G, Kissling E (1996) Geophysical-geological transect and tectonic evolution of the Swiss-Italian Alps. *Tectonics* 15(5):1036–1064
- Sengör AMC (1979) The North Anatolian transform fault: its age, offset and tectonic significance. *J Geol Soc* 136:269–282. doi:10.1144/gsjgs.136.3.0269
- Shaw PR (1987) Investigations of relative plate motions in the South Atlantic using SEASAT altimeter data. *J Geophys Res* 92(B9):9363–9375
- Smythe DK (1989) Rockall Trough-Cretaceous or late Palaeozoic? *Scott J Geol* 25(1):5–43
- Solomon SC, Sleep NH (1974) Some simple physical models for absolute plate motions. *J Geophys Res* 79(17):2557–2567. doi:10.1029/JB079i017p02557
- Solomon SC, Sleep NH, Jurdy DM (1977) Mechanical models for absolute plate motions in the early Tertiary. *J Geophys Res* 82(2):203–212. doi:10.1029/JB082i002p00203
- Treves B (1984) Orogenic belts as accretionary prisms: the example of the northern Apennines. *Ofioliti* 9:577–618
- Vigny C, Socquet A, Rangin C, Chamot-Rooke N, Pubellier M, Bouin M-N, Bertrand G, Becker M (2003) Present-day crustal deformation around Sagaing fault, Myanmar. *J Geophys Res* 108:2533. doi:10.1029/2002JB001999,B11
- Wernicke B (1985) Uniform-sense normal simple shear of the continental lithosphere. *Can J Earth Sci* 22:108–125
- Whittaker JM, Müller RD, Roest WR, Wessel P, Smith WHF (2008) How supercontinents and superoceans affect seafloor roughness. *Nature* 456:938–941. doi:10.1038/nature07573

**Abstract**

This is the first of four chapters devoted to the techniques for collecting and analysing kinematic data. The fundamental source of data for determining past plate motions is furnished by rock magnetization. In this chapter, I describe the different classes of rock magnetism, the main magnetic minerals, and the origin of ferromagnetism.

**3.1 Electric Currents**

An *electric current* is a material flow of electric charges. These charges may be single particles (e.g., electrons) that move through a medium or in the empty space or, in a continuous mechanics representation (see Sect. 2.1), we can think about an electric current as a flow of volume elements  $dV$  having electric charge. Given an arbitrary surface  $S$  in the physical space, the current  $I(S)$  across this surface is defined as the charge  $q$  that walks through  $S$  in the unit time:

$$I(S) = \frac{dq}{dt} \quad (3.1)$$

Let us consider an infinitesimal surface element of  $S$ . This quantity can be represented by a vector  $d\mathbf{S} = \mathbf{n}dS$ , whose versor  $\mathbf{n}$  is normal to the surface, and having magnitude  $dS$  equal to the infinitesimal area of the surface element (Fig. 3.1).

The charge  $dq$  crossing  $d\mathbf{S}$  in an infinitesimal time interval  $dt$  coincides with the electric charge that originally was contained in a small cylinder

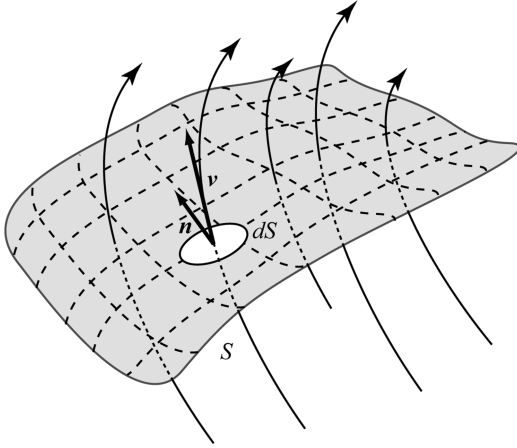
$dV$ , having base  $dS$  and height equal to the distance  $h$  travelled by the charge in the direction  $\mathbf{n}$ . Clearly,  $h = (\mathbf{v} \cdot \mathbf{n})dt$ . Then:

$$\begin{aligned} dq &= \rho dV = \rho (\mathbf{v} \cdot \mathbf{n}) dt dS \\ &= \rho \mathbf{v} \cdot d\mathbf{S} dt \equiv \mathbf{j} \cdot d\mathbf{S} dt \end{aligned} \quad (3.2)$$

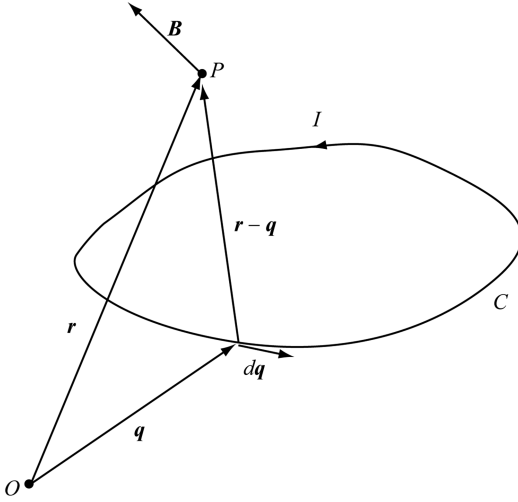
Here, the scalar quantity  $\rho$  and the vector  $\mathbf{j} = \rho \mathbf{v}$  are respectively the *density of charge* and the *density of current*. Combining this expression with (3.1), we have:

$$I(S) = \int_S \mathbf{j} \cdot d\mathbf{S} \quad (3.3)$$

The electric charge is expressed in Coulomb [C], while currents are measured in Ampere [A = 1 °C s<sup>-1</sup>]. *Magnetic fields* are force fields generated by charges in motion and that exert a force only on moving charges. The most simple example is represented by the field generated by a flow of electrons through a copper cable. If  $I$  is the current through the conductor, the law of *Biot-Savart* allows to determine the magnetic



**Fig. 3.1** A flow of electric charges walks through a surface  $S$  with local velocity  $\mathbf{v}$



**Fig. 3.2** Geometry of a current loop

(induction) field  $\mathbf{B} = \mathbf{B}(\mathbf{r})$  at any position  $\mathbf{r}$  as a function of the cable geometry (Fig. 3.2):

$$\mathbf{B}(\mathbf{r}) = -\frac{\mu_0 I}{4\pi} \int_C \frac{(\mathbf{r} - \mathbf{q}) \times d\mathbf{q}}{\|\mathbf{r} - \mathbf{q}\|^3} \quad (3.4)$$

where the line integral is calculated along the circuit  $C$  and  $\mu_0$  is the *magnetic permeability in the vacuum*:  $\mu_0 = 4\pi \times 10^{-7}$  H/m. In the SI, the unit of  $\mathbf{B}$  is the Tesla [ $1 \text{ T} = 1 \text{ Vsm}^{-2}$ ]. The magnetic force exerted on a moving charge is determined by the equation of Lorentz:

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}) \quad (3.5)$$

Here,  $\mathbf{v}$  and  $q$  are respectively the velocity of the particle and its charge. Equation 3.5 shows that the magnetic force is always orthogonal both to the direction of motion of the particle and to the magnetic field vector  $\mathbf{B}$ . Furthermore,  $\mathbf{F} = \mathbf{0}$  when the particle moves in the same direction of the field. Thus, the equation of motion for a particle that is moving in a magnetic field assumes the form:

$$m \frac{d\mathbf{v}}{dt} = q\mathbf{v} \times \mathbf{B} \quad (3.6)$$

An interesting class of solutions for Eq. (1.6) can be obtained easily for a homogeneous magnetic field, for example when  $\mathbf{B} = B\mathbf{k}$  ( $\mathbf{k}$  being the base vector in the direction  $z$ ). In this case (3.6) assumes the following simple form:

$$\begin{cases} \dot{v}_x = \Omega v_y \\ \dot{v}_y = -\Omega v_x \\ \dot{v}_z = 0 \end{cases} \quad (3.7)$$

where the quantity:

$$\Omega \equiv \frac{qB}{m} \quad (3.8)$$

is called *cyclotron frequency*. A solution to this system of differential equations, with the initial condition  $\mathbf{v}_0 = \mathbf{v}(0)$ , can be obtained in a few steps. The third equation implies that  $v_z$  is constant:

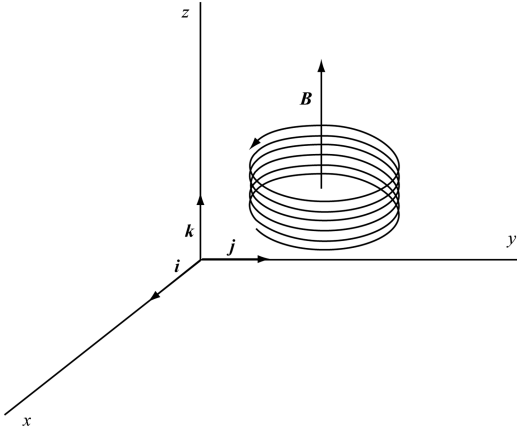
$$v_z(t) = v_{0z} \quad (3.9)$$

Furthermore, taking the derivative of the first two equations we easily obtain a separation of the equations:

$$\begin{cases} \ddot{v}_x = -\Omega^2 v_x \\ \ddot{v}_y = -\Omega^2 v_y \end{cases} \quad (3.10)$$

Therefore, we obtain:

$$\begin{cases} v_x = A \sin(\Omega t + \varphi) \\ v_y = A \cos(\Omega t + \varphi) \end{cases} \quad (3.11)$$



**Fig. 3.3** An electric charge in motion through an homogeneous magnetic field follows a helical trajectory

Applying the initial conditions we can solve for the amplitude,  $A$ , and the phase parameter,  $\varphi$ :

$$A = \sqrt{v_{x0}^2 + v_{y0}^2} ; \quad \varphi = \arctan\left(\frac{v_{x0}}{v_{y0}}\right) \quad (3.12)$$

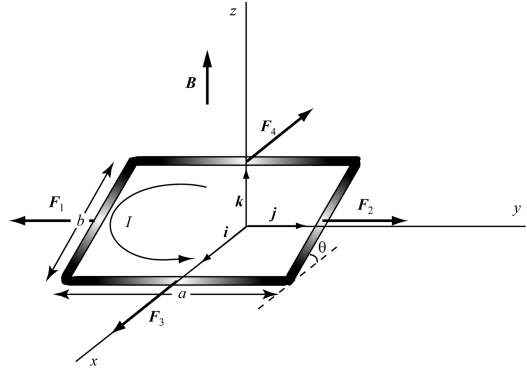
Therefore, the particle will follow a helical trajectory with radius  $A/\Omega$  (Fig. 3.3). This important example explains why charged particles that move in the Earth's magnetic field travel in spiral paths about the geomagnetic field lines.

### 3.2 Magnetic Moments

Let us consider now a rectangular coil having sides  $a$  and  $b$ , placed in a homogeneous magnetic field  $\mathbf{B} = B\mathbf{k}$  (Fig. 3.4). Let us also assume that the sides having length  $a$  are aligned with direction  $y$ , and that the coil is inclined by an angle  $\theta$  with respect to the horizontal plane. If  $I$  is the current flowing in clockwise sense along the coil, then the magnetic force (Eq. 3.5) exerted on a coil element with orientation  $\mathbf{n}$  and length  $dl$  is given by:

$$d\mathbf{F} = dq(\mathbf{v} \times \mathbf{B}) = Idl(\mathbf{n} \times \mathbf{B}) \quad (3.13)$$

We can integrate this formula to calculate the forces exerted on the coil. The forces on the inclined sides,  $\mathbf{F}_1$  and  $\mathbf{F}_2$ , have the same



**Fig. 3.4** A coil crossed by a current  $I$  is subject to a net torque when it is placed in a magnetic field

magnitude and are opposite each other (Fig. 3.4). Conversely, forces  $\mathbf{F}_3$  and  $\mathbf{F}_4$ , which are applied to the sides aligned with the  $y$  direction, also have the same magnitude, but are not lined up. Therefore, they generate a force couple that tends to align the coil with the  $xy$  plane. Since  $\mathbf{F}_3 = IaB\mathbf{i}$ , and  $\mathbf{F}_4 = -IaB\mathbf{i}$ , then the total torque is given by:

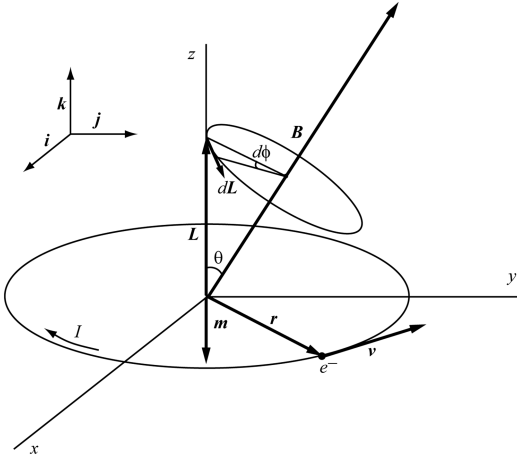
$$\mathbf{N} = \frac{b}{2}F_3 \sin\theta \mathbf{i} + \frac{b}{2}F_4 \sin\theta \mathbf{i} = SIB \sin\theta \mathbf{i} \quad (3.14)$$

where  $S = ab$ . Assuming that  $\mathbf{S}$  is a vector orthogonal to the coil and having magnitude  $S$ , we can introduce a vector  $\mathbf{m}$  such that:  $\mathbf{m} = I\mathbf{S}$ . In this instance, (3.14) can be rewritten as follows:

$$\mathbf{N} = \mathbf{m} \times \mathbf{B} \quad (3.15)$$

The vector  $\mathbf{m}$  is called *magnetic moment* of the coil. Equation 3.15 has general validity and does not depend from the specific geometry of the coil. It implies that the field exerts a torque that tends to align the magnetic moment to the external field. Let us consider now the example of an electron that is moving along a circular orbit with radius  $r$  with velocity  $v$  (Fig. 3.5). If  $T$  is the orbital period, then the quantity of charge that crosses an arbitrary surface orthogonal to the orbit in a unit time is:

$$I = \frac{dq}{dt} = \frac{e}{T} = \frac{e}{2\pi r/v} = \frac{ev}{2\pi r} \quad (3.16)$$



**Fig. 3.5** The counterclockwise rotation of an electron about the  $z$  axis is equivalent to a coil crossed by a current flowing in the opposite direction. If the direction of the external magnetic field  $\mathbf{B}$  does not coincide with that of the magnetic moment associated with the current loop,  $\mathbf{m}$ , then the orbit is perturbed and the angular momentum  $\mathbf{L}$  will be subject to a precession about the direction of  $\mathbf{B}$

Thus, the orbital motion of the electron generates a current loop, whose strength of interaction with the external magnetic field is described by the magnetic moment  $\mathbf{m}$ . Since the electron has negative charge, in this case the vector  $\mathbf{m}$  will be opposite to the angular momentum  $\mathbf{L}$  associated with the orbit. Therefore, if  $m_e$  is the electron mass and  $\mathbf{L} = \mathbf{r} \times m_e \mathbf{v}$  is its angular momentum, then:

$$\begin{aligned} \mathbf{m} &= -\pi r^2 I \hat{\mathbf{L}} = -\frac{1}{2} e v r \hat{\mathbf{L}} = -\frac{1}{2} e \mathbf{r} \times \mathbf{v} \\ &= -\frac{e}{2m_e} \mathbf{L} \end{aligned} \quad (3.17)$$

The ratio  $-e/2m_e$  is called *classic gyromagnetic ratio of the electron*. Combining (3.17) with (3.15), we obtain a simple expression for the torque exerted on the electron orbit:

$$\mathbf{N} = \frac{d\mathbf{L}}{dt} = -\frac{e}{2m_e} \mathbf{L} \times \mathbf{B} \quad (3.18)$$

Therefore,  $d\mathbf{L} \perp \mathbf{L}$ . This implies that the angular momentum  $\mathbf{L}$ , and consequently the magnetic moment  $\mathbf{m}$ , precesses about the vector  $\mathbf{B}$ , as shown in Fig. 3.5.

To calculate the angular frequency of precession, we observe that:

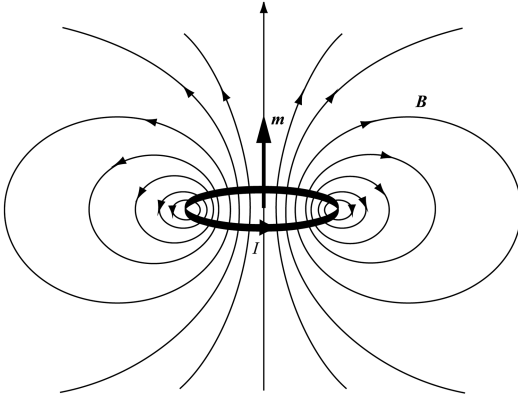
$$dL = L \sin \theta d\phi \quad (3.19)$$

where  $\theta$  is the angle between  $\mathbf{L}$  and  $\mathbf{B}$  and  $d\phi$  is the infinitesimal angle of precession. Combining (3.19) with (3.18) we obtain:

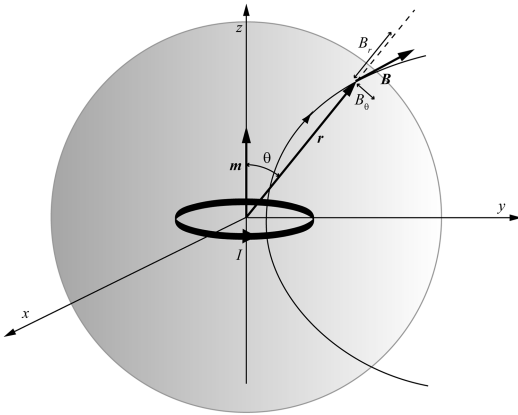
$$\omega_L = \frac{d\phi}{dt} = \frac{eB}{2m_e} \quad (3.20)$$

This quantity, which is independent from the angle  $\theta$ , is known as the *Larmor angular frequency*. The precession of the magnetic and angular moments around the magnetic field vector is the basic principle underlying the construction of *proton precession magnetometers*, which form an important class of *scalar magnetometers*. These devices furnish measurements of the *magnitude* of the Earth's magnetic field. We shall see that these are the raw data of plate kinematics. The operation of these instruments is based on the proportionality between Larmor frequency and magnitude of the external field predicted by Eq. (3.20), so that a measurement of frequency can be converted to a measurement of field intensity.

We have seen that current loops are subject to the action of external magnetic fields. However, a current loop generates itself a special kind of magnetic field that is called *dipole field*, whose structure is illustrated in Fig. 3.6. We note that in this case the field lines are arranged symmetrically with respect to the direction of  $\mathbf{m}$  (dipole axis). The dipole field describes with good approximation the present day Earth's magnetic field, which will be considered in the next chapter, and with excellent approximation the *paleomagnetic fields* (Chap. 6). The current loops that generate these fields are referred to as *magnetic dipoles*. To obtain an expression for  $\mathbf{B}$ , we choose a coordinate system with the  $z$ -axis aligned in the direction of  $\mathbf{m}$  and with the origin placed at the centre of the current loop. At great distance, the magnetic field generated by this magnetic dipole has the following approximate expression (e.g., Panofsky and Phillips 2005):



**Fig. 3.6** Force lines of the magnetic dipole field  $\mathbf{B}$  generated by a small current loop



**Fig. 3.7** Components of the dipole field generated by a magnetic dipole directed as the  $z$ -axis

$$\mathbf{B}(\mathbf{r}) \cong \frac{\mu_0}{4\pi} \left( 3 \frac{\mathbf{m} \cdot \mathbf{r}}{r^5} \mathbf{r} - \frac{\mathbf{m}}{r^3} \right) \quad (3.21)$$

This expression shows that at any location  $\mathbf{r}$  the field vector  $\mathbf{B}(\mathbf{r})$  can be decomposed in two orthogonal components, one directed radially as  $\mathbf{r}$ , and a component that is tangent to the sphere of radius  $r$  (Fig. 3.7). They are, respectively,

$$B_r(\mathbf{r}) = \frac{\mu_0 m \cos \theta}{2\pi r^3} \quad (3.22)$$

$$B_\theta(\mathbf{r}) = \frac{\mu_0 m \sin \theta}{4\pi r^3} \quad (3.23)$$

where  $\theta$  is the angle between  $\mathbf{r}$  and  $\mathbf{m}$ . In the case of the geomagnetic field, this angle coincides

with the magnetic colatitude. These expressions show that the magnitude of  $\mathbf{B}$  depends from the inverse cube of the distance from the origin:

$$B(\mathbf{r}) = \frac{\mu_0 m}{4\pi r^3} (3\cos^2\theta + 1)^{1/2} \quad (3.24)$$

We note that at any fixed distance from the origin, the magnitude of  $\mathbf{B}$  is maximum at the poles ( $\theta = 0$  or  $\theta = \pi$ ) and attains its minimum along the magnetic equator ( $\theta = \pi/2$ ). Equations 3.22 and 3.23 can be used to determine the components of the Earth's magnetic field in a local reference frame (see Sect. 2.3). In the next chapter, we shall see discuss the application of the magnetic dipole model to the representation of the geomagnetic field. For the moment, it is sufficient to say that in a local coordinate system the term (3.22) corresponds to the vertical component of the field,  $Z$ , while the horizontal component,  $H$ , will be obtained by (3.23).

### 3.3 Maxwell's Equations for the Magnetic Field

The four *Maxwell's equations* of classical Electrodynamics express, in a concise form, all the basic features of the electromagnetic fields and their relation with the electric and magnetic sources. Together with Lorentz's equation, they furnish a complete description about the origin of the electromagnetic fields, their interaction with charged particles, and their evolution in time. Two of these equations describe the sources of magnetic fields. The first of them is a differential form of the *law of Gauss*:

$$\nabla \cdot \mathbf{B} = 0 \quad (3.25)$$

This equation simply states that there are no magnetic charges in the physical world. It also implies that a vector field exists,  $\mathbf{A} = \mathbf{A}(\mathbf{r})$ , such that:

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (3.26)$$

The field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$  is called *vector potential*. The second of Maxwell's equations devoted to

the magnetic fields is known as *Ampere's law*. It relates the spatial structure of the magnetic field  $\mathbf{B}$  to the geometry of the electric currents and to the temporal variations of the electric field.

Assuming that the latter does not change in time, the differential form of Ampere's law can be written as follows:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} \quad (3.27)$$

If  $\mathbf{j} = \mathbf{0}$  in a region  $\mathbf{R}$ , hence in absence of currents, we have that  $\nabla \times \mathbf{B} = \mathbf{0}$  in  $\mathbf{R}$ . Therefore, for any point in a region where the current density is zero, there exists a scalar field  $V = V(\mathbf{r})$  such that:

$$\mathbf{B} = -\nabla V \quad (3.28)$$

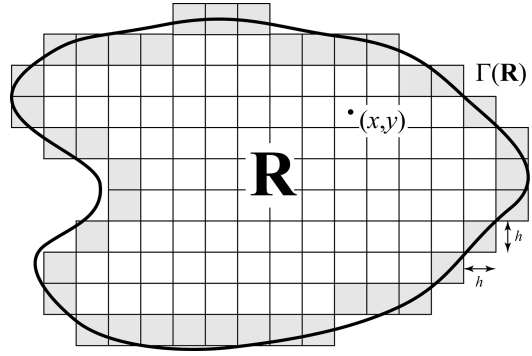
The field  $V$  is called *scalar magnetic potential*. If we combine (3.28) with (3.25), we obtain the following fundamental equation, which is valid in current-free regions:

$$\nabla^2 V = 0 \quad (3.29)$$

This second-order differential equation is called *Laplace's equation*, and its solutions are *harmonic functions*. It represents the fundamental tool for the study of the Earth's magnetic and gravity fields, which form a substantial portion of a branch of geosciences known as *potential fields geophysics*. In Chap. 4, we shall prove that a solution of this equation in a region  $\mathbf{R}$  can be found if *boundary conditions* have been assigned along the frontier  $\Gamma(\mathbf{R})$  of  $\mathbf{R}$ , that is, if the values of  $V$  are known on the closed surface  $\Gamma(\mathbf{R})$ . Once a solution  $V = V(\mathbf{r})$  has been determined in  $\mathbf{R}$ , we can use (3.28) to determine uniquely the vector field  $\mathbf{B} = \mathbf{B}(\mathbf{r})$  in that region.

To understand the basic properties of the harmonic functions, let us consider now the example of a two-dimensional potential, depending only from  $x$  ed  $y$ . In this instance, Laplace's equation assumes the form:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \quad (3.30)$$



**Fig. 3.8** Discretization of Laplace's equation. The region  $\mathbf{R}$  is divided into a set of squared grid cells of dimension  $h$ . It is assumed that the values of the potential  $V$  are known along the boundary  $\Gamma(\mathbf{R})$ , which is represented by the grayed cells. The equation is solved calculating iteratively the values of  $V$  at each point  $(x, y)$  in the interior of  $\mathbf{R}$  (white cells)

A numerical solution to this equation can be found by discretization of the domain  $\mathbf{R}$  through squared grid cells. In this approach, we subdivide the region  $\mathbf{R}$  in small grid cells of dimension  $h$  (Fig. 3.8) and search for an approximate solution at the centre of each cell. Let  $V = V(x, y)$  be a solution of (3.30).

If we expand  $V$  in a Taylor series with respect to variable  $x$ , it results:

$$\begin{cases} V(x+h, y) = V(x, y) + h \frac{\partial V}{\partial x} + \frac{1}{2} h^2 \frac{\partial^2 V}{\partial x^2} \\ \quad + \frac{1}{6} h^3 \frac{\partial^3 V}{\partial x^3} + \dots \\ V(x-h, y) = V(x, y) - h \frac{\partial V}{\partial x} + \frac{1}{2} h^2 \frac{\partial^2 V}{\partial x^2} \\ \quad - \frac{1}{6} h^3 \frac{\partial^3 V}{\partial x^3} + \dots \end{cases} \quad (3.31)$$

Summing these two expressions, it results:

$$V(x+h, y) + V(x-h, y) = 2V(x, y) + h^2 \frac{\partial^2 V}{\partial x^2} + O(h^4) \quad (3.32)$$

An analog expression can be found by expanding  $V$  in a Taylor series with respect to variable  $y$ :

$$V(x, y + h) + V(x, y - h) = 2V(x, y) + h^2 \frac{\partial^2 V}{\partial y^2} + O(h^4) \quad (3.33)$$

Now we add expressions (3.32) and (3.33). Applying the equation of Laplace (3.30), we have:

$$V(x, y) = \frac{1}{4} [V(x + h, y) + V(x - h, y) + V(x, y + h) + V(x, y - h)] + O(h^4) \quad (3.34)$$

This expression shows that the potential at any point  $(x, y)$  equals, up to high-order terms, the average of  $V$  over a neighbor of  $(x, y)$ . This is a general property of the harmonic functions, which will be proved rigorously in Chap. 4. Therefore,  $V$  cannot have maxima or minima within the region  $\mathbf{R}$ . The numerical approach also allows to calculate easily the values of  $V$  in  $\mathbf{R}$ , starting from the boundary values. This algorithm assigns the initial value of each grid cell internal to  $\mathbf{R}$  to an arbitrary constant value  $V = V_0$ , while the points along the frontier of  $\mathbf{R}$ ,  $(x, y) \in \Gamma(\mathbf{R})$ , are set through the boundary conditions (Fig. 3.8). At the next step, for each point  $(x, y) \in \mathbf{R} - \Gamma(\mathbf{R})$  we calculate iteratively more precise values of the field by the following assignment:

$$V_{n+1}(x, y) = \frac{1}{4} [V_n(x + h, y) + V_n(x - h, y) + V_n(x, y + h) + V_n(x, y - h)] \quad (3.35)$$

The algorithm terminates when the standard deviation of the values over the neighbor of each point, with respect to the central value, falls below an assigned threshold. To reduce the number of iterations, it is possible to use a recurrence formula more sophisticated than (3.35), which includes an *over-relaxation factor*  $0 \leq \lambda < 1$ . In this approach, the update expression (3.35) is modified by the addition of an

over-relaxation terms depending from the weighting factor  $\lambda$  and the deviation of the neighbor average from the central value:

$$V_{n+1}(x, y) = V_n(x, y) + (\lambda + 1) \Delta_n(x, y) \quad (3.36)$$

where the deviation factor  $\Delta_n(x, y)$  at  $(x, y)$  at step  $n$  is given by:

$$\Delta_n(x, y) = \frac{1}{4} [V_n(x + h, y) + V_n(x - h, y) + V_n(x, y + h) + V_n(x, y - h)] - V_n(x, y) \quad (3.37)$$

This procedure, which can be easily generalized to the three-dimensional space, does not require an explicit analytic solution of Laplace's equation. It is a practical method to find the values of  $V$  in  $\mathbf{R}$  when it is not possible to determine an exact solution. However, in the next chapter we shall consider a general class of exact solutions of this equation that are commonly used to represent the geomagnetic potential. Then, a complete analytical procedure for finding the solutions of Laplace's equation in spherical coordinates will be described. Finally, in Chap. 14 we shall see that also the Earth's gravity potential is a harmonic function outside the Earth's surface.

---

## 3.4 Magnetization

All ordinary materials, when placed in a magnetic field, acquire a *magnetization*, which is a manifestation of the presence of a large number of magnetic dipoles at atomic scale. These dipoles result from microscopic currents, associated with the motion of electrons within the atoms and the with the intrinsic magnetic moments of the elementary particles (spin). The magnetic field generated by these microscopic sources adds to that produced by macroscopic currents (flows of electrons in conductors, motion of electrically charged fluids, etc.). For many substances, the net magnetic moment of the individual atoms

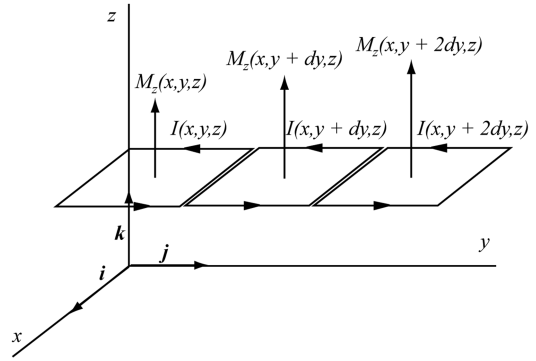
is zero in absence of external field, whereas small currents are induced when a magnetic field is applied. These currents are associated with magnetic moments that *oppose* the external field, and this kind of response of the matter is called *diamagnetism*. Other substances do have atoms with permanent magnetic moments, but these vectors average to zero in absence of external field. However, when we place these substances in a magnetic field, the permanent magnetic moments have the tendency to align with the external field, so that the diamagnetic response is flanked by an additional form of magnetism, which enforces the applied field and has greater magnitude. This response, which is called *paramagnetism*, is quite weak because the aligning forces are small compared with the forces associated with thermal agitation, which tend to destroy the alignment. Therefore, this kind of magnetism strongly depends from temperature. The third kind of magnetism of matter is called *ferromagnetism* and has the greatest importance in Earth Sciences. In this instance, the magnetization induced by external fields has very strong magnitude and adds a significant contribution to the total magnetic field. This behaviour can be observed only in some crystalline solids and originates from the combination of a well-known quantum mechanics phenomenon and electrostatic interactions at atomic scale. The total magnetization of a body having volume  $V$  is defined as the average magnetic moment per unit volume:

$$\mathbf{M} = \frac{1}{V} \sum_i \mathbf{m}_i \text{ [A/m]} \quad (3.38)$$

where the vectors  $\mathbf{m}_i$  are magnetic moments at atomic scale. To understand the effect of magnetization, let us consider a body composed by a continuous distribution of magnetic moments, such that the magnetic moment associated with a volume element  $dV$  at position  $\mathbf{r}$  is:

$$d\mathbf{m}(\mathbf{r}) = \mathbf{M}(\mathbf{r})dV \quad (3.39)$$

We are going to show that this body generates a magnetic field, which is equivalent to the field



**Fig. 3.9** Equivalent distribution of atomic dipoles associated with the component  $z$  of magnetization of a body

that would be generated by a macroscopic current  $\mathbf{j}_m$  that we shall call *magnetization current*. This current is calculated as follows:

$$\mathbf{j}_m = \nabla \times \mathbf{M} \quad (3.40)$$

A comparison of this equation with Ampere's law (3.27) shows that the equivalent field is directed as  $\mathbf{M}$  and has magnitude  $\mu_0 \mathbf{M}$ . The equivalence between magnetized materials and currents was observed for the first time by Ampere. The expression (3.40) for  $\mathbf{j}_m$  implies that the current of magnetization goes to zero in regions of homogeneous magnetization, because in this case the derivatives of  $M_x$ ,  $M_y$ , and  $M_z$  in (3.40) are all zero. Therefore, we attain the fundamental result that the current  $\mathbf{j}_m$  is the result of a non-uniform magnetization at atomic scale. To prove (3.40), let us consider a thin slice of the body, of thickness  $dz$ , parallel to the plane  $xy$  at elevation  $z$  (Fig. 3.9).

By (3.39), at any location  $(x, y, z)$  the presence of a non-zero component  $M_z(x, y, z)$  of the local magnetization vector implies the existence of an infinitesimal current loop, having area  $dxdy$ , which has magnetic moment  $dm_z = M_z(x, y, z)dxdydz$ . Then, the microscopic current flowing in the coil will be given by:

$$\begin{aligned} I(x, y, z) &= \frac{dm_z}{dxdy} = \frac{M_z(x, y, z)dxdydz}{dxdy} \\ &= M_z(x, y, z)dz \end{aligned} \quad (3.41)$$

Similarly, at position  $(x, y + dy, z)$  we have:

$$\begin{aligned} I(x, y + dy, z) &= M_z(x, y + dy, z) dz \\ &= \left[ M_z(x, y, z) + \frac{\partial M_z}{\partial y} dy \right] dz \end{aligned} \quad (3.42)$$

Therefore, it is evident from Fig. 3.9 that a first contribution to the net macroscopic current flowing in direction  $x$  is:

$$\delta I_1^{(x)} = -\frac{\partial M_z}{\partial y} dy dz \quad (3.43)$$

Another contribution to the current flowing in the  $x$  direction can be obtained considering a thin slice of thickness  $dy$ , parallel to plane  $xz$  at elevation  $y$ . At position  $(x, y, z)$ , the existence of a non-zero component of magnetization  $M_y(x, y, z)$  is equivalent to the existence of an infinitesimal rectangular coil having area  $dx dz$  and magnetic moment  $dm_y = M_y(x, y, z) dx dy dz$ . By the same reasoning as before, we see that an additional contribution to the macroscopic current in the  $x$  direction will be:

$$\delta I_2^{(x)} = \frac{\partial M_y}{\partial z} dy dz \quad (3.44)$$

Therefore, the total macroscopic current in the  $x$  direction is given by:

$$\begin{aligned} \delta I^{(x)} &= \delta I_1^{(x)} + \delta I_2^{(x)} = \left( \frac{\partial M_y}{\partial z} - \frac{\partial M_z}{\partial y} \right) dy dz \\ &\equiv j_{mx} dy dz \end{aligned} \quad (3.45)$$

This proves that the components of the curl of  $\mathbf{M}$  coincide with the components of a net macroscopic current density, resulting from the microscopic currents that flow within the body.

The magnetization acquired by a body in presence of an external magnetic field is called *induced magnetization* and is in general a function of the applied field  $\mathbf{B}_{ext}$ . The relation between  $\mathbf{M}$  and  $\mathbf{B}_{ext}$  is different among the three classes of magnetic behavior of matter. We have mentioned that all the ordinary substances acquire a weak magnetization that opposes  $\mathbf{B}_{ext}$  (diamagnetism).

Now we shall prove that this is a consequence of the Larmor precession discussed in Sect. 3.2. We have shown in Sect. 3.2 that the electron orbits are magnetic dipoles, with a definite magnetic moment that depends from the angular momentum (Eq. 3.17). In normal conditions, the atoms of a diamagnetic substance have a net magnetic moment equals zero, because the various magnetic moments, associated with orbits, electron spins, etc., balance out. However, when we apply an external magnetic field, the precession of an electron orbit about the field direction is equivalent to an additional microscopic current loop having radius  $a$  and an intensity that according to (3.16) is given by:

$$I_L = -\frac{e\omega_L}{2\pi} \quad (3.46)$$

This current is associated with an extra angular momentum, resulting from the twist of the electron orbit, which is parallel to the external field:

$$\delta \mathbf{L} = m_e a v_L \hat{\mathbf{B}}_{ext} = m_e a^2 \omega_L \hat{\mathbf{B}}_{ext} = \frac{1}{2} e a^2 \mathbf{B}_{ext} \quad (3.47)$$

where we have used (3.20). It should be noted that  $\delta \mathbf{L}$  represents an additional vector of angular momentum and *not* the variation of  $\mathbf{L}$  in a small time interval, which is indicated as  $d\mathbf{L}$  (see Fig. 3.5). According to (3.17), this extra angular momentum must be associated with an additional magnetic moment,  $\delta \mathbf{m}$ , given by:

$$\delta \mathbf{m} = -\frac{e}{2m_e} \delta \mathbf{L} = -\frac{e^2 a^2}{4m_e} \mathbf{B}_{ext} \quad (3.48)$$

Thus, the extra magnetic moment  $\delta \mathbf{m}$  opposes the external field  $\mathbf{B}_{ext}$ . Regarding the quantity  $a^2$  that appears in (3.48), it can be shown on the basis of quantum mechanics considerations that it represents the averaged square distance of the electron from the axis of  $\mathbf{B}_{ext}$  (e.g., Feynman et al. 2006). If the external field is aligned with the  $z$ -axis and  $\mathbf{r} \equiv (x, y, z)$  is the instantaneous position of the electron, then  $a^2 = \langle x^2 \rangle + \langle y^2 \rangle$ . For a spherically symmetric atom,  $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle = \langle r^2 \rangle$ , where  $r$  is the distance of the electron from

the nucleus. Therefore,  $a^2 = (2/3)\langle r^2 \rangle$  and (3.48) can be rewritten as follows:

$$\delta \mathbf{m} = -\frac{e^2 \langle r^2 \rangle}{6m_e} \mathbf{B}_{ext} \quad (3.49)$$

The total magnetization can be calculated using the definition (3.38). If  $N$  is the total number of atoms and  $Z$  is the average number of electrons per atom, we have:

$$\begin{aligned} \mathbf{M} &= \frac{1}{V} \sum_i m_i = -\frac{NZe^2}{6m_e V} \langle r^2 \rangle \mathbf{B}_{ext} \\ &= -\frac{ne^2}{6m_e} \langle r^2 \rangle \mathbf{B}_{ext} \end{aligned} \quad (3.50)$$

where  $n = NZ/V$  is the density of electrons. Therefore,  $\mathbf{M}$  has opposite direction with respect to  $\mathbf{B}_{ext}$ . This relation can be rewritten as follows:

$$\mathbf{M} = \frac{\chi}{\mu_0} \mathbf{B}_{ext} \quad (3.51)$$

The dimensionless quantity  $\chi$ , which is negative in the case of diamagnetic materials, is called *magnetic susceptibility* and its value depends from the substance. Common diamagnetic substances are water, wood, many organic materials and most metals.

Let us consider now the case of paramagnetic materials. In this instance, there are permanent magnetic moments at atomic scale, which are independent each other and in normal conditions are randomly oriented due to thermal agitation. Therefore, the total magnetization (3.38) is zero in absence of external field. However, if we apply a magnetic field, the magnetic moments tend to align to the external field, determining a non-zero net magnetization. The magnitude of the induced magnetization is determined by the *equation of Langevin*:

$$\frac{M}{M_s} = L(\xi) \equiv \coth(\xi) - \frac{1}{\xi} ; \quad \xi \equiv \frac{mB_{ext}}{kT} \quad (3.52)$$

where  $m$  is the permanent magnetic moment of the atoms,  $k$  is the Boltzmann constant,  $T$  is the

absolute temperature of the body, and  $M_s$  is the *saturation value* of magnetization. The function  $L$  in (3.52) is called *Langevin function* and tends asymptotically to unity, so that  $M \rightarrow M_s$  as  $\xi \rightarrow \infty$ . The existence of a saturation magnetization indicates that the induced magnetization cannot increase arbitrarily as we raise the magnitude of the external field or decrease the temperature. Given an atomic magnetic moment  $m$ , there is always a maximum value of magnetization that can be attained by the substance, which corresponds to a perfect alignment of the elementary atomic moments.

The shape of the Langevin function is illustrated in Fig. 3.10. If  $n$  indicates the density of magnetic dipoles, then:

$$M_s = mn \quad (3.53)$$

For  $M = M_s$  (that is, for  $L(\xi) = 1$ ) all the atomic dipoles are aligned with the external field  $\mathbf{B}_{ext}$ . Equation (3.52) implies that for  $mB_{ext} \sim 100kT$  the induced magnetization practically coincides with the saturation magnetization. Conversely, for  $mB_{ext} < kT$  the function  $L$  is approximately a linear function with slope  $\sim 1/3$ :

$$\frac{M}{M_s} \cong \frac{mB_{ext}}{3kT} \quad (3.54)$$

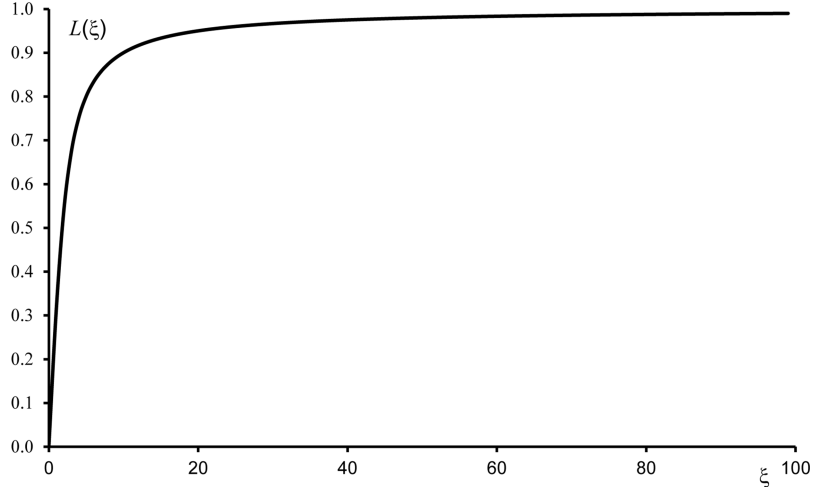
Substituting (3.53) gives:

$$M \cong \frac{nm^2 B_{ext}}{3kT} \quad (3.55)$$

This equation is *Curie's law*. It predicts that the magnetization is inversely proportional to the temperature. A comparison of this result with (3.51) shows that the magnetic susceptibility of a paramagnetic material is positive and assumes the value:

$$\chi = \frac{\mu_0 n m^2}{3kT} \quad (3.56)$$

We have seen that the magnetization of diamagnetic and paramagnetic materials is a vector having the same direction of the external field  $\mathbf{B}_{ext}$

**Fig. 3.10** The Langevin function

and a magnitude approximately proportional to the magnitude of this field. However, the magnetization current  $\mathbf{j}_m$  generates itself a magnetic field,  $\mu_0 \mathbf{M}$ , which contributes with  $\mathbf{B}_{ext}$  to form the *total field*  $\mathbf{B}$ . Therefore, Maxwell's Eq. (3.27) can be rewritten to show explicitly the contribution of both the macroscopic and microscopic currents:

$$\nabla \times \mathbf{B} = \mu_0 (\mathbf{j} + \nabla \times \mathbf{M}) \quad (3.57)$$

where  $\mathbf{j}$  represents the macroscopic currents. This equation implies that:

$$\nabla \times (\mathbf{B} - \mu_0 \mathbf{M}) = \mu_0 \mathbf{j} \quad (3.58)$$

Let us introduce now a new vector field, which is the *magnetic field intensity*:

$$\mathbf{H} \equiv \frac{1}{\mu_0} (\mathbf{B} - \mu_0 \mathbf{M}) = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \quad (3.59)$$

Note that  $\mathbf{H}$  has the same units of  $\mathbf{M}$ , namely A/m. Using this field in the (3.58) we obtain the following simple form for the Maxwell Eq. (3.27):

$$\nabla \times \mathbf{H} = \mathbf{j} \quad (3.60)$$

Now, considering that the field  $\mathbf{H}$  is a way to represent the *external* field, and the corresponding macroscopic currents, we can write:

$$\mathbf{M} = \chi \mathbf{H} \quad (3.61)$$

Therefore, the total magnetic field  $\mathbf{B}$  assumes the following simple expression:

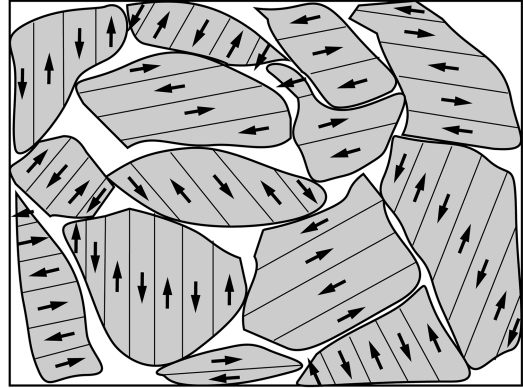
$$\begin{aligned} \mathbf{B} &= \mathbf{B}_{ext} + \mu_0 \mathbf{M} = \mu_0 (\mathbf{H} + \mathbf{M}) \\ &= \mu_0 (1 + \chi) \mathbf{H} \equiv \mu \mathbf{H} \end{aligned} \quad (3.62)$$

The quantity  $\mu$  is called *absolute magnetic permeability*. For some substances the law (3.62), which establishes a simple relation of proportionality between magnetic field intensity and total field, is not valid. In this case, a more complex tensor expression describes the relation between the two fields and we say that the material has *magnetic anisotropy*. For this class of substances,  $\mathbf{B}$  and  $\mathbf{H}$  are *not* parallel, and a field applied in the  $x$ -direction determines an induced magnetization also in the  $y$  and  $z$  directions. Therefore, in general the magnetic permeability (and the susceptibility) is described by a  $3 \times 3$  tensor. The magnetic susceptibility of paramagnetic materials is on average from 50 to 150 times greater, in absolute value, than the susceptibility of diamagnetic materials. However, for both classes of substances the removal of the external field determines the disappearance of the induced magnetization. In the next section, we shall consider a third class of materials, the ferromagnetic solids, which retain some magnetization even in absence of external magnetic fields.

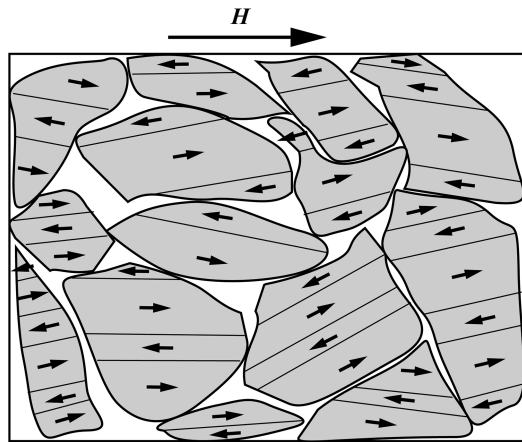
### 3.5 Magnetic Properties of Rocks

It is not exaggerated to say that the ferromagnetic properties of some crystalline solids are at the base of the plate tectonics revolution during the 1960s. In ferromagnetic materials, the permanent magnetic moments of neighbor atoms are not independent as a consequence of electrostatic interactions and quantum phenomena, which determine their alignment and a magnetization that can be several orders of magnitude greater than the paramagnetic response induced by the same external field. Ultimately, such a large magnetization is a consequence of the tendency of unpaired electrons of neighbor atoms to avoid sharing of their orbits, in which case they would acquire opposing spins, and to align their intrinsic magnetic moments.

This interaction exists independently from the application of an external field, and determines a spontaneous alignment of the spins through distinct regions of each crystal called *magnetic domains* (Fig. 3.11). Therefore, a single magnetic domain has a net non-zero *spontaneous magnetization* even in absence of external field. The reason for which a ferromagnetic substance does not reveal, in normal conditions, any apparent magnetization is illustrated in Fig. 3.11. Each crystal grain has a preferred direction of magnetization, and is divided into a series of magnetic domains whose spontaneous magnetization is alternate and parallel to this direction. At macroscopic scale, these preferred directions are randomly distributed. Therefore, the net magnetization is zero. When we apply an external field  $H$ , the domain walls start moving to favor the growth of domains with a direction of magnetization close to that of the applied field and the simultaneous reduction of size for the other domains. For small values of  $H$ , this process is reversible, so that if we remove the external field the magnetization returns to zero. If the applied field increases sufficiently, the domain walls are progressively destroyed, as illustrated in Fig. 3.12, until the total magnetization reaches a saturation value, which corresponds to a complete alignment of

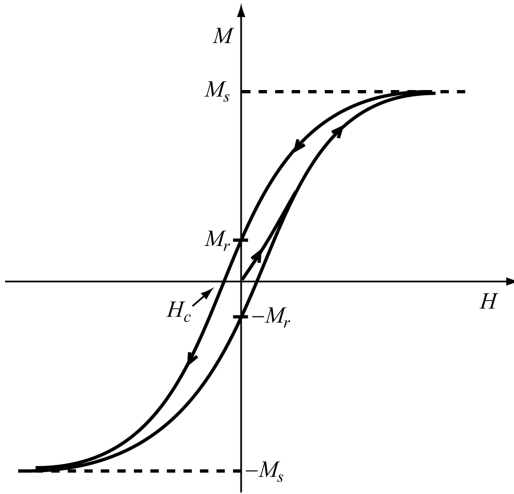


**Fig. 3.11** Arrangement of mineral grains (in gray) and magnetic domains (regions separated by thin lines) in an unmagnetized polycrystalline solid. Arrows are spin directions



**Fig. 3.12** Arrangement of mineral grains and magnetic domains in a magnetized polycrystalline solid close to saturation

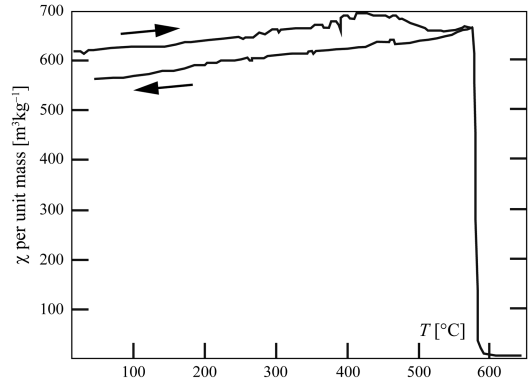
the spins. Now let us imagine to reduce progressively the intensity of the external field. In this instance, the domains start reforming, but the progression is not exactly the reverse of the previous one. In fact, a key feature of the process described above is represented by the relevant loss of internal energy in so far as the external field increases and the mineral grains rearrange their magnetic domains. This is a consequence of the presence of crystal defects, which prevents a continuous adaptation of the domain geometry as the magnitude of the external field increases. The jerky rearrangement of the domains wall geometry determines the formation of eddy currents that



**Fig. 3.13** Magnetization curve of a ferromagnetic mineral, showing the characteristic hysteresis loop of these materials.  $M_s$  is the saturation magnetization,  $M_r$  is the remanent magnetization,  $H_c$  is the coercive value of the external field

dissipate internal energy during the progressive magnetization resulting from increasing external field intensity. Therefore, from some point onward the process is *irreversible*. For elevated values of the applied field, most of the work is done to rotate a little bit the spins and obtain a better alignment to the external field axis. In these conditions, there are small increments of magnetization even when the increase of magnitude of  $H$  is large. Consequently, the magnetization curve asymptotically converges to a saturation value in a similar way as paramagnetic materials (Fig. 3.10). However, in the case of ferromagnetic minerals the irreversibility of the process for large external fields determines the appearance of a hysteresis loop in the magnetization curve, as illustrated in Fig. 3.13. The presence of this loop implies that some residual magnetization persists even when we remove completely the external field. It is called *remnant magnetization* of the sample. Figure 3.13 shows that in order to remove completely this remanent magnetization we must apply an inverse field of magnitude  $H_c$ , which is called *coercive field*.

In general, a ferromagnetic solid that has experienced one or more phases of strong magnetization has a total magnetization  $M_T$  that is the sum



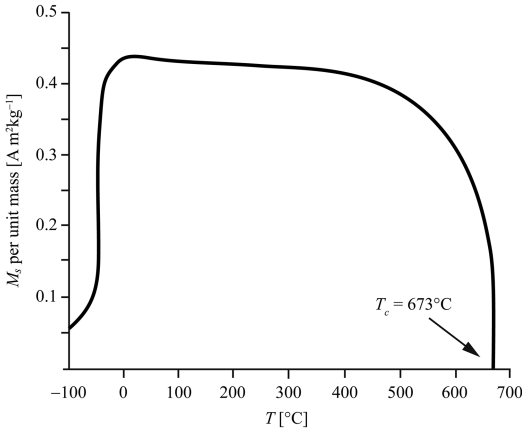
**Fig. 3.14** Magnetic susceptibility per unit mass,  $\chi$  [ $\text{m}^3/\text{kg}$ ] vs. temperature ( $^{\circ}\text{C}$ ) for magnetite ( $\text{Fe}_3\text{O}_4$ ). Arrows indicate the directions of heating and cooling (Redrawn from Harrison and Putnis 1996)

of two components: an induced magnetization  $M$ , which is proportional to  $H$  for small values of the external field, and a remanent magnetization  $M_r$ . Therefore,

$$M_T = M + M_r = \chi H + M_r \quad (3.63)$$

The magnetic susceptibility  $\chi$  is positive and depends upon temperature, just as in the case of paramagnetic substances, but in a more complicate way. Figure 3.14 illustrates the result of measurements of susceptibility on a pure magnetite ( $\text{Fe}_3\text{O}_4$ ) sample (Harrison and Putnis 1996). The sample was first heated, so that the temperature raised from room conditions to  $650^{\circ}\text{C}$ , then cooled back to room temperature at a rate of  $11^{\circ}\text{C}/\text{min}$ . It is probably not much surprising that also this plot shows a hysteresis loop, given the irreversibility of the process of magnetization. The temperature associated with the sharp drop of susceptibility at  $\sim 585^{\circ}\text{C}$  is called *Curie temperature*  $T_c$  and depends from the material. For  $T > T_c$  any ferromagnetic solid is converted into a paramagnetic material. This is a consequence of the fact that at high temperatures the thermal energy exceeds the spin coupling energy, determining a random arrangement of the magnetic moments.

For any ferromagnetic material, the saturation value of magnetization  $M_s$  is a decreasing function of temperature, as illustrated in Fig. 3.15.



**Fig. 3.15** Saturation magnetization vs  $T$  for hematite ( $\text{Fe}_2\text{O}_3$ ) (Redrawn from Hunt et al. (1995))

The curve  $M_s = M_s(T)$  shows a drop in correspondence with the Curie temperature  $T_c$ . Note that  $T_c \cong 680^\circ\text{C}$  for hematite.

The relative importance of the remanent magnetization with respect to the induced magnetization is measured by the *Koenigsberger ratio*:

$$Q \equiv \frac{M_r}{M} = \frac{M_r}{\chi H} \quad (3.64)$$

Therefore,  $Q \gg 1$  indicates that the remanent magnetization dominates, which is usually a desirable attribute in marine geophysics studies. Values of  $Q$  for several kinds of rocks are listed in Table 3.1. These values have been determined assuming that the magnetizing field  $H$  coincides with the Earth's magnetic field at the Earth's surface ( $H = 24 - 48 \text{ Am}^{-1}$ ). In general, mafic rocks have a larger spontaneous magnetization. For example, basalts are generally more magnetic than rhyolites, and gabbros are more magnetic than granites. Furthermore, the chemical composition being equal, extrusive rocks have larger remnant magnetization and lesser susceptibility than intrusive rocks. Finally, sedimentary and metamorphic rocks generally have low values of remnant magnetization and susceptibility.

An important aspect of the process of formation of magnetic domains is the relation between the number of domains within a mineral grain and the *grain size*. In general, the magnetization of

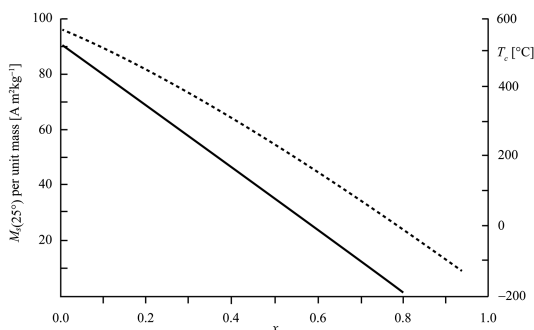
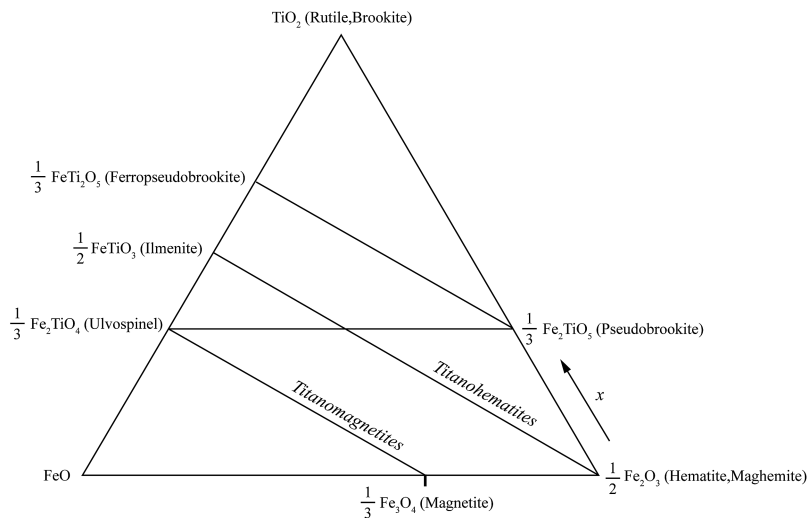
**Table 3.1** Koenigsberger ratios of selected rocks (From Hunt et al. 1995)

Roccia	Q
Marine sediments	5
Red sediments	1.6–6
Siltstone	0.02–2
Silty shale	5
Granite	0.1–28
Ganodiorite	0.1–0.2
Dolerite	2–3.5
Diabase	0.2–4
Gabbro	1–9.5
Oceanico Gabbro	0.1–58.4
Intrusive rocks	0.1–20
Volcanic rocks	30–50
Subaerial basalts	1–116
Oceanic basalts	1–160
Seamounts	8–57
Granulites	0.003–50

rocks arises from the presence of small grains of ferromagnetic minerals dispersed within a matrix of diamagnetic and paramagnetic minerals. The size of these grains strongly affects the magnetic behaviour of the rock, just because of its influence on the number of magnetic domains.

In geology, the most important magnetic minerals are undoubtedly iron-titanium ( $\text{FeTi}$ ) oxides, whose ternary diagram is shown in Fig. 3.16. Two classes of solid solutions are particularly important: the *titanomagnetites* series and the *titanohematites*, which represent primary phases of crystallization of igneous rocks (1–5 % vol). In titanomagnetites,  $\text{Ti}^{+4}$  substitutes  $\text{Fe}^{+3}$  as the Ti content increases. The crystal structure of these minerals is the spinel structure. The addition of Ti decreases progressively the saturation magnetization and the Curie temperature (Fig. 3.17), to the point that the ulvospinel is paramagnetic at ambient temperature. The general formula of titanomagnetites is:  $\text{Fe}_{3-x}\text{Ti}_x\text{O}_4$ , where  $x$  varies between zero (magnetite) and 1 (ulvospinel). The ionic substitution is  $2\text{Fe}^{+3} \rightarrow \text{Fe}^{+2} + \text{Ti}^{+4}$ . However, any titanomagnetite with  $x > 0.8$  is paramagnetic at ambient temperature. Regarding titanohematites, these minerals have a corundum structure and constitute a lesser

**Fig. 3.16** Ternary diagram of the iron-titanium oxides and their solid-solution series. Variable  $x$  is the composition parameter (Ti content). The ratio of ferric iron ( $\text{Fe}^{+3}$ ) to ferrous iron ( $\text{Fe}^{+2}$ ) increases from left to right. The Ti content increases from *bottom* to *top*. The diagram is normalized to one cation Fe, so that oxidation lines (increasing  $\text{Fe}^{3+}/\text{Fe}^{2+}$  ratio) are horizontal



**Fig. 3.17** Variation of room-temperature saturation magnetization (*solid line*) and Curie temperature (*dashed line*) with composition ( $x$  parameter) in the titanomagnetites. End members are magnetite ( $x = 0$ ) and ulvospinel ( $x = 1$ ) (Redrawn from Hunt et al. (1995))

portion of ferromagnetic minerals, although in highly silicic or highly oxidized igneous rocks hematite can give a major contribution to the rock ferromagnetism. Furthermore, hematite is the dominant or exclusive ferromagnetic mineral in red beds, which represent an important sedimentary source of paleomagnetic data. The general formula of titanohematites is  $\text{Fe}_{2-x}\text{Ti}_x\text{O}_3$ , where hematite ( $x = 0$ ) and ilmenite ( $x = 1$ ) are the end-members of the series. For a more in-depth discussion about the mineral physics of titanomagnetites and titanohematites, the reader is referred to Butler (1992).

Both titanomagnetites and titanohematites crystallize early, at a temperature of  $\sim 1,300^\circ\text{C}$ .

The distribution of grain sizes, hence the structure of the magnetic domains, depends strongly from the cooling rate. Titanomagnetites that form in rapidly cooling volcanic rocks (e.g., oceanic pillow lavas) are fine-grained, because they generally contain a significant fraction of grains with size of  $1\ \mu\text{m}$  or smaller. Conversely, the grain size is larger in the case of slowly cooled intrusive rocks, where it may exceed  $100\ \mu\text{m}$ . Paleomagnetists consider fine-grained ferromagnetic particles as the best magnetic recorders (e.g., Butler 1992). Therefore, volcanic rocks are generally preferred over intrusive rocks as targets for paleomagnetic studies.

The number of magnetic domains is an increasing function of the grain size. When the grains are sufficiently small, they will contain just one domain. These grains, which are referred to as *single-domain* (SD) grains, have magnetic properties that are dramatically different from those of *multi-domain* (MD) grains. The threshold grain diameter,  $d_0$ , below which we have only SD grains, depends essentially from the grain shape and from the saturation magnetization  $M_s$ . For hematite,  $d_0 = 15\ \mu\text{m}$ , thereby in most cases these minerals are SD. Conversely, only fine-grained magnetite is SD. In general, cubic magnetite particles must have a diameter  $d < 0.1\ \mu\text{m}$  to be SD, although elongated SD particles can have a length of up to  $1\ \mu\text{m}$ . Our interest into SD grains arises from their property of being

very efficient carriers of remnant magnetization. In fact, while the magnetization of MD grains tends to decay with time, that of SD particles is stable and can carry paleomagnetic information over a long time interval (up to billion years).

When the size of magnetite grains is between 1 and 10  $\mu\text{m}$ , the particles are called *pseudo-single-domain* (PSD) grains. They form a class of grains exhibiting intermediate values of the ratio  $M_r/M_s$  and the coercive field  $H_c$ . These grains contain a small number of domains and may carry a substantial and stable remnant magnetization. Grain size distributions of many igneous and sedimentary rocks have a peak within the magnetite PSD field, although they have only a small percentage of particles within the true SD field. Therefore, PSD grains can be important carriers of paleomagnetic information.

So far, we have been mainly concerned with the phenomenological aspect of ferromagnetism, after having mentioned that it arises from the coupling of electron spins between neighbor atoms. However, we have not yet explained the nature of this interaction, and the different ways through which the electron spins are coupled within a crystal lattice. Understanding this subject will require some basic quantum mechanics concepts, because ultimately the magnetism of matter cannot be explained by classical electrodynamics (see e.g., Feynman et al. 2006). At first glance, we could think that the alignment of magnetic moments in ferromagnetic materials results from their magnetic interaction. However, we are going to prove that this is not the case. Let us consider the angular momentum of an electron in an atom, which includes a component of orbital motion about the nucleus and one associated with the spinning about its own axis. These components also originate magnetic moments, one arising from the orbital motion,  $\mathbf{m}$ , and the other associated with spinning, which will be indicated as  $\mathbf{m}_S$ . A famous theorem, due to Bohr and van Leeuwen, shows that the orbital moment  $\mathbf{m}$  cannot produce a net magnetization, even in presence of external field  $\mathbf{B}_{\text{ext}}$ . Thus, let us focus on the spin moment  $\mathbf{m}_S$ . Let  $\mathbf{S}$  be the intrinsic angular momentum associated with the spinning

of the electron. Quantum mechanics shows that in this case the relation between magnetic moment and angular momentum is slightly different from (3.17), so that the ratio of  $\mathbf{m}_S$  to  $\mathbf{S}$  is twice the ratio of  $\mathbf{m}$  to  $\mathbf{L}$ :

$$\mathbf{m}_S = -\frac{e}{m_e} \mathbf{S} \quad (3.65)$$

Another difference with respect to classical mechanics is that in the case of atomic systems it is not possible to determine unambiguously the absolute direction of  $\mathbf{S}$  (or  $\mathbf{L}$ ) at any given time  $t$ . However, it is possible to show that at time  $t$  the *projection* of  $\mathbf{S}$  onto any arbitrary axis  $\mathbf{n}$  can assume only a finite number of values:

$$\mathbf{S} \cdot \mathbf{n} = (s - k) \hbar ; \quad k = 0, 1, \dots, 2s \quad (3.66)$$

where the quantity  $s$  is called *spin* of the electron (or, in general, of the particle) and the constant  $\hbar = h/2\pi = 1.054571726(47) \times 10^{-34}$  [Js] is the *reduced Planck's constant*. A similar relation can be written for the orbital momentum  $\mathbf{L}$ . Therefore, at any given time there are only  $2s + 1$  possible values for the component of  $\mathbf{S}$  along an arbitrary axis, for example the  $z$ -axis. Such discretization of the physical variables is one of the consequences of quantum mechanics. Equation (3.66) indicates that the maximum magnitude of the projection of  $\mathbf{S}$  onto  $\mathbf{n}$  is  $s\hbar$ , whereas classical mechanics would give  $S$ . Therefore, we would expect that  $s = S/\hbar$ . However, this conclusion would be wrong. The average squared magnitude of  $\mathbf{S}$  along an arbitrary axis  $\mathbf{n}$  is given by:

$$\begin{aligned} \langle (\mathbf{S} \cdot \mathbf{n})^2 \rangle &= \frac{1}{2s+1} \sum_{k=0}^{2s} (s-k)^2 \hbar^2 \\ &= \frac{\hbar^2}{2s+1} \left[ s^2 (2s+1) + \sum_{k=0}^{2s} k^2 - 2s^2 (2s+1) \right] \\ &= -\hbar^2 s^2 + \frac{\hbar^2}{2s+1} \sum_{k=0}^{2s} k^2 \\ &= -\hbar^2 s^2 + \frac{\hbar^2}{2s+1} \left[ \frac{1}{6} 2s (2s+1) (4s+1) \right] \\ &= \frac{1}{3} \hbar^2 s (s+1) \end{aligned} \quad (3.67)$$

This quantity does not depend from the selected axis of projection  $\mathbf{n}$ . In particular, we have:

$$\langle (\mathbf{S} \cdot \mathbf{i})^2 \rangle = \langle (\mathbf{S} \cdot \mathbf{j})^2 \rangle = \langle (\mathbf{S} \cdot \mathbf{k})^2 \rangle$$

Therefore,

$$\begin{aligned} \langle S^2 \rangle &= \langle S_x^2 + S_y^2 + S_z^2 \rangle = \langle (\mathbf{S} \cdot \mathbf{i})^2 \rangle \\ &+ \langle (\mathbf{S} \cdot \mathbf{j})^2 \rangle + \langle (\mathbf{S} \cdot \mathbf{k})^2 \rangle = 3 \langle (\mathbf{S} \cdot \mathbf{k})^2 \rangle \end{aligned}$$

Finally, using (3.67) and taking into account that  $\langle S^2 \rangle$  does not depend from the direction of any axis, so that  $\langle S^2 \rangle = S^2$ , we obtain the following general result:

$$S^2 = \hbar s(s+1) \quad (3.68)$$

This important relation shows that for an atomic system the maximum projection of  $\mathbf{S}$  along an arbitrary axis is never  $S$ , but it is smaller than this quantity, because  $s\hbar$  is always less than the square root of  $\hbar s(s+1)$ . Therefore, in the quantum mechanical world the *observed* projection of angular momentum can never be exactly aligned with any particular axis! An electron has spin  $s = 1/2$ , so that there are only two possible values for the projection of  $\mathbf{S}$  along an arbitrary axis:  $(\mathbf{S} \cdot \mathbf{n}) = \pm \hbar/2$ , which correspond to the two possible (“up” and “down”) states associated with its spinning. Now let us consider the interaction between the magnetic moment of an electron and a magnetic field. When the electron is placed in an external magnetic field, it acquires an additional potential energy that depends from the component of  $\mathbf{m}_S$  along  $\mathbf{B}_{ext}$ :

$$U_m = -\mathbf{m}_S \cdot \mathbf{B}_{ext} \quad (3.69)$$

This equation shows that the magnetic energy depends from the orientation of the magnetic moment with respect to the external field: if  $\mathbf{m}_S \perp \mathbf{B}_{ext}$ , then the potential energy is maximum ( $U_m = 0$ ), whereas  $U_m$  assumes its minimum value ( $U_m = -m_S B_{ext}$ ) when the magnetic moment is aligned with the field. Substituting (3.65) gives:

$$U_m = \frac{e}{m_e} \mathbf{S} \cdot \mathbf{B}_{ext} \quad (3.70)$$

Equation (3.66) puts a constrain on the possible values of  $\mathbf{S} \cdot \hat{\mathbf{B}}_{ext}$ , which are  $\pm \hbar/2$ . Therefore, the magnetic energy, just like the projection of  $\mathbf{S}$  along an axis, can assume only two values:

$$U_m = \pm \frac{e\hbar}{2m_e} B \equiv \pm \mu_B B \quad (3.71)$$

where the quantity:

$$\mu_B \equiv \frac{e\hbar}{2m_e} = 9.27400968(20) \times 10^{-24} [\text{JT}^{-1}] \quad (3.72)$$

is called the *Bohr magneton*. Let us consider now a neighbor magnetic dipole  $\mathbf{m}'_S$  at position  $\mathbf{r}$ . In this instance,  $\mathbf{B}_{ext}$  has the form (3.21) and  $U_m$  represents the magnetic potential of dipole-dipole interaction. We have:

$$\begin{aligned} U_m &= -\mathbf{m}'_S \cdot \mathbf{B}_{ext}(\mathbf{r}) \\ &\cong -\frac{\mu_0}{4\pi} \left[ 3 \frac{(\mathbf{m}_S \cdot \mathbf{r})(\mathbf{m}'_S \cdot \mathbf{r})}{r^5} - \frac{\mathbf{m}_S \cdot \mathbf{m}'_S}{r^3} \right] \end{aligned} \quad (3.73)$$

As we may have expected, this expression is symmetric with respect to  $\mathbf{m}_S$  and  $\mathbf{m}'_S$ . If this were the interaction responsible for the alignment of spins in ferromagnetic materials, then the thermal energy corresponding to the Curie temperature  $T_c$ , which is  $kT_c$  ( $k = 1.3806488(13) \times 10^{-23} [\text{JK}^{-1}]$  being the Boltzmann constant), would have the same order of magnitude of the strongest field energy of dipolar interaction. In fact, the former represents the threshold thermal energy separating ordered states from paramagnetic behavior, while the latter would be a measure of the aligning energy. A rough estimate can be obtained considering that the nearest neighbor separation coincides with the lattice constant  $a$ . Therefore, setting  $m_S = m'_S$  in (3.73), we would have, for in-plane interactions:

$$kT_c \approx U_{m,\max} = n \frac{\mu_0}{4\pi} \frac{m_S^2}{a^3} \quad (3.74)$$

where  $n$  is the number of nearest in-plane neighbors. From (3.65), (3.68), and (3.72) it results:

$$m_S^2 = \frac{e^2 S^2}{m_e^2} = \frac{e^2 \hbar^2 s^2 (s+1)^2}{m_e^2} = \mu_B^2 \quad (3.75)$$

Therefore,

$$kT_c \approx U_{m,\max} = n \frac{\mu_0}{4\pi} \frac{\mu_B^2}{a^3} \quad (3.76)$$

Taking  $n = 4$  and  $a = 0.8393$  nm (as appropriate for magnetite crystals), we get  $T_c \sim 0.0042$  K. However, actually magnetite becomes ferromagnetic at  $\sim 858$  K! Clearly, the dipole-dipole interaction is too weak to explain the ferromagnetic behavior of this mineral. It can only be a correction to the actual *exchange energy*  $\mu_B B \sim kT_c$ , and in theory represents a source of anisotropy. Ironically, the ultimate source of the strongest form of magnetism results to be a combination of Coulomb (i.e., electric, *not* magnetic) interaction between neighbor electrons and the necessity for these particles to satisfy the *Pauli exclusion principle* of quantum physics. This principle states that in an atomic system the *quantum state* of an electron, which is specified by a set of four *quantum numbers*, must be different from any other electron in the system. The four quantum numbers are:

- The *principal quantum number*,  $n$ , which defines the *size* of the shell where an electron moves;
- The *orbital quantum number*,  $l$ , associated with the angular momentum, which determines the *shape* of the shell;
- The *orbital magnetic quantum number*,  $m$ , associated with the orbital magnetic moment;
- The *spin magnetic quantum number*,  $m_s$ , which is associated with the spin magnetic moment

Therefore, as  $m_S = \pm 1/2$ , given a triplet of quantum numbers  $(n, l, m)$ , we can find at most two electrons in an atomic system having these quantum numbers and opposite spins. In some

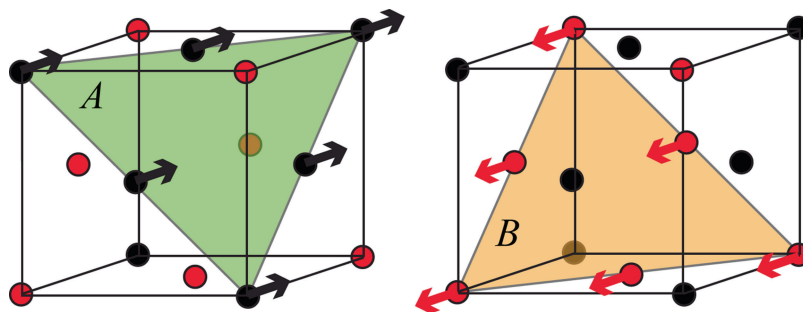
materials, the atoms are packed within the crystal lattice in such a way that highly eccentric orbital shells of adjacent atoms overlap. In this instance, the electrons that move across these shells will be forced to satisfy simultaneously the exclusion principle of *both* atoms. This phenomenon can be described intuitively by the so-called *exchange interaction*, which was proposed in 1928 by Heisenberg to explain the very large electromagnetic fields that form in ferromagnetic materials. Let us consider two atoms with unpaired electrons and assume that they get close each other. By Pauli's principle, if the spins of the two electrons align antiparallel to each other, these electrons will be able to share a common orbital shell, and this event would *increase* the electrostatic Coulomb energy. Conversely, if the spins align parallel to each other, Pauli's exclusion principle will prevent the formation of a shared orbit, so that the two electrons would move apart along separate orbits, thus *reducing* the normal Coulomb interaction. The latter solution is clearly *favoured* by nature. The order of magnitude of the Coulomb energy that would be required by a shared orbit is given by:

$$U_e = \frac{e^2}{4\pi\epsilon_0 r} \quad (3.77)$$

where  $r$  is the average distance between the two electrons and the constant  $\epsilon_0 = 8.8542 \times 10^{-12}$  C<sup>2</sup>N<sup>-1</sup>m<sup>-2</sup> is the *free space permittivity*. Taking  $r = 1$  Å gives:

$$T_c \approx \frac{e^2}{4\pi\epsilon_0 k r} = 1.67 \times 10^5 \text{ K} \quad (3.78)$$

This value is  $10^5$  times larger than the magnetic dipole interaction calculated from (3.76) using the same distance. Therefore, if the Coulomb energy is, on average, reduced by the Pauli exclusion principle to a small fraction of the value required by (3.77), say 0.5 %, the variation of electrostatic energy after the formation of the new atomic system will give a Curie temperature of  $\sim 835$  K, which can explain the magnitude of the molecular field. Thus, the parallel alignment of electron spins in ferromagnetic materials results



**Fig. 3.18** Antiferromagnetic arrangement of Nickel<sup>II</sup> oxide (NiO), which has the fcc crystal structure of NaCl, with octahedral Ni<sup>II</sup> (black and red circles) and O<sup>2-</sup> sites

(not displayed). The A and B sublattices are shown in green and yellow. Arrows show the spin directions of Ni atoms

from the combined effect of electrostatic interaction and a quantum-mechanical phenomenon, the latter being the Pauli exclusion principle.

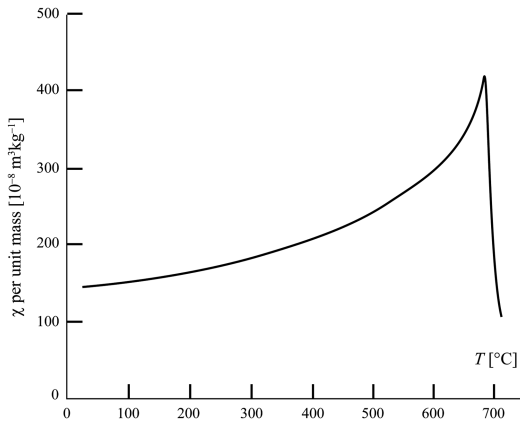
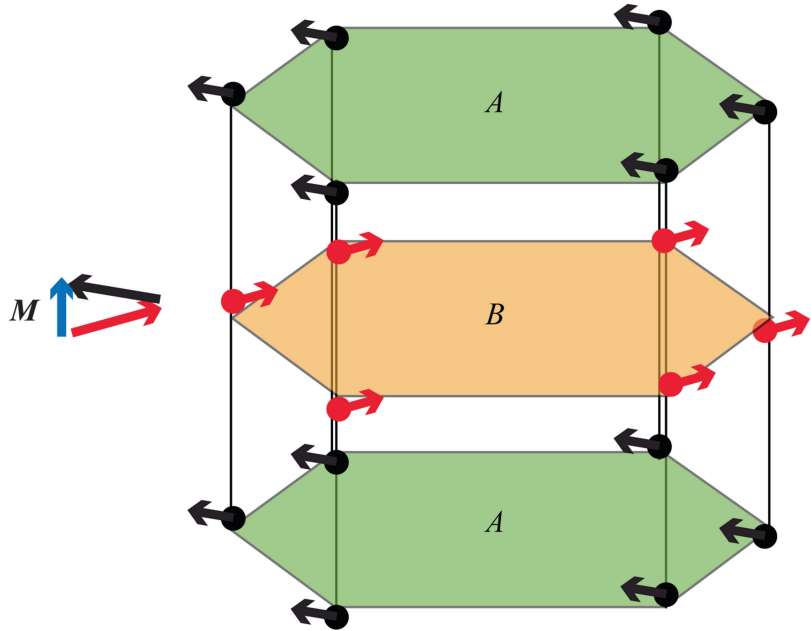
Although this description of ferromagnetism explains the magnetic behavior of many metals, the two series of magnetic minerals that are of interest for plate tectonics, titanomagnetites and titanohematites, belong to two special classes of ferromagnetic materials, characterized by an “unusual” alignment of the atomic spins. Hematite is an example of *antiferromagnetic* mineral. In antiferromagnetism, the crystal lattice can be ideally divided in two disjoint subsystems (or *sublattices*), A and B, in such a way that neighboring spins are coupled only within each subsystem, where they have parallel alignment. Furthermore, the magnetic moments belonging to adjacent sublattices have antiparallel orientation, as shown in the example of Fig. 3.18. Therefore, if the two subsystems have equal magnetic moment, the resulting net magnetization will be zero.

In general, antiferromagnetic minerals have no spontaneous magnetization and display a weak magnetism, with a susceptibility ranging from  $10^{-5}$  to  $10^{-2}$ , similar to that of paramagnetic materials. However, differently from the latter substances, antiferromagnetic materials present an ordered structure. In some cases the magnetic moments of the A and B layers are not perfectly antiparallel, so that a non-zero net magnetization results from the vector difference between the two spins. This kind of ferromagnetism is referred to as *canted antiferromagnetism*. Figure 3.19 shows the canted antiferromagnetic

arrangement of hematite, in which the small net magnetization resulting from the imperfect alignment of Fe<sup>3+</sup> spins is nearly perpendicular to the crystal sublattices. If we place an antiferromagnetic solid in an external magnetic field parallel to the spin axes, the torque exerted on the elementary current loops is nearly zero, so that the ordered spin arrangement is not disturbed. Therefore, the magnetic susceptibility is smaller than that of a normal paramagnetic substance. However, in so far as we increase the temperature, the ordered arrangement tends to be destroyed, so that the resulting canted structure determines an *increase* of susceptibility, contrarily to the case of paramagnetic materials. However, above a critical value of temperature the spin ordering disappears completely and the solid behaves like a paramagnetic substance. This temperature, which is called *Néel temperature*,  $T_n$ , is more appropriate than the Curie temperature to describe the transition point from antiferromagnetic arrangement to paramagnetic disorder. The Néel temperature is associated with a maximum of magnetic susceptibility, as shown in Fig. 3.20. In the case of hematite, we have  $T_n \cong 673^\circ\text{C}$ .

The other special class of ferromagnetic materials that are characterized by “unusual” alignment of the atomic spins is represented by the *ferrimagnetic* solids. Magnetite is an important example of this class of magnetic minerals. In ferrimagnetic substances, the A and B sublattices of an antiferromagnetic spin arrangement are occupied by different magnetic atoms and sometimes

**Fig. 3.19** Canted antiferromagnetic arrangement of  $\text{Fe}^{3+}$  spins unit cells of hematite. Red and black circles are  $\text{Fe}^{3+}$  ions with opposite magnetic moments. Note that the ions are displaced upward or downward along the  $c$ -axis of the crystal. Red and black arrows indicate spin directions. The resulting net magnetization is represented by the blue arrow



**Fig. 3.20** Magnetic susceptibility per unit mass,  $\chi$ , vs. temperature ( $^{\circ}\text{C}$ ) for a sample of hematite ( $\text{Fe}_2\text{O}_3$ ) (Modified from de Boer et al. 2001)

by a different number of atoms. The result is a net spontaneous magnetization, which may be of relevant magnitude. In the case of magnetite, the spinel crystal structure of this mineral yields a preferred direction of magnetization along the cube diagonal  $[111]$ . Other examples of ferrimagnetic minerals are titanohematites,  $\text{Fe}_{2-x}\text{Ti}_x\text{O}_3$ , when  $x > 0.45$ , whereas for  $0 < x < 0.45$  these substances conserve the canted antiferromagnetic arrangement of hematite.

The *natural remnant magnetization* (NRM) of a rock is the remanent magnetization  $M_r$  that can be measured on a rock specimen through a laboratory magnetometer. The magnitude and direction of this vector depend from the geographic location of the rock at the time of its formation and eventually from subsequent geological processes. The original NRM acquired during the process of formation of the rock is referred to as the *primary* NRM. Any successive component that contributes to the observed remnant magnetization is called a *secondary* component. Thus,

$$M_r = M_{\text{primary}} + M_{\text{secondary}} \quad (3.79)$$

There are three basic kinds of primary NRM. They are:

- *Thermoremanent magnetization* (TRM), which is acquired during cooling from temperatures above the Curie temperature. This is the typical form of remanent magnetization of igneous rocks;
- *Chemical remanent magnetization* (CRM), which forms by growth of ferromagnetic grains below the Curie temperature and is the typical kind of remanent magnetization in sedimentary rocks. It results from chemical

reactions that form new ferromagnetic minerals by alteration of existing minerals or precipitation;

- *Detrital remanent magnetization* (DRM), which forms during the process of accumulation and lithification of sediments containing detrital ferromagnetic minerals.

In the next chapters, we shall apply these concepts to the study of the Earth's magnetic field in the geologic past, which furnishes the basic class of data used in plate kinematics.

## Problems

1. Calculate the maximum diamagnetic magnetization of 1 kg Hydrogen in an external field  $B$  with magnitude  $B = 0.1$  T;
2. The saturation magnetization per unit mass for hematite is  $M_s \cong 0.48 \text{ A m}^2 \text{ kg}^{-1}$ . Assuming that a hematite assemblage is formed by  $10 \text{ }\mu\text{m}$  SD grains, determine the magnetic moment of a single grain at saturation;
3. A compass needle having volume  $V = 20 \times 10^{-9} \text{ m}^3$  has magnetization  $M = 300 \text{ kA/m}$ . Assuming that at some location the Earth's magnetic field has strength  $F = 40,000 \text{ nT}$  and inclination  $I = 35^\circ$ , determine the maximum torque exerted on the needle;
4. Determine the maximum magnitude of dipolar interaction energy for hematite;

## References

- Butler RF (1992) *Paleomagnetism: magnetic domains to Geologic terranes*. Blackwell Scientific Publications, Boston, p 238
- de Boer CB, Mullender TAT, Dekkers MJ (2001) Low-temperature behaviour of haematite: susceptibility and magnetization increase on cycling through the Morin transition. *Geophys J Int* 146:201–16
- Feynman RP, Leighton RB, Sands M (2006) *The Feynman lectures on physics, The definitive edition, vol 2, Chapter 34*. Pearson Education Inc., California Institute of Technology, Pasadena, CA
- Harrison RJ, Putnis A (1996) Magnetic properties of the magnetite-spinel solid solution: curie temperatures, magnetic susceptibilities, and cation ordering. *Am Miner* 81:375–84
- Hunt CP, Moskowitz BM, Banerjee SK (1995) Magnetic properties of rocks and minerals. In: Ahrens TJ (ed) *Rock physics and phase relations: a handbook of physical constants*, 3rd edn. AGU, Washington, DC, pp 189–204. doi:10.1029/RF003p0189
- Panofsky WKH, Phillips M (2005) *Classical electricity and magnetism*. Courier Dover, Mineola, NY



---

**Abstract**

The Earth's magnetic field represents the ultimate cause of rock magnetization. Here I describe the main internal source, or geodynamo, associated with convective currents in the external core, as well as the crustal field, generated by magnetized rocks, and the external magnetospheric and ionospheric sources. The concepts discussed in this chapter are essential for the correct use of magnetic data in plate kinematics. In particular, it is considered the secular variation of the core field, which represents the ultimate cause of most difficulties in collecting and analysing paleomagnetic data on continents. Furthermore, the concepts of polarity inversion and chron are introduced, along with the techniques used to build geomagnetic timescales. These topics are particularly important for the analysis of marine magnetic anomalies, a major source of data in plate kinematics. Finally, this chapter introduces the potential field technique, which represents the fundamental mathematical tool for describing the Earth's magnetic field.

---

**4.1 Source of the Main Geomagnetic Field**

The first systematic studies on the Earth's magnetic field go back to the beginning of the 1800s. Gauss, von Humboldt, and Weber in Germany, and Arago in France, were the pioneer of the modern techniques of observation of the Earth's magnetic field time variations (Courtillot and Le Mouél 1988). In addition, Gauss introduced a new theoretical method to represent the field in terms of *spherical harmonics*, a technique that is still in use. However, another century occurred to have the

first realistic models about the *origin* of the field. Joseph Larmor first proposed in 1919 that the geomagnetic field originates from electric currents associated with fluid motions in the Earth's core. Today we know that the observed magnetic field at the Earth's surface (or close to it) is mainly produced within the *external* core, but also includes a small component of crustal origin and an even smaller "external" component, whose source is in the Earth's ionosphere (between 85 and 600 km altitude) and in the magnetosphere (a region that extends beyond the atmosphere). On average, the field that originates from magnetization of crustal rocks is two orders of magnitude lower than

the main field, while the mantle is essentially non-magnetic. The exclusion of any mantle contribution is also supported by the fact that mantle convection is too slow to account for the observed rate of variation of the geomagnetic field. Therefore, the field must be associated with *macroscopic* currents somewhere below the CMB. Current models about the origin of the Earth's main field invoke the existence of convective motions within the outer liquid layer of the Earth's core, which are driven by thermal and compositional variations just above the inner core–outer core boundary (ICB) at 2,891 km depth (e.g., Glatzmaier and Roberts 1995). The Earth's core is made by an alloy of iron and lighter elements. Iron has a good electrical conductivity  $\sigma \sim 4 \times 10^5 \text{ S m}^{-1}$ . As the Earth cools, iron freezes at the base of the outer core and accretes to the inner core, determining its progressive growth. During this process, the latent heat of crystallization heats the residual liquid, which is also enriched in light elements, determining its positive buoyancy and triggering the convective process. The upward motion of the buoyant fluid does not occur along straight lines, because the Coriolis' force associated with the Earth's rotation determines the formation of helical flows aligned as the Earth's rotation axis. Now we are going to examine the basic physical laws underlying the formation and the maintenance of a magnetic field by electric currents in the outer core. To this purpose, it is necessary to consider again the fundamental laws of classical electrodynamics, including the Maxwell equations.

In Chap. 3, we have seen that charged particles in motion generate magnetic fields. Charged particles also generate *electric fields*, independently from being at rest or not, the strength of which depends on their electric charge. An electric field,  $\mathbf{E}$ , is a conservative force field that can be felt by any other charged particle and represents the electric force per unit charge on a test particle. Its units are  $[\text{Vm}^{-1}] = [\text{kg} \cdot \text{m}^2 \cdot \text{s}^{-3} \cdot \text{A}^{-1}]$ . Let  $\rho = \rho(\mathbf{r})$  be the density function for a distribution of electric charges (that is, the electric charge per unit volume). If the charges are at rest in the laboratory frame of reference, the corresponding

electric field is said to be *electrostatic* and the expression for  $\mathbf{E}$  in the vacuum is given by *Coulomb's law*:

$$\mathbf{E}(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0} \int_{\mathbf{R}} \frac{\rho_e(\mathbf{q})(\mathbf{r} - \mathbf{q}) dV}{\|\mathbf{r} - \mathbf{q}\|^3} \quad (4.1)$$

where  $\mathbf{R}$  is the region of distribution of the electric charges and  $\epsilon_0 = 1/(c^2\mu_0) = 8.8542 \times 10^{-12} \text{ F/m}$  ( $c$  being the speed of light) is the *vacuum permittivity*. The forces associated with electric fields give rise to flow of electric charges in the free space and in *conductors*. The constitutive law that describes the resistance of materials to be traversed by electric currents is the empirical *Ohm's law*, which states that in isotropic materials the current density  $\mathbf{j}$  at a location  $\mathbf{r}$  is proportional to the electric field  $\mathbf{E}$ :

$$\mathbf{j}(\mathbf{r}) = \sigma \mathbf{E}(\mathbf{r}) \quad (4.2)$$

where the *electric conductivity* of the material,  $\sigma$ , is measured in  $[\text{S m}^{-1}]$ . If the material is moving with uniform velocity  $\mathbf{u}$ , the law (4.2) is still valid for an observer who is in motion with it, but *not* for one that is at rest in the laboratory frame of reference. If  $\rho$  is the density of stationary charges in a reference frame fixed to the moving material, then this distribution of charge is viewed, from the perspective of the laboratory frame, as an electric current having density  $\rho\mathbf{u}$ .

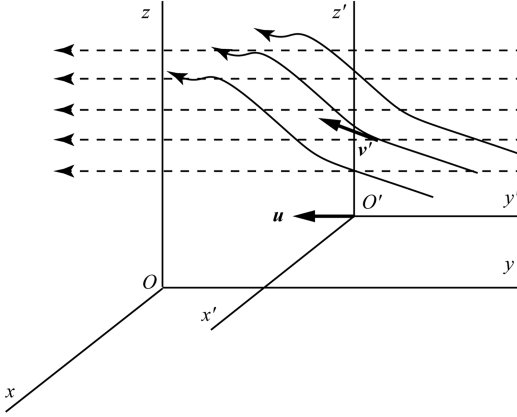
The quantity:

$$\mathbf{j}_c(\mathbf{r}) = \rho(\mathbf{r})\mathbf{u} \quad (4.3)$$

is called *convective current density*. Let  $\mathbf{j}'$  and  $\mathbf{E}'$  be respectively the current density and the electric field measured in the moving frame. The invariance of (4.2) implies that:

$$\mathbf{j}' = \sigma \mathbf{E}' \quad (4.4)$$

What is the relation between the currents  $\mathbf{j}$  and  $\mathbf{j}'$ ? The response is simple. In general, in the moving frame we have electric charges with local density  $\rho$  that move with some velocity  $\mathbf{v}'$ . Therefore, in this frame we have a density of current  $\mathbf{j}' = \rho\mathbf{v}'$  (Fig. 4.1). However, from the



**Fig. 4.1** Flow of electric charges through a moving conductive fluid. The fluid moves leftward with uniform velocity  $\mathbf{u}$  (dashed lines). The charges have local density  $\rho$  and velocity  $\mathbf{v}'$  in the reference frame  $x'y'z'$  of the fluid, so that the flow of charges through the conductor is represented by a density of current  $\mathbf{j}' = \rho \mathbf{v}'$

point of view of an observer in the laboratory frame, the velocity of the charges is  $\mathbf{v} = \mathbf{v}' + \mathbf{u}$ , so that the density of current will be given by:

$$\mathbf{j} = \rho \mathbf{v} = \rho (\mathbf{v}' + \mathbf{u}) = \mathbf{j}' + \mathbf{j}_c \quad (4.5)$$

Now we want to determine the relation between the electric field  $\mathbf{E}'$  in (4.4) and the electric field  $\mathbf{E}$  observed in the laboratory frame. In this frame, a test charge that is at rest in the moving frame appears to be subject to the simultaneous action of an electrostatic field  $\mathbf{E}$  and a Lorentz force per unit charge, which is  $\mathbf{u} \times \mathbf{B}$  by (3.5). Therefore, the total force exerted on a charge  $q$  is given by:

$$\mathbf{F} = q (\mathbf{E} + \mathbf{u} \times \mathbf{B}) \quad (4.6)$$

This is in fact the complete form of the Lorentz force. Now let us consider an observer in the moving frame. From the point of view of this person, the test particle  $q$  is at rest, so that the unique force that comes to play is the electrostatic force  $q\mathbf{E}'$ . Therefore, in the moving frame it results that  $\mathbf{F}' = q\mathbf{E}'$ . The two forces  $\mathbf{F}$  and  $\mathbf{F}'$  must be clearly equal, because the acceleration is the same in the two reference frames ( $\mathbf{u}$  is a constant vector). Consequently, we have:

$$\mathbf{E}' = \mathbf{E} + \mathbf{u} \times \mathbf{B} \quad (4.7)$$

Combining (4.3) and (4.4) with (4.7) we obtain the following expression for the density of current in the laboratory frame:

$$\mathbf{j} = \mathbf{j}' + \rho \mathbf{u} = \sigma \mathbf{E}' + \rho \mathbf{u} = \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}) + \rho \mathbf{u} \quad (4.8)$$

This is the general form of Ohm's law in the laboratory frame. It establishes the dependency of the current density through a conductor that is moving with velocity  $\mathbf{u}$  from the electric and magnetic fields and from the local scalar charge density. The convective term  $\rho \mathbf{u}$  in (4.8) is often negligible in comparison with the transport by conduction, which is proportional to  $\sigma$ . For example, in the case of the Earth's core the conductivity is thought to be very large:  $\sigma \sim 4 \times 10^5 \text{ Sm}^{-1}$ . In this instance, Ohm's law assumes the following simpler form:

$$\mathbf{j} = \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}) \quad (4.9)$$

Substituting this expression in Ampere's law (3.27), gives:

$$\nabla \times \mathbf{B} = \mu_0 \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}) \quad (4.10)$$

This form of the Maxwell-Ampere equation combines a macroscopic current, depending on  $\mathbf{E}$ , with a term depending from the velocity  $\mathbf{u}$  of the conductor. In the case of the Earth's magnetism, the electric field  $\mathbf{E}$  at the right-hand side of (4.10) cannot be a conservative field, generated by a distribution of charges as in Eq. (4.1). In fact, the Earth's mantle is essentially an electric insulator, with a conductivity ranging between  $10^{-4}$  and  $1 \text{ Sm}^{-1}$  (Dobson and Brodholt 2000), thereby, the source of the electric field in (4.10) must be somewhere within the Earth's core. A possibility could be a distribution of electric charges in the solid inner core, a sort of giant battery, but this solution would require rather special conditions that cannot be reasonably hypothesized (for an in-depth discussion, see Elsasser 1939). Therefore, the problem is: what is the origin of the currents flowing through the

outer core? The response to this question comes from one of the fundamental laws of classical electrodynamics. In Sect. 3.3 we introduced only two of the four Maxwell equations, in the special case of stationary fields in the empty space. Now we must consider a third equation in the more general context of time-varying fields. This equation, which is known as *Faraday's law*, predicts the existence of a *non-conservative* (i.e., not coming from a potential) electric field  $\mathbf{E}$  associated with temporal variations of  $\mathbf{B}$ :

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (4.11)$$

This equation describes the *electromagnetic induction* determined by a time-varying magnetic field. Soon we shall further investigate the consequences and the significance of this fundamental law. For the moment, it will be used just to eliminate  $\mathbf{E}$  from (4.10):

$$\begin{aligned} \frac{\partial \mathbf{B}}{\partial t} &= -\nabla \times \left( \frac{1}{\mu_0 \sigma} \nabla \times \mathbf{B} - \mathbf{u} \times \mathbf{B} \right) \\ &= -\frac{1}{\mu_0 \sigma} \nabla \times \nabla \times \mathbf{B} + \nabla \times (\mathbf{u} \times \mathbf{B}) \end{aligned} \quad (4.12)$$

The first term at the right-hand side of (4.12) can be simplified using a simple rule of vector calculus (see Appendix 1) and Gauss' law (3.25):

$$\nabla \times \nabla \times \mathbf{B} = \nabla (\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B} = -\nabla^2 \mathbf{B} \quad (4.13)$$

Substituting in (4.12) gives the fundamental equation of magnetohydrodynamics (MHD):

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B} \quad (4.14)$$

where  $\eta \equiv 1/(\mu_0 \sigma)$  is called *magnetic diffusivity*. Equation 4.14 is known as the *magnetic induction equation* and plays a key role in the study of the Earth's magnetic field and in plasma physics. It allows to determine the magnetic field associated with currents that are originated by electromagnetic induction as a consequence of convective motions within the Earth's core. If the velocity

$\mathbf{u}$  is zero, the first term at the right-hand side of (4.14) disappears. In this *diffusion limit* we have:

$$\frac{\partial \mathbf{B}}{\partial t} = \eta \nabla^2 \mathbf{B} \quad (4.15)$$

This equation is well known in mathematics. It is called the *diffusion equation* and is found in many applications. For example, in Chap. 12 we shall see that it describes the non-stationary conduction of heat. To understand the significance of the diffusion term in the magnetic induction equation, let us consider a field depending only from  $x$  and assume that at time  $t = 0$  we have:

$$B(x, 0) = \begin{cases} +B_0; & x > 0 \\ -B_0; & x < 0 \end{cases} \quad (4.16)$$

Let us also assume that the field is held fixed at points  $\pm L$ , so that:

$$B(L, t) = -B(-L, t) = B_0 \quad (4.17)$$

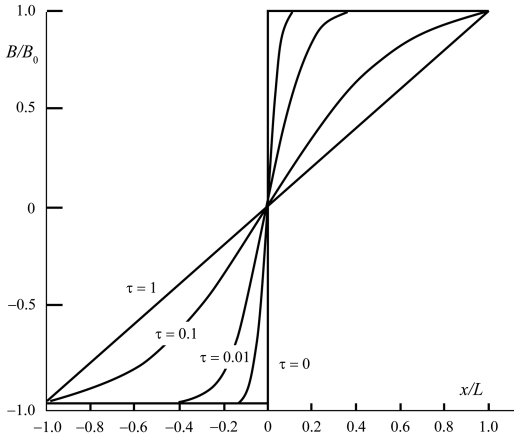
In this instance, it is possible to show that the solution to the diffusion equation is (e.g., Wilmot-Smith et al. 2005):

$$\begin{aligned} B(x, t) &= B_0 \frac{x}{L} + \frac{2B_0}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \\ &\times \exp(-n^2 \pi^2 \eta t / L^2) \sin\left(\frac{n\pi x}{L}\right) \end{aligned} \quad (4.18)$$

A graphical representation of this solution is illustrated in Fig. 4.2. It suggests that the field diffuses gradually into the fluid, thus removing local inhomogeneities. If  $L$  is interpreted as the length scale of magnetic inhomogeneities, then the *diffusion time* is defined as follows:

$$\tau_D \equiv \frac{L^2}{\eta} \quad (4.19)$$

This is the time interval required to smooth away any local anomaly of the field. For example, Fig. 4.2 shows that after a time  $t = \tau_D$  the magnetic field distribution is within a factor  $10^{-4}$  from the stationary solution  $B(x) = B_0 x/L$ . In the general case of an infinite range ( $L \rightarrow \infty$ )



**Fig. 4.2** Distribution of the magnetic field for different values of the non-dimensional time  $\tau = \eta t / L^2$  predicted by (4.18). It is apparent the progressive smoothing of the field with time

and an arbitrary initial profile  $B(x, 0) = f(x)$ , it is possible to show that the field decays rapidly to zero on a time scale given by  $\tau_D$ . In the case of the Earth, the geomagnetic field would disappear within  $10^5$  years. Consequently, the induction term  $\nabla \times (\mathbf{u} \times \mathbf{B})$  in (4.14) is effective to contrast the decay associated with diffusion. For example, if we set the diffusion term in (4.14) to zero, which is equivalent to assume a very high conductivity of the fluid, then the magnetic field lines would be “frozen” into the fluid and would always be moving with it.

So far, we have considered the effect of convective motions for the maintenance of a magnetic field within the outer core. Now we want to briefly mention the action exerted to the fluid back by the magnetic field. We know that a Lorentz force is exerted on a moving charged particle in presence of magnetic field (Eq. 3.5). In the case of a fluid, which can be represented as a continuous distribution of mass (see Sect. 2.1), the Lorentz force per unit volume,  $\mathbf{f}$ , will given by:

$$\mathbf{f} = \rho (\mathbf{v} \times \mathbf{B}) = \mathbf{j} \times \mathbf{B} \quad (4.20)$$

This force must be incorporated into the fluid dynamics equations describing the relation between forces and accelerations in the liquid

core (we shall face the dynamics of fluids in Chap. 13). Therefore, a combined system of differential equations is necessary to determine the evolution of the Earth’s magnetic field. The Earth’s magnetic field model of Glatzmaier and Roberts (1995) is precisely the result of a numerical solution to the induction equation and related fluid dynamics and electromagnetic equations. We shall not investigate further this rather complex subject. However, in the next section, we are going to discuss a conceptual (analog) model for the generation of the geomagnetic field, starting from Faraday’s law.

## 4.2 The Geodynamo

Let us consider first the electrostatic field  $\mathbf{E} = \mathbf{E}(\mathbf{r})$  generated by a system of electric charges (Eq. 4.5). This is a *conservative field*, such that the work  $W$  done in moving a particle from a point  $P_1$  to another point  $P_2$  does not depend from the path  $\Gamma$  between the two points. In this case, a scalar function  $V = V(\mathbf{r})$  exists such that:

$$\mathbf{E} = -\nabla V \quad (4.21)$$

The function  $V$  is called *electrostatic potential* and its units are [V]. Then, the work per unit charge will be given by:

$$\begin{aligned} W(P_1, P_2) &= \int_{\Gamma} \mathbf{E}(\mathbf{r}) \cdot d\mathbf{r} = - \int_{P_1}^{P_2} \nabla V \cdot d\mathbf{r} \\ &= - \int_{P_1}^{P_2} \frac{\partial V}{\partial r} dr = V(P_1) - V(P_2) \end{aligned} \quad (4.22)$$

Therefore, the integral of  $\mathbf{E}$  along any closed loop  $\Gamma$  is zero:

$$\oint_{\Gamma} \mathbf{E} \cdot d\mathbf{r} = 0 \quad (4.23)$$

By Stokes' theorem (see Appendix 1), this relation implies, in turn, that:

$$\nabla \times \mathbf{E} = \mathbf{0} \quad (4.24)$$

This equation implies that the electric field generated by a system of charges is *irrotational*, so that it is not possible to maintain a stationary current along a closed circuit. However, Faraday discovered in 1840 that a *variable* magnetic field could induce an electric current through a conductor cable. In particular, he found that if he moved a magnet near an electric circuit, it was possible to *induce* a stationary current through the circuit. This phenomenon also occurred if he varied the current through a nearby wire. Conversely, Faraday observed that the presence of constant magnetic fields did not induce any current, independently from the field strength. We call *electromotive force* (emf)  $\mathcal{E}$  the energy that must be supplied by a source to move a unit of charge along a closed walk through the circuit. This quantity can be thought as the work done by a special kind of electric field, which does not originate from charge distributions.

In this representation, the emf  $\mathcal{E}$  coincides with the work done by a *non-conservative* electric field  $\mathbf{E}$  to move the charge along a closed loop  $\Gamma$ :

$$\mathcal{E}(C) = \oint_{\Gamma} \mathbf{E} \cdot d\mathbf{r} \quad (4.25)$$

Faraday's experiments showed that the existence of this non-conservative electric field was related to *variations* of magnetic field in a way that will be clarified now. Let  $C$  be a wire loop and  $S(C)$  any open surface bounded by  $C$ . Also, let  $\mathbf{B} = \mathbf{B}(\mathbf{r}, t)$  a *variable* magnetic field. Then, *Faraday's law* states that a non-conservative electric field  $\mathbf{E} = \mathbf{E}(\mathbf{r}, t)$  and an emf through  $C$  are generated, such that:

$$\mathcal{E}(C) = \oint_C \mathbf{E} \cdot d\mathbf{r} = - \int_{S(C)} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \quad (4.26)$$

By Stokes's theorem, the Maxwell-Faraday Eq. (4.11) follows. The quantity:

$$\Phi(C) = \int_{S(C)} \mathbf{B} \cdot d\mathbf{S} \quad (4.27)$$

is called *magnetic flux* through the surface  $S$ . Eq. (4.26) says that the emf associated with the current through the circuit  $C$  is the opposite of magnetic flux variation through  $S(C)$ :

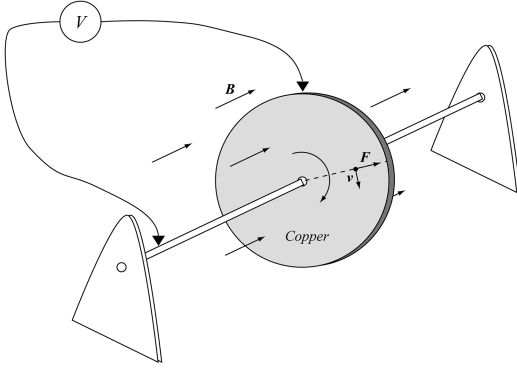
$$\mathcal{E}(C) = - \frac{\partial \Phi}{\partial t} \quad (4.28)$$

This equation has great practical importance, because it is the fundamental operating principle for the construction of a variety of electrical devices (e.g., electrical motors and generators). In fact, it implies that we can convert *mechanical energy*, which is necessary to move the magnet or the conductor, into *electrical energy*. One may wonder: what happens if we keep the magnetic field constant, so that  $\partial \mathbf{B} / \partial t = 0$ , while moving the wire? Of course, if the conductor  $C$  moves, then the flux  $\Phi(S)$  must change, because the integration surface in (4.27) also varies its orientation in space. Furthermore, the wire will be subject to a Lorentz force (4.20), which is non-conservative and can maintain a stationary current.

Therefore, it is possible to show that also in this case the induction law (4.28) holds, according to Faraday's experiments, and we have:

$$\mathcal{E}(C) = \oint_C (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{r} = - \frac{\partial}{\partial t} \int_{S(C)} \mathbf{B} \cdot d\mathbf{S} = - \frac{\partial \Phi}{\partial t} \quad (4.29)$$

The induction law (4.28), which links the emf to flux variations, does *not* have general validity, because in some circumstances a true circuit does not exist, and currents can flow through a volume. Faraday himself illustrated some of these situations, one of which is shown in Fig. 4.3. The device in Fig. 4.3 is called *Faraday's dynamo*. It is made by a disk that rotates about its axis in presence of a constant magnetic field. The angular velocity of the disk,  $\boldsymbol{\omega}$ , and the magnetic field,  $\mathbf{B}$ , are both aligned with the rotation axis. Any point on the disk surface has velocity  $\mathbf{v} = \mathbf{v}(r)$  lying in the disk plane, with magnitude  $v = \omega r$ ,



**Fig. 4.3** A rotating disk placed in a constant and uniform magnetic field produces an emf, which can be measured by the difference of potential,  $V$ , between the centre and the periphery of the disk

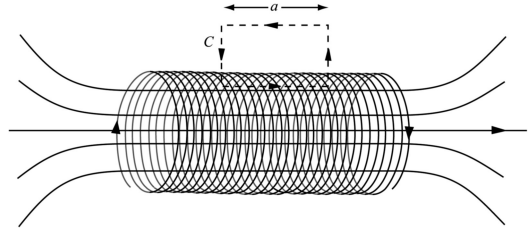
being the distance from the rotation axis. Because  $\omega$  and  $\mathbf{B}$  have the same direction, a unit charge at distance  $r$  will be subject to a Lorentz force  $\mathbf{F} = v\mathbf{B} = \omega r\mathbf{B}$ , directed radially toward the disk border. The total emf can be obtained integrating over the disk radius,  $a$ :

$$\mathcal{E} = \int_0^a vB dr = \omega B \int_0^a r dr = \frac{1}{2} \omega B a^2 \quad (4.30)$$

Clearly, we have also a constant magnetic flux through the disk surface, given by:

$$\Phi = \pi B a^2 \quad (4.31)$$

Thus, it is possible to have a non zero emf even in presence of constant flux. Another interesting phenomenon occurs when we consider a *solenoid* (Fig. 4.4). This is a coil wound into a tightly packed helix, having a diameter small with respect to its length. When a current passes through the coil, the resulting magnetic field is rather intense and uniform within the helix and negligible outside, with flow lines that are almost parallel to the solenoid axis (the divergence of  $\mathbf{B}$  must be zero everywhere). To calculate the field within the solenoid, we can apply the integral version of Ampere's law (3.27) to the closed loop  $C$  in Fig. 4.4. Taking the line integral of  $\mathbf{B}$  along an arbitrary closed loop  $C$  gives:



**Fig. 4.4** Flow lines of the magnetic field produced by a solenoid when a current passes through it. A path,  $C$ , can be used together with Ampere's law to determine the magnetic field within the solenoid

$$\begin{aligned} \oint_C \mathbf{B} \cdot d\mathbf{r} &= \int_{S(C)} \nabla \times \mathbf{B} \cdot d\mathbf{S} \\ &= \mu_0 \int_{S(C)} \mathbf{j} \cdot d\mathbf{S} = \mu_0 I(S) \end{aligned} \quad (4.32)$$

where we have used (3.3). This is the integral form of Ampere's law. It states that the closed loop integral of the magnetic field is always proportional to the current flowing through an arbitrary surface bounded by  $C$ . When we apply (4.32) to the loop  $C$  in Fig. 4.4, we note that  $C$  has an edge of length  $a$  within the solenoid, running parallel to  $\mathbf{B}$ , two edges that are perpendicular to  $\mathbf{B}$ , and one edge outside the wire, where the field strength is negligible. If  $I$  is the current flowing through the coil and  $M$  is the number of turns for a length  $a$ , then the current passing through a surface bounded by  $C$  is  $MI$ . Therefore, applying (4.32) we have:

$$\mu_0 MI = \oint_C \mathbf{B} \cdot d\mathbf{r} \cong Ba \quad (4.33)$$

Consequently, the field within a solenoid is approximately given by:

$$B \cong \mu_0 n I \quad (4.34)$$

where  $n = M/a$  is the number of turns per unit length. In so far as the number of turns tends to infinity, solution (4.34) tends to be an exact solution. Conversely, for a finite solenoid the axial magnetic field in proximity of the two ends is half

of that predicted by (4.34). The flux through any cross-section  $S$  of the solenoid is:

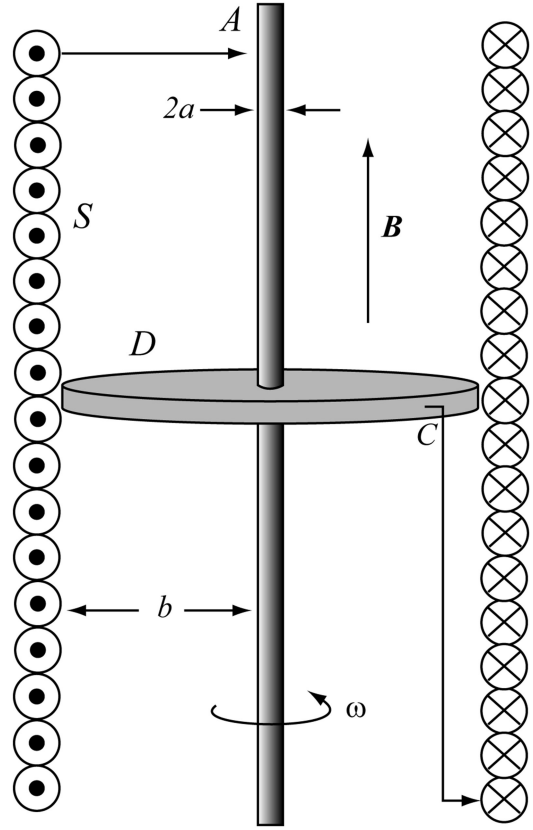
$$\Phi(S) = \int_S \mathbf{B} \cdot d\mathbf{S} = BA = \mu_0 \frac{NA}{L} I \quad (4.35)$$

where  $A$  is the cross-section area,  $N$  is the total number of turns, and  $L$  is the length of the solenoid, so that  $n = N/L$ . Now, if we vary the current  $I$  through the coil, we obtain a variation of flux for each turn. Consequently, by Faraday's law (4.28) an emf  $\mathcal{E}$  appears, which *opposes* the variations of current from which it was generated (remind the minus sign in Faraday's law). The resulting additional current is said to be *self-induced*. To calculate the emf through the solenoid, we multiply the emf of each turn by  $N$ :

$$\mathcal{E} = -N \frac{\partial \Phi}{\partial t} = -\mu_0 \frac{N^2 A}{L} \frac{dI}{dt} \equiv -\mathcal{L} \frac{dI}{dt} \quad (4.36)$$

The quantity  $\mathcal{L}$  in (4.36), which depends from the solenoid geometry, is called *self-inductance* of the solenoid and has units of Henry [H]. A very interesting phenomenon occurs if we use a solenoid to generate the magnetic field of Fig. 4.3. It is not difficult to show that in this instance the system becomes unstable. The modified circuit is illustrated in Fig. 4.5. We have again a rotating disk and an external initial field  $\mathbf{B}_0$ . As in the example of Fig. 4.3, the induced current flows toward the periphery of the disk.

However, in this new system, the contacts  $C$  and  $A$  and the disk shaft allow the current to flow externally through a solenoid  $S$ . This flow generates a new axial magnetic field  $\mathbf{B}$  having a magnitude determined by (4.34). At this point, removing the initial field  $\mathbf{B}_0$  and keeping the disk in motion, we expect that the spontaneous field  $\mathbf{B}$  continues to be sustained by the current that it itself provides to generate. In this instance, apart from a possibly weak initial magnetic field  $\mathbf{B}_0$  seeding the system, no other external field is needed to sustain the dynamo. Therefore, after removal of the seed field  $\mathbf{B}_0$ , the dynamo only



**Fig. 4.5** Self-exciting dynamo. A disk  $D$  rotates with angular velocity  $\omega$  in the axial magnetic field  $\mathbf{B}$ , generated by the current flowing through the solenoid  $S$ . Note the direction of the current flow in the wire

requires a continuous supply of mechanical energy to drive the electrical conductor sufficiently fast for self-excitation to be possible.

To understand what really happens, let us assume that before the initial time  $t = t_0$  the circuit be kept open, for example by disconnecting  $C$  or  $A$ , so that  $B(0) = 0$  and the emf is only manifest through a difference of potential  $V$  between the contacts  $C$  and  $A$ . From (4.30) it results:

$$V_0 = V(0) = \frac{1}{2} \omega B_0 (b^2 - a^2) \quad (4.37)$$

At  $t = t_0$  the circuit is closed and a current will flow through the solenoid, generating a magnetic field  $B = B(t)$ .

In this conditions, the disk works as a battery supplying an emf given by:

$$\mathcal{E}(t) = \frac{1}{2}\omega [B(t) + B_0] (b^2 - a^2) \quad (4.38)$$

This expression shows that an increase of  $B$  determines an increase of electromotive force, which in turn implies an increase of the current flowing through the solenoid. The increase of current determines in turn an increase of  $B$  and so on. We say that the system has *positive feedback*. Let  $R$  and  $\mathcal{L}$  be, respectively, the electric resistance (in Ohm) of the circuit and the self-inductance of the solenoid. Clearly, the total emf (4.38) must be equal to the electromotive force that is necessary to allow flowing of a current  $I$  through a circuit having electric resistance  $R$ , plus a component that compensates the emf (4.36) associated with current variations. The former is given by *Ohm's law*:

$$\mathcal{E}(t) = RI(t) \quad (4.39)$$

Therefore, we have:

$$\begin{aligned} \mathcal{L} \frac{dI}{dt} + RI &= \frac{1}{2}\omega [B(t) + B_0] (b^2 - a^2) \\ &\cong \frac{1}{2}\omega [B(t) + B_0] b^2 \quad (\text{for } b \gg a) \end{aligned} \quad (4.40)$$

Using (4.34) for the solenoid field  $B$ , gives:

$$\frac{dI}{dt} + \frac{R}{\mathcal{L}} \left(1 - \frac{\omega}{\omega_0}\right) I(t) = \frac{\omega B_0 b^2}{2\mathcal{L}} \quad (4.41)$$

where:

$$\omega_0 \equiv \frac{2R}{\mu_0 n b^2} \quad (4.42)$$

Equation 4.41 is a linear first-order differential equation with constant coefficients. The solution is:

$$I(t) = \frac{\omega B_0 b^2}{2R(1 - \omega/\omega_0)} \left\{ 1 - \exp \left[ -\frac{R}{\mathcal{L}} \left(1 - \frac{\omega}{\omega_0}\right) t \right] \right\} \quad (4.43)$$

where we have used the initial condition  $I(0) = 0$ . Let us assume now that  $\omega/\omega_0 > 1$ . In this instance, the argument of the exponential is positive and the current increases until it reaches some value, say  $I_1$ , at time  $t = t_1$ . This current flows through the solenoid in the direction indicated in Fig. 4.5, thereby, the induced magnetic field  $B$  has the same direction of the external field  $B_0$ . If at time  $t = t_1$  the field  $B_0$  is instantaneously removed, although in reality the presence of self-induction effectively prevents an instantaneous removal of the external field, the Eq. (4.41) for  $t \geq t_1$  becomes a homogeneous equation:

$$\frac{dI}{dt} + \frac{R}{\mathcal{L}} \left(1 - \frac{\omega}{\omega_0}\right) I(t) = 0 \quad (4.44)$$

Assuming the initial condition:  $I(t_1) = I_1$ , we have the following solution:

$$I(t) = I_1 \exp \left[ -\frac{R}{\mathcal{L}} \left(1 - \frac{\omega}{\omega_0}\right) (t - t_1) \right]; \quad t \geq t_1 \quad (4.45)$$

This is an exact solution only for times  $t$  much greater than the decay constant of the external field  $B_0$ . It implies that even in absence of external field the system sustains a self-excited magnetic field  $B = B(t)$ . By (4.34) this field is given by:

$$B(t) = B_1 \exp \left[ -\frac{R}{\mathcal{L}} \left(1 - \frac{\omega}{\omega_0}\right) (t - t_1) \right]; \quad t \geq t_1 \quad (4.46)$$

Therefore, we can draw the following conclusions:

1. If  $\omega/\omega_0 < 1$ , both  $I$  and  $B$  decay exponentially;
2. If  $\omega = \omega_0$ , we have a stationary field  $B(t) = B_1$ ;
3. If  $\omega/\omega_0 > 1$ , both  $I$  and  $B$  increase exponentially.

Thus, the simple analog model of a self-excited dynamo composed by a solenoid and a rotating disk furnishes a theoretical background for the existence of self-sustained, persistent,

magnetic fields. The more realistic mechanism through which Earth's rotation and convective fluid motions in the outer core generate and sustain the main geomagnetic field is known as the *geodynamo*. The fundamental laws of this model are the MHD induction Eq. (4.14) and the equations of motion of fluid mechanics (Chap. 13). However, we have shown that the basic operating principle of a self-sustained magnetic field is conceptually simple and can be described through an analog model formed solely by a solenoid and a rotating disk. In the next sections, we shall focus on the time variations of the magnetic field generated by the geodynamo mechanism.

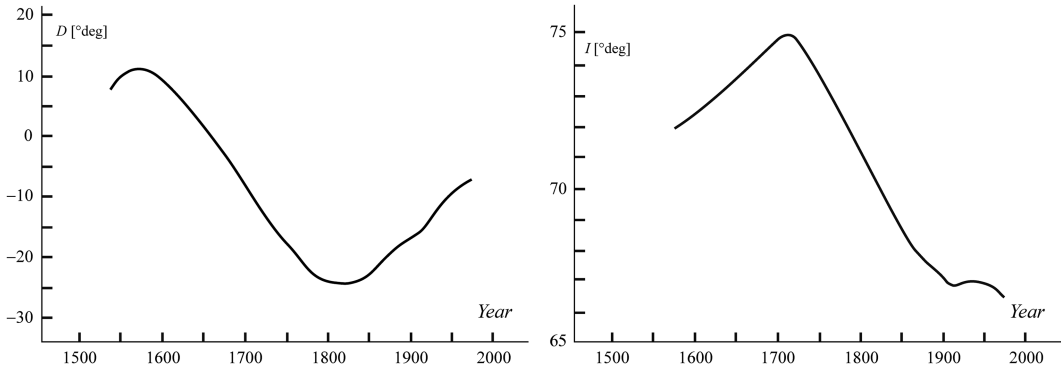
### 4.3 Secular Variation of the Core Field

The magnetic field generated within the external Earth's core by MHD processes is the *main geomagnetic field*, as it represents the major component of the magnetic field observed at the Earth's surface. The convective currents flowing in the external core have an estimated velocity of the order of 10 km/year (e.g., Buffet 2000, and references therein). The resulting field is subject to time variations that are significant at the scale of months. This variability of the main field is said *secular variation* and is of the order of 80 nT/year. The conventions used in geomagnetism to designate the components of the main field,  $\mathbf{F}$ , have been illustrated in Sect. 2.3. This field resembles the field that would be generated by a magnetic dipole placed at the Earth's centre and inclined by  $\sim 11^\circ$  with respect to the Earth's spin axis. As mentioned in Sect. 2.3, the axis of this dipole intersects the Earth's surface at the geomagnetic poles. If the field were perfectly dipolar, and in absence of crustal and external contributes, then the inclination at the geomagnetic poles would be  $I = \pm 90^\circ$ . The points at the Earth's surface where the *observed* field is effectively vertical are called *magnetic poles*. At present, the distance of these poles from the geomagnetic poles is several hundreds km and they are by no means aligned

with an axis passing through the Earth's centre. In fact, a better fit to the observed geomagnetic field can be obtained using a magnetic dipole located  $\sim 400$  km from the centre of mass in the direction of the positive  $z$ -axis. This model field is known as the *eccentric dipole field*. The magnitude of the observed geomagnetic field varies considerably along the Earth's surface and in time. The field vectors are sub horizontal close to the Equator, where  $F \sim 30,000$  nT, whereas they tend to be vertical in proximity of the geomagnetic poles, with magnitudes of the order of  $\sim 60,000$  nT. Regarding the variability in time, at the scale of months and years we observe that the inclined geocentric dipole, which represents about 90 % of the main field, has no fixed direction. The magnetic moment of this dipole precesses irregularly about the geographic North Pole, thereby determining variations of declination and inclination, and has variable magnitude. Although there is some superposition between the spectrum of time variations associated with external sources and the variability of the core field, in general the former are much more rapid and have much smaller magnitude. The secular variation of the Earth's magnetic field was first recognized by Gellibrand in 1634, who observed that the declination in London had decreased by more than  $7^\circ$  since the first measurement in 1580. Figure 4.6 shows the secular variations of declination and inclination recorded in London since the sixteenth century. In general, the geomagnetic field variations are not constant in time and have variable amplitude from place to place.

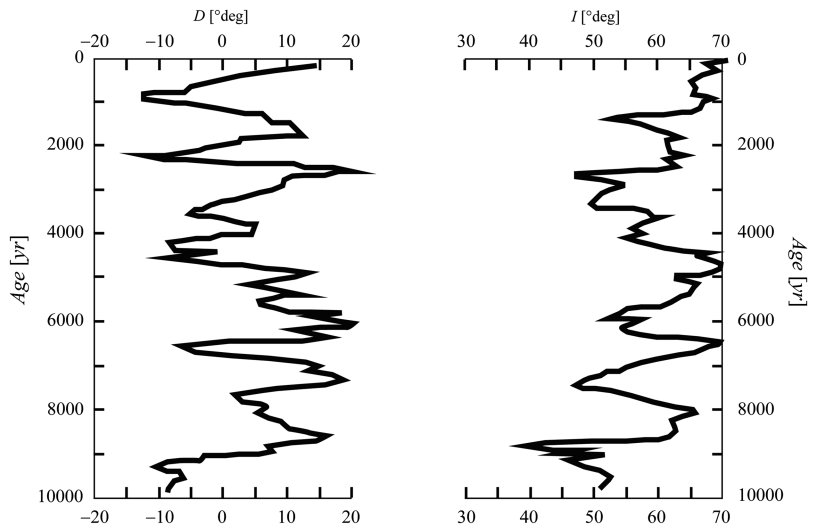
There are three basic phenomena associated with the secular variation of the geomagnetic field, which affect both the dipolar component, over long time intervals, and the non-dipolar part of the field on the short period. They are:

- (a) A continuous change of direction of the magnetic dipole moment and the consequent migration of the geomagnetic poles;
- (b) Variations of magnetic moment magnitude;
- (c) A westward drift of some non-dipolar features of the field at a rate of  $\sim 0.4^\circ$  longitude per year;



**Fig. 4.6** Historical spline regression curves of declination and inclination for London, England (Data from Malin and Bullard (1981))

**Fig. 4.7** Holocene secular variation, recorded by lacustrine sediments in southeastern Oregon. Ages come from radiocarbon dating (Redrawn from Butler (1992))



The geomagnetic poles' wandering has been observed over a longer time interval than that shown in Fig. 4.6. For example, the archeomagnetism of rapidly deposited lacustrine and cave sediments has given further evidence to the existence of an irregular precession of the geomagnetic dipole axis about the spin axis (e.g., Creer 1977). Figure 4.7 shows the Holocene record of variations of declination and inclination at a site in NW United States. Regarding the field magnitude, archeomagnetic studies have shown that it decreased by 50 % during the last 2,000 years, at a rate of  $\sim 6.3$  % per century. In particular, the magnetic moment was  $\sim 9.4 \times 10^{22}$  Am<sup>2</sup> in 1600 ac and decreased almost linearly to  $\sim 7.9 \times 10^{22}$  Am<sup>2</sup> in 1980 ac.

If this rate were maintained, the magnetic moment would go to zero within 1,600 years! Finally, the westward migration of some magnetic features (such as declination contour lines) was observed for the first time by Halley in 1683 around the Atlantic Ocean. However, although the phenomenon seems to involve some independent variables, it is *not* observed in the Pacific region, thereby, it could be associated with a process that operates at regional scale.

The discovery of the secular variation of the geomagnetic field had an indirect but dramatic consequence for the success of the plate tectonics paradigm during the 1960s. The apparently random wandering of the dipole axis about the Earth's spin axis, along with the recognition that

the centered magnetic dipole represents  $\sim 90\%$  of the field intensity, led to the hypothesis that the *time-averaged geomagnetic field* could be approximated by a *geocentric axial dipole* (GAD). In this hypothesis, if we consider a statistically significant data set of  $N$  magnetic field directions,  $F_i$ , spanning several thousand years, the *declination anomaly*,  $\Delta D$ , will be approximately zero:

$$\Delta D = \frac{\sum_{i=1}^N w_i D_i}{\sum_{i=1}^N w_i} \approx 0 \quad (4.47)$$

where the  $w_i$  are weighting factors depending from measurement uncertainties and the  $D_i$  are observed declinations (see McElhinny et al. 1996). In fact, if we assume a magnetic moment aligned with the Earth's rotation axis (as in Fig. 3.7), the theoretical declination of the field vectors will be zero everywhere. Therefore, the time-averaged declination of a large data set should also be close to zero.

However, differently from the example of Fig. 3.7, a representation of the present day time-averaged geomagnetic field by a centered axial dipole requires a *southward* directed magnetic moment. Therefore, by (3.22) and (3.23) we see

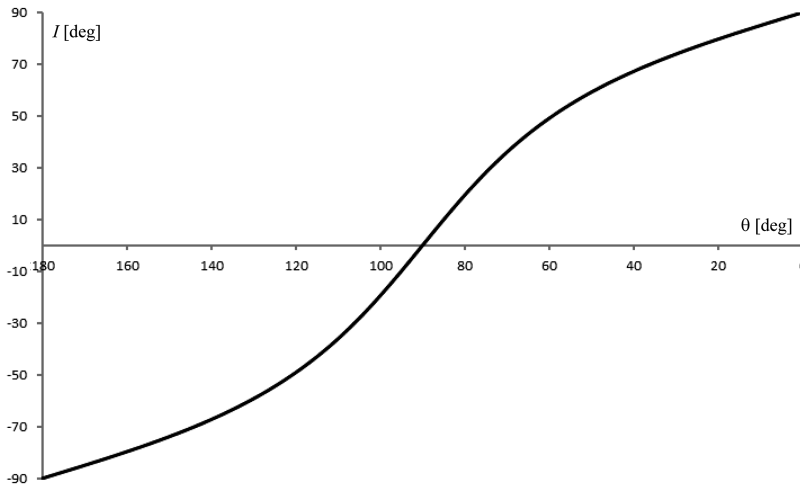
that in this instance the horizontal and vertical components of  $F$  are given by:

$$\begin{cases} Z(r, \theta) = -B_r(r, \theta) = \frac{\mu_0 m \cos \theta}{2\pi r^3} \\ H(r, \theta) = -B_\theta(r, \theta) = \frac{\mu_0 m \sin \theta}{4\pi r^3} \end{cases} \quad (4.48)$$

where  $m$  is the dipole magnetic moment and  $(r, \theta)$  are spherical coordinates (distance from the Earth's centre and colatitude). Thus, the inclination  $I$  will be given by:

$$\tan I = \frac{Z}{H} = 2 \cot \theta \quad (4.49)$$

This is the famous *dipole equation*, which allows to convert a magnetic inclination into a geographic colatitude and vice versa. In Chap. 6, we shall appreciate the importance of this equation for plate tectonics. For the moment, we can only anticipate that it allows to convert the magnetic inclination observed on an ancient rock sample into a *paleolatitude* of the site from which the sample was obtained. Equation (4.49) shows that the inclination of the field vector associated with a GAD field depends in a simple way from the site colatitude only. A plot of  $I$  versus  $\theta$  is shown in Fig. 4.8. Therefore, if we take the time average of a set of field directions, observed in



**Fig. 4.8** Inclination of GAD field versus site colatitude

a narrow colatitude band about  $\theta$ , in the GAD hypothesis also the *inclination anomaly*,  $\Delta I(\theta)$ , will be approximately zero:

$$\Delta I(\theta) = I - \tan^{-1}(2 \cot \theta) \approx 0 \quad (4.50)$$

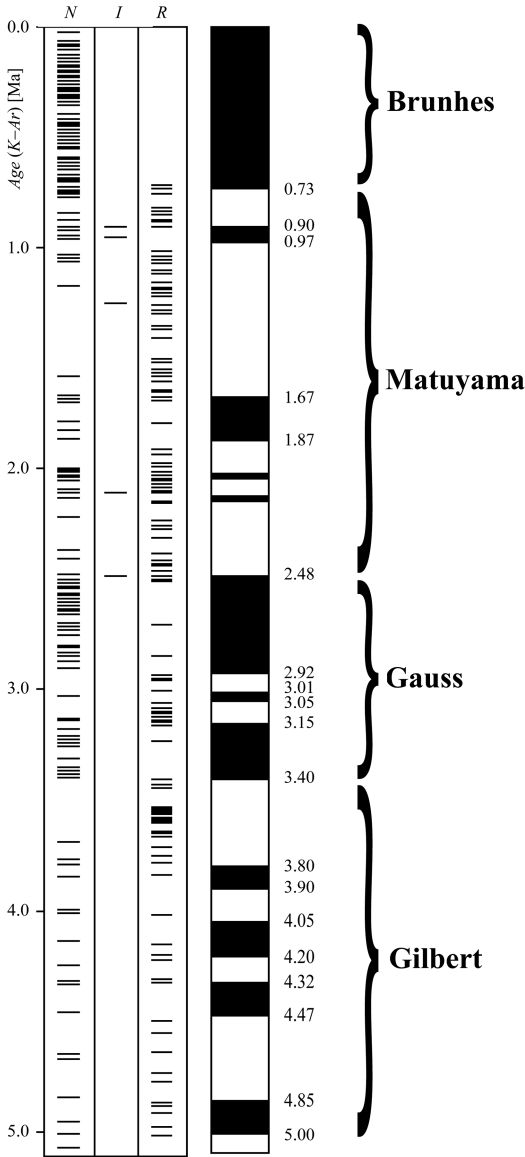
where  $I$  is the inclination of the time-averaged field vector  $\overline{\mathbf{F}}$ . To date, many studies have shown that the time-averaged geomagnetic field is essentially a GAD field, possibly with small additional non-dipole axial components that will be considered in Chap. 6. Therefore, the average position of the geomagnetic poles over a time interval of the order of hundreds of thousands of years is indistinguishable from the geographic poles. In general, on a time scale of thousands of years, the magnetic moment wobbles (so that the field inclination changes), precesses, and changes its intensity (e.g., Merrill and McFadden 2003). The characteristic periodicity of these processes is several times longer than that of the non-dipolar component. For an in-depth discussion about the secular variation of the geomagnetic field the reader is referred to the books of Butler (1992) and McElhinny and McFadden (2000), and to the paper of McElhinny et al. (1996).

#### 4.4 Polarity Inversions, Chrons, and Geomagnetic Timescales

On a time scale larger than that associated with secular variation, we observe another phenomenon that has had a dramatic impact for the construction of the plate tectonics paradigm. It consists into the recurrent inversion of the GAD field polarity, with a periodicity between  $10^4$  and  $10^8$  years. A time interval such that the GAD magnetic moment has the same direction as the present day field, pointing southward, is said to be of *normal polarity*, while the opposite configuration defines a *reversed polarity* time interval. Although magnetic reversals determine a change of declination of  $180^\circ$  at any point on the Earth's surface, the fact that an observed paleomagnetic field direction at a site is approximately opposite to the present day

field direction could not be sufficient to prove that the rock formed during a time interval of reversed polarity. In fact, it is known that the non-dipole field components can produce large local deviations of the observed declination values. Therefore, the evidence for magnetic polarity reversals must be based upon many independent observations distributed over the Earth's surface.

The recognition of the existence of recurrent events of field inversion led to the development of *geomagnetic polarity time scales* since the 1960s. In these time scales, a time interval with constant magnetic polarity (normal or inverted) is called *chron* and the time boundaries are established on the basis of radiometric dating. The first time scales had a rather restricted temporal range, which was based on potassium–argon (K–Ar) dating of Pliocene and Pleistocene igneous rocks. Therefore, they spanned the interval 0–5 Ma. An example of this class of polarity time scales is illustrated in Fig. 4.9. This time scale was built using 354 igneous rock samples, from which K–Ar ages and magnetic polarity had been determined. Figure 4.9 shows that the average duration of the chrons was  $\sim 0.25$  Myrs since 5 Ma. The number of observations that were classified as “intermediate polarity” was  $\sim 1.5\%$  of the sample. These field directions were probably acquired during the short phases of switching of the GAD polarity at the transition between two adjacent chrons. It is estimated that these phases have duration between 1,000 and 8,000 years (see McElhinny and McFadden 2000, and references therein). Therefore, in terms of geological time they are global synchronous events. An important feature of the alternate sequence of polarity chrons in Fig. 4.9 is represented by their variable length, so that the specific *pattern* of normal and reversed chrons for a given time interval can be used as a distinctive “fingerprint” that identifies the interval. We shall make extensive usage of this feature in the procedure of identification of marine magnetic anomalies (Chap. 5), which is the primary tool for dating the oceanic sea floor. In the case of the time scale 0–5 Ma, historically four major *magnetic polarity epochs* were identified, along with shorter intervals that were called *events*. These epochs, to which was given the name of

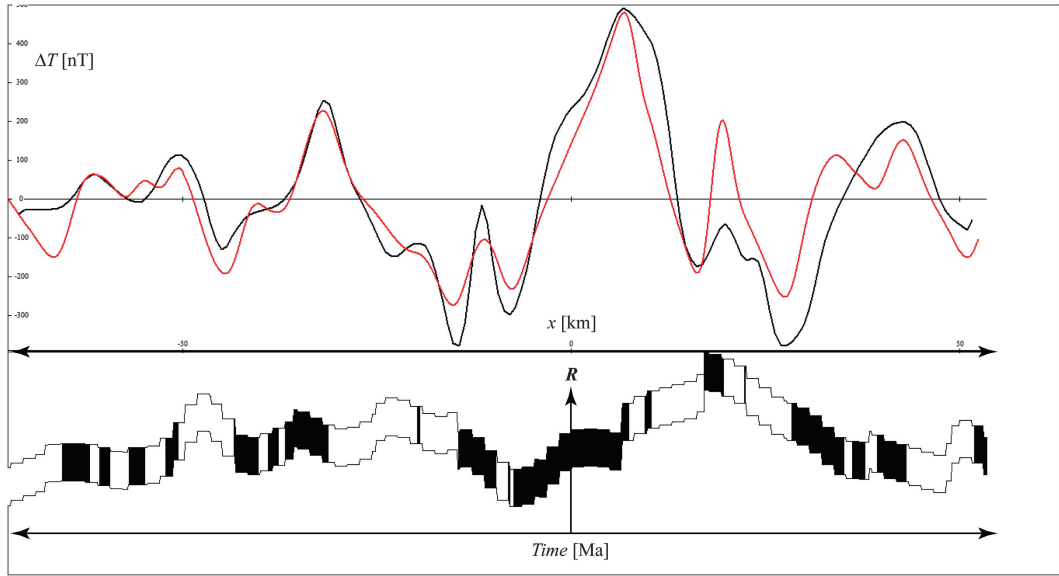


**Fig. 4.9** Magnetic polarity time scale of Mankinen and Dalrymple (1979). Columns *N*, *I*, and *R* show respectively the occurrence of samples with normal, intermediate, and inverse polarity. Normal and reversed polarity intervals are shown, respectively, in *black* and *white*. Numbers are K–Ar radiometric ages (in Ma)

pioneering scientists in geomagnetism (*Brunhes*, *Matuyama*, *Gauss*, and *Gilbert*), are shown in Fig. 4.9. Although today this nomenclature is considered obsolete, it is still used by many geologists. The extension of the geomagnetic polarity time scale to times older than the Pliocene

required a different technique, because the uncertainty of radiometric dating increases with the age of the samples and approaches the typical duration of chrons. The method used today for the time interval between the Miocene and the middle Jurassic was proposed for the first time by Heirtzler et al. (1968) and is based on the analysis of marine magnetic anomalies from the World’s oceans. The technique can be described as follows. As new oceanic crust forms at a spreading ridge by cooling of MORB, it acquires a TRM having the same direction of the ambient field at that time (see Sect. 3.5). Therefore, when the main field is subject to a polarity reversal, this event is recorded in the oceanic crust, which acquires an inverted polarity with respect to the previous chron. This implies that the magnetization pattern of the World’s oceans is formed by long parallel stripes having alternate polarity, which move away from spreading ridges.

This structure was described for the first time by Vine and Matthews (1963) in a seminal paper that can be considered as a milestone in the development of the plate tectonics theory. We shall see in Chap. 5 that the alternated arrangement of stripes with opposite TRM polarity can be detected by the analysis of the magnetic anomalies that are produced by the “anomalous” field associated with crustal magnetization. Figure 4.10 shows an example of magnetization pattern in the central Atlantic and the associated magnetic anomalies. The technique used by Heirtzler et al. (1968) to build a geomagnetic polarity time scale for times older than 5 Ma is a milestone in the history of marine geophysics. These authors started from some ship-track magnetic profiles acquired in the North and South Pacific, in the South Indian Ocean, and in the South Atlantic. For each magnetic profile, they built a magnetization model such that the predicted magnetic anomalies matched the observed profiles. An example of this procedure is shown in Fig. 4.10. In order to fit the theoretical and observed profiles, they changed repeatedly the width of the blocks, until the two curves overlapped. The procedure furnished a sequence of distances  $x_1, x_2, \dots, x_n$  of the blocks from the ridge. These distances are



**Fig. 4.10** An example of sea floor magnetization and magnetic anomaly signal in the central Atlantic. The *black* and *red* lines show respectively the measured and model anomalies, while the *lower panel* shows the resulting

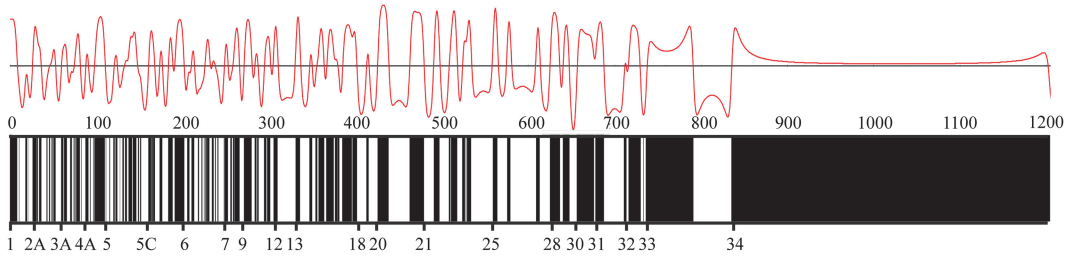
predicted magnetization of the oceanic crust, with the pattern of normal (*black*) and reversed (*white*) blocks. *R* is the ridge

referred to as *anomalies* and are numbered progressively according to a nomenclature that will be described later. Assuming that the spreading velocity  $v$  was constant, Heirtzler and colleagues could have converted promptly the anomalies  $x_i$  into ages of the top of each chron,  $T_i$ , obtaining a geomagnetic polarity time scale. In fact, the scaling formula is simply:

$$T_i = \frac{2x_i}{v} \quad (4.51)$$

However, there was no warranty that the spreading rate had remained constant for any of the magnetic profiles. Heirtzler and colleagues had some *calibration points*, consisting into sea floor locations with known age. For example, they knew that the magnetic basement at anomaly six in the North Pacific and South Atlantic had an age of 20–22 Ma. Therefore, these authors performed a comparative analysis of the magnetic profiles, concluding that the South Pacific spreading rate had varied with time, whereas the South Atlantic and North Pacific velocities might be considered approximately constant. In particular, an  $x_i - x'_i$  plot of South

Atlantic and North Pacific anomalies showed a linear correlation between the two data sets. Consequently, the South Pacific was eliminated as a standard for the construction of a time scale. Finally, they selected the South Atlantic profile as a standard, because of its length and better anomaly pattern. The assumed constant velocity was  $19 \text{ mm year}^{-1}$  for the last 80 Myrs. Testing the predicted ages of the corresponding anomalies was a major objective of the Deep Sea Drilling Project (DSDP) from the late 1960s to the early 1980s. Further biostratigraphic calibration points were provided by magnetostratigraphic investigations of marine sedimentary sequences. The most important of these studies was undoubtedly that performed on the Late Mesozoic and Cenozoic pelagic limestone sequences in the Umbria-Marche Apennine, Italy (Lowrie and Alvarez 1977). The new data allowed the construction of more refined geomagnetic polarity time scales, notably the Mesozoic time scale of Gradstein et al. (1994), the Late Cretaceous – Cenozoic time scale of Cande and Kent (1995), and the more recent Phanerozoic time scale of Gradstein et al. (2004).



**Fig. 4.11** C-series geomagnetic polarity time scale of Cande and Kent (1995). Only the major anomalies are labeled. Numbers on the upper scale are distances from the ridge when the spreading rate is  $20 \text{ mm year}^{-1}$ . The red line shows the theoretical magnetic signal associated with this time scale in the case of an N-S profile in the northern hemisphere. Ages of the labeled anomalies

are: 2A = 2.58 Ma, 3A = 5.89 Ma, 4A = 8.70 Ma, 5 = 10.95 Ma, 5C = 16.01 Ma, 6 = 20.13 Ma, 7 = 24.73 Ma, 9 = 27.03 Ma, 12 = 30.48 Ma, 13 = 33.06 Ma, 18 = 40.13 Ma, 20 = 42.54 Ma, 21 = 47.91 Ma, 25 = 55.90 Ma, 28 = 62.50 Ma, 30 = 65.58 Ma, 31 = 67.74 Ma, 32 = 71.07 Ma, 33 = 73.62 Ma, and 34 = 83.50 Ma

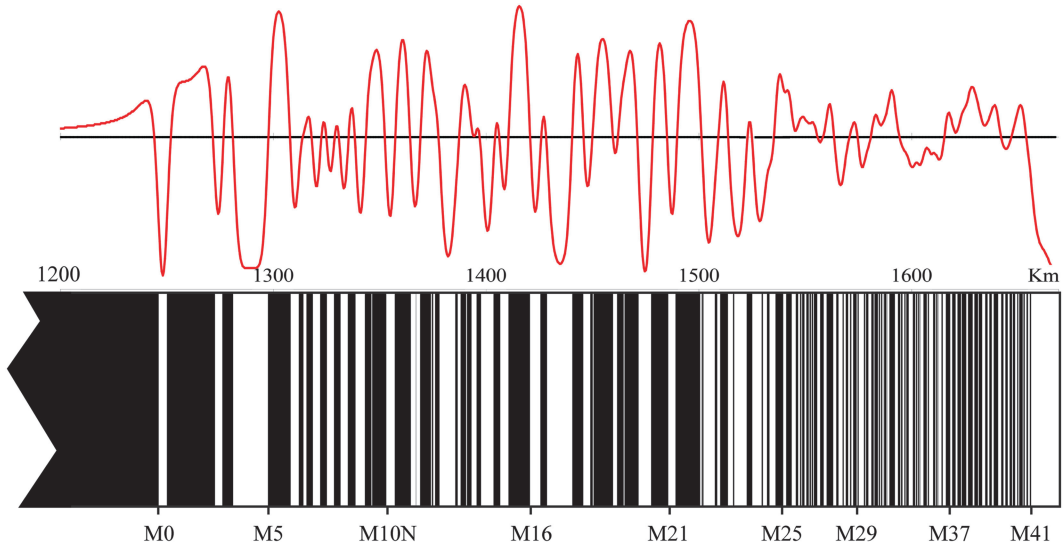
In these time scales, the polarity chrons for latest Cretaceous and Cenozoic are numbered within the *C-series*, from C34 (the oldest) to C1 (the youngest). For older times, the middle Jurassic through Early Cretaceous sequence, which is used in marine geophysics studies, consists of the *M-series* (polarity chrons from M0 to M41). This segment of the time scale is based upon the assumption that the Hawaiian sequence of marine magnetic anomalies in the western Pacific formed at constant spreading rate (Gradstein et al. 1994). Also in this case the sequence was pinned to tie points with known radiometric age.

The naming conventions of magnetochronology are simple. For each normal polarity chron, there is a corresponding older chron with reversed polarity. Then, a suffix “n” or “r” is used to distinguish the two time intervals. For instance, C13n identifies the normal polarity chron within the interval C13. A complication arises from the presence in the *C-series* of smaller intervals with inverted polarity within a chron. For example, C2r starts at 2.581 Ma and ends at 1.950 Ma in the scale of Cande and Kent (1995). This reversed chron includes a short normal polarity sub-chron between 2.140 and 2.150 Ma, which was discovered later. Therefore, it is now divided into three sub-chrons, named respectively C2r.1r, C2r.1n, and C2r.2r. Figures 4.11 and 4.12 show respectively the geomagnetic polarity time scales for the *C*- and *M*-series.

An important feature of the late Cretaceous – Cenozoic geomagnetic polarity time scale is the presence of a long chron with normal polarity (C34), from  $\sim 120$  to  $\sim 84$  Ma (Fig. 4.11). This is called the *Cretaceous normal-polarity super-chron* or, alternatively, *Cretaceous Quiet Zone* (CQZ). During this time interval, which lasted  $\sim 36$  Myrs, the geomagnetic polarity remained fixed. Another interesting feature can be observed on the Jurassic – early Cretaceous time scale (Fig. 4.12). It is represented by a long sequence of short polarity chrons before M25, between the Oxfordian and the Callovian, which determines blurring of the magnetic signal. This time interval partly coincides with the so-called *Jurassic Quiet Zone* (JQZ), which is observed offshore North America, along the coast of northwest Africa, and in the western Pacific. This zone includes anomalies older than M29 and is characterized by the low amplitude of the magnetic signal, which makes the analysis difficult.

## 4.5 Ionosphere and Magnetosphere

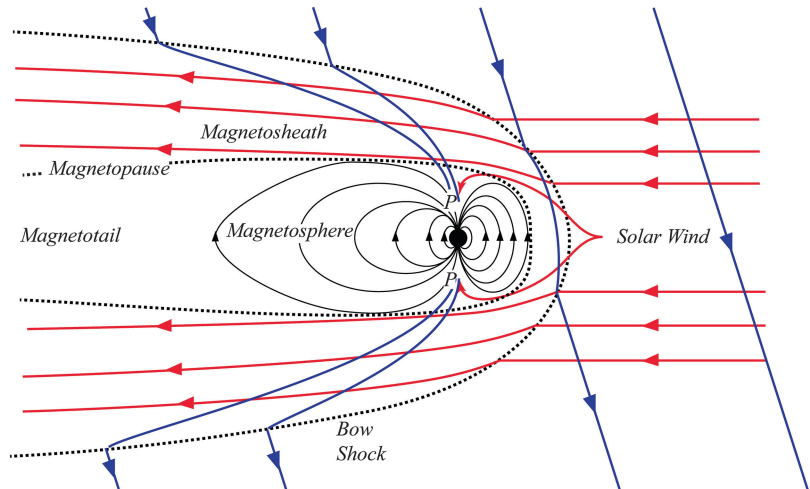
The large region around the Earth where the geomagnetic field dominates is called *magnetosphere*. It is occupied by *plasma*, that is, high-energy charged particles, of solar wind provenance or terrestrial origin. The *solar wind* is a



**Fig. 4.12** *M*-series segment of the geomagnetic polarity time scale of Gradstein et al. (2004). Only the major anomalies are labeled. Numbers on the upper scale are distances from the ridge when the spreading rate is  $20 \text{ mm year}^{-1}$ . The red line shows the theoretical magnetic signal associated with this

time scale in the case of an N–S profile in the northern hemisphere. Ages of the labeled anomalies are: M0 = 124.61 Ma, M5 = 129.76 Ma, M10N = 135.28 Ma, M16 = 142.06 Ma, M21 = 148.54 Ma, M25 = 154.37 Ma, M29 = 157.51 Ma, M37 = 162.04 Ma, and M41 = 165.61 Ma

**Fig. 4.13** Earth's magnetosphere and interplanetary magnetic field (IMF). Solar wind streamlines are shown in red. Points *P* are polar cusps. Field lines of the Earth's magnetic field and the IMF are shown respectively by solid black and blue lines



highly conductive plasma flow ejected by the Sun, which travels at a velocity of  $\sim 500 \text{ km/s}$  as a result of the supersonic expansion of the solar corona. High conductivity of this material implies that the magnetic field flow lines are frozen into the solar wind plasma (see Sect. 4.1). Therefore, in the region where the charged particles strike against the Earth's magnetic field, the particles slow down and to a large extent are deflected

without penetrating the magnetic shield. At the same time, the flow lines of the Earth's magnetic field are compressed in the dayside region and stretched out along the nightside area, generating a wide *magnetotail* (Fig. 4.13). As a consequence, the Earth's magnetosphere is a magnetic cavity shaped like a comet head and tail, and only at distances less than  $5R_e$ ,  $R_e$  being the Earth's radius, it can be considered approximately dipolar.

A small portion of solar wind plasma enters the magnetosphere along the dayside region near the *polar cusps*, two funnel-shaped areas between the dayside and the magnetotail. Here, the solar plasma particles follow the magnetic field lines towards the Earth's atmosphere (Fig. 4.13).

Three major regions can be distinguished in the space around the Earth. In the outer space, the *interplanetary magnetic field* (IMF) has sub parallel field lines, and the solar wind streamlines are straight lines toward the Earth. The *bow shock* is the shock surface where the solar wind is suddenly slowed from supersonic to sonic velocities. It marks the transition to turbulent flow in the *magnetosheath*, the intermediate region where IMF geometry and solar wind streamlines are affected by the presence of the Earth's magnetic field. The *magnetopause* is the inner boundary that separates the magnetosphere from the outer region. Along the magnetopause, the inward *dynamic pressure* of the advancing solar wind plasma is balanced by the outward *magnetic pressure* of the magnetosphere. To understand this important concept, let us consider again the magnetic field inside a long solenoid (Fig. 4.4), and let us assume that the coil is infinitely thin with square cross-section of dimension  $h$ . In this instance, the current through the solenoid is related to the current density by:  $I = jh^2$ . If  $n$  is the number of turns per unit length, then  $h = 1/n$ . Therefore, the magnitude of the current density in the solenoid is given by:

$$j = In^2 \quad (4.52)$$

We know that the magnetic field  $B$  is zero outside the solenoid and assumes the uniform value (4.34) in the internal region. The region occupied by the coil is a transition zone where the average field is given by:

$$\langle B \rangle = \frac{1}{2}B \cong \frac{1}{2}\mu_0 nI \quad (4.53)$$

The presence of a magnetic field within the coil generates a Lorentz force that can be calculated using (4.20). To evaluate the cross prod-

uct in (4.20) we take into account that the current density vector is always tangent to the coil. Therefore, the Lorentz force per unit volume,  $\mathbf{f}$ , is directed radially outwards and has magnitude:

$$f = I \langle B \rangle n^2 = \frac{2\langle B \rangle^2 n}{\mu_0} = \frac{B^2 n}{2\mu_0} \quad (4.54)$$

where we have used (4.53). If  $L$  and  $r$  are respectively length and radius of the solenoid, then the volume of a small slice is  $dV = rLh d\alpha$ , where  $d\alpha$  is an arc element of the coil circumference. The area of its internal surface will be  $dA = rL d\alpha$ . Therefore, the force on the volume element is:

$$dF = f dV = \frac{B^2 r L d\alpha}{2\mu_0} \quad (4.55)$$

This force determines an outward *pressure* on the internal surface of the solenoid, which will be given by the force per unit area:

$$P_m = \frac{dF}{dA} = \frac{B^2}{2\mu_0} \quad (4.56)$$

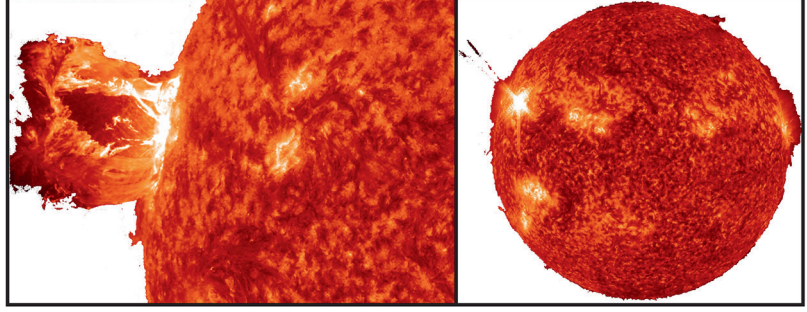
The quantity  $P_m$  is called *magnetic pressure* and can be quite effective. For example, for  $B = 1$  T, we have  $P_m \cong 4.0 \times 10^5$  Pa = 4 bar. Now let us consider the *dynamic pressure*, which for any fluid is defined as the *density* of kinetic energy:

$$P_d = \frac{1}{2}\rho_m u^2 \quad (4.57)$$

where  $\rho_m$  is the mass density and  $u$  is the fluid velocity. This quantity contributes to the total fluid pressure just like the usual hydrostatic pressure. The solar wind is a neutral fluid formed by positive ions and electrons. For a flow dominated by protons and electrons, with a typical density of five protons per  $\text{cm}^3$  and an average velocity  $u \sim 500$  km  $\text{s}^{-1}$ , it results:

$$\begin{aligned} P_d &= \frac{1}{2} (5\text{cm}^{-3}) \times (1.67 \times 10^{-27}\text{kg}) \\ &\times (500\text{km s}^{-1})^2 \cong 1.04 \times 10^{-9} \text{ Pa} \end{aligned} \quad (4.58)$$

**Fig. 4.14** Examples of Solar eruptions. *Left*: Coronal mass ejection on April 16, 2012. *Right*: X 1.9 class solar flare on November 3, 2011 (Credit: NASA Solar Dynamics Observatory)



At the magnetopause, the magnetic and dynamic pressures balance out. Therefore, we can use (4.58) and (4.56) to estimate the distance of the magnetopause from the Earth's centre. To this purpose, we approximate the Earth's magnetic field by a GAD dipole (Eq. 3.24). In the equatorial plane, where the value (4.58) is more appropriate, the magnitude  $B$  is given by:

$$B(r) = \frac{\mu_0 m}{4\pi r^3} \cong 7.9 \times 10^{15} \frac{1}{r^3} \quad (4.59)$$

where we have assumed that  $m = 7.9 \times 10^{22} \text{ Am}^2$ . Let  $R_p$  be the distance of the magnetopause from the Earth's centre in the dayside equatorial plane. Equating (4.58) with (4.56) gives an estimate of this quantity:

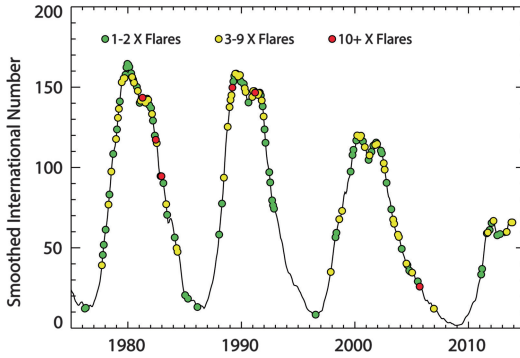
$$R_p = \left( \frac{m^2 \times 10^{-7}}{8\pi P_d} \right)^{1/6} \cong 8.5 R_e \quad (4.60)$$

where  $R_e \cong 6,371 \text{ km}$  is the Earth's radius. This is the upstream distance of the magnetopause in normal conditions. However, strong solar wind conditions can push the magnetopause well inside the geostationary orbit of satellites ( $\sim 6.6 R_e$ ). In the nightside region, the magnetopause is on average  $\sim 30 R_e$  from the ecliptic plane.

The importance for plate tectonics practitioners to have a basic understanding of the magnetosphere and related processes arises from the influence that short-period geomagnetic field time variations have on the measurement of magnetic anomalies produced by sea floor crustal magnetization. Marine geophysics campaigns require precise determinations of the total geomagnetic field strength of internal origin (which includes

core and crustal components), thereby, the presence of strong external field components impedes a correct application of the method of calculation of magnetic anomalies that will be described later in this chapter. The strongest source of geomagnetic field disturbance is represented by *geomagnetic storms*. These events are associated with large sudden variations of solar wind dynamic pressure at the magnetopause, which follow the impact of *coronal mass ejections* (CME) and *solar flare* particles (Fig. 4.14).

A CME is a form of extensive and explosive solar mass release that produces strong perturbations of the solar wind, which reaches speeds as high as  $2,800 \text{ km s}^{-1}$  during these events. A solar flare is a more local event than CMEs, which produces flashes of light for short time intervals ranging from a few minutes to a few hours (Schunk and Nagy 2009). These explosions can send bursts of energetic particles into the solar wind, determining magnetic storms. A storm results from compression of the magnetosphere due to the arrival of the shockwave and can be particularly strong when the increased solar wind pressure is associated with a large southward IMF component (Yokoyama and Kamide 1997). The typical time that a CME takes to reach the Earth is 2–3 days. The resulting geomagnetic field disturbance can have serious consequences for the human electric infrastructures and networks, and it is hardly a good idea to take magnetic measurements during a storm. At mid-latitudes, about one storm per year produces an external field whose horizontal component  $H > 250 \text{ nT}$ , and about ten storms per year have  $H > 50 \text{ nT}$  (Campbell 2003). Some events can produce extremely strong variations of geomagnetic field



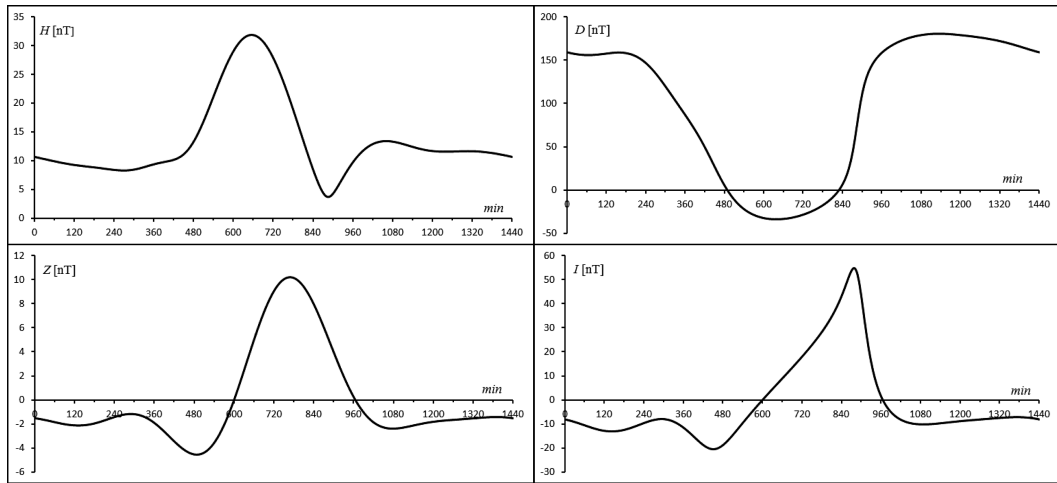
**Fig. 4.15** Monthly X-class flares and International Sunspot Number and 11-year solar cycle. The plot shows the increased number of X-class flares during the peaks of solar activity (higher number of sunspots). However, flares can also occur when the sunspot cycle is at the minimum (Original figure in Hathaway 2010, updated to 2014 by courtesy of the author)

strength. Tsurutani et al. (2003) report 12 large magnetic storms from 1859 to 1989, with  $H$  range (difference between the maximum and minimum of  $H$  during the storm) varying from 450 to 1,720 nT. The duration of the main phase of these events is of the order of a few hours. In general, during the main phase the magnetic field at the Earth's surface is significantly depressed. This depression has a strong latitudinal dependence and is determined by the injection of magnetotail plasma into the nightside magnetosphere. Within the inner magnetosphere, the Lorentz force drives the high-energy protons to the west, whereas electrons move eastward, thus feeding and strengthen a current loop around the Earth that is known as the *ring current*. This current determines the formation of a northward directed magnetic moment that opposes the Earth's dipole moment, thus decreasing the net magnitude of the geomagnetic field measured at near-equatorial magnetic stations.

Regarding the recurrence interval of geomagnetic storms, it is strictly related to the *solar cycle*, which consists into a periodic change of solar activity with average duration of  $\sim 11$  years (Fig. 4.15). The solar cycle is a magnetic phenomenon that originates by dynamo processes within the Sun. During an 11-year cycle, it is possible to observe a rise and fall in the number

of solar flares, CMEs, and other visible changes (sunspots). To measure the solar activity and the corresponding geomagnetic disturbance, several indices have been designed. In marine geophysics, the most useful measure of magnetic activity is the global 3-h-range  $Kp$  index. This index is determined as the arithmetic mean of  $K$  index values from 13 different geomagnetic observatories. The  $K$  index is a quasi-logarithmic scale that quantifies the disturbance in the  $H$  component of the geomagnetic field in a single observatory. It is calculated from the amplitude range of  $H$  in a 4-h interval, after removal of diurnal and secular variations by fitting a smooth curve to the observed values. The  $Kp$  scale ranges between 0 and 9, and is expressed in thirds of a unit: 0, 1/3, 2/3, 1, 4/3, ... These quantities are usually indicated as: 0,0+, 1-, 1,1+, etc. Usually,  $Kp \leq 2+$  indicates quietness, whereas  $Kp \geq 5$  is indicative of a geomagnetic storm.

A less intense but still significant contribution to the external geomagnetic field comes from currents driven by tidal forces and winds circulating in the Earth's *ionosphere*, a conducting region of the terrestrial atmosphere where a large density of free low-energy electrons and ions is present. This region extends upwards from 70 to 80 km elevation to considerable altitudes, where it merges with the magnetosphere. The free electrons and ions that form the ionosphere plasma are produced by ionization of neutral atoms, as a consequence of both extreme ultraviolet radiation from the Sun and collisions with high-energy particles that penetrate the atmosphere. The ionosphere currents are revealed by a recurrent pattern on observatory magnetograms during quiet days, with characteristic 24/ $k$  hour period spectral components, where  $k = 1, 2, 3, 4$  (Campbell 2003). Therefore, the typical magnetogram pattern associated with these currents is referred to as *quiet daily variations*. When the small but persistent component arising from tidal forces is removed, the changes are called *Sq variations* (i.e., solar quiet). A theoretical plot of *Sq* components is illustrated in Fig. 4.16. The *Sq* field varies slowly through the months of the year and its contribution to magnetic measurements changes with the geographic location, but the



**Fig. 4.16** Theoretical 24-h plot of Sq field components at  $0^{\circ}\text{N}$ ,  $15^{\circ}\text{E}$ , calculated using the SQ1 model of Campbell et al. (1989). The horizontal axis shows local time (in min after midnight) on June 1st 2005

prevalence of the 24, 12, 8, and 6 h harmonics shows that the driving mechanism of the source currents depends strongly from the Earth's rotation. In fact, it is known that this field results from enhancement of conductivity of the  $E$  region of the ionosphere, an intermediate layer between 90 and 140 km altitude, induced by solar radiation (e.g., Hitchman et al. 1998).

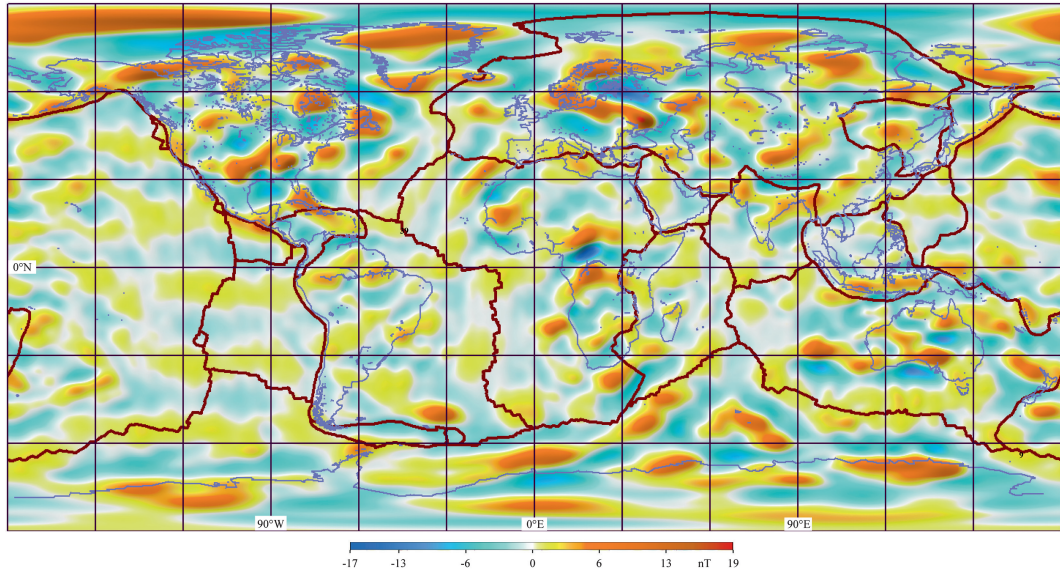
In general, at high latitudes the  $Sq$  field does not give a significant contribution to the external field, whereas at the magnetic equator the highest magnitude is of a few tens of nT.

## 4.6 Crustal Magnetic Field

The magnetization of crustal rocks represents an important source for the Earth's internal field, originating near-surface anomalies between  $-3,600$  and  $+8,500$  nT. Conversely, both the mantle and the unconsolidated sediments are essentially non-magnetic. The crustal sources are located in regions where the temperature field is below the Curie point of the magnetic minerals (see Sect. 3.5). Therefore, the lower boundary of the so-called *magnetic basement*, which is the region of crustal rocks having relatively large magnetic susceptibilities, approximately coincides with the Curie isotherm for magnetite.

This isotherm is typically more than 20 km depth in stable continental regions, but may be as shallow as 2 km in young oceanic regions. Although both remnant and induced magnetization contribute to the crustal field, only the former is important for plate kinematics. In paleomagnetic studies performed on continents, the remnant components of magnetization are isolated directly on rock samples through laboratory procedures. Conversely, in the case of marine geophysics, it is not a simple task to separate these components from a data set of total field measurements, thereby, our interpretations often rely on the hypothesis (more or less justified) that the induced component is small. In general, igneous rocks have the highest Koenigsberger ratios (often between 5 and 50, see Table 3.1), and in the important case of MORBs  $Q$  may reach values as high as 160. Therefore, in the oceanic regions, where it is not generally possible to separate the remnant magnetization from the induced component, the crustal field can be considered with good approximation as the product of *time independent* TRM at the historical scale.

In most places, the crustal field is less than 1 % of the total magnetic field, but locally may represent up to 20 % of the observed field. The first global maps of the crustal field were built



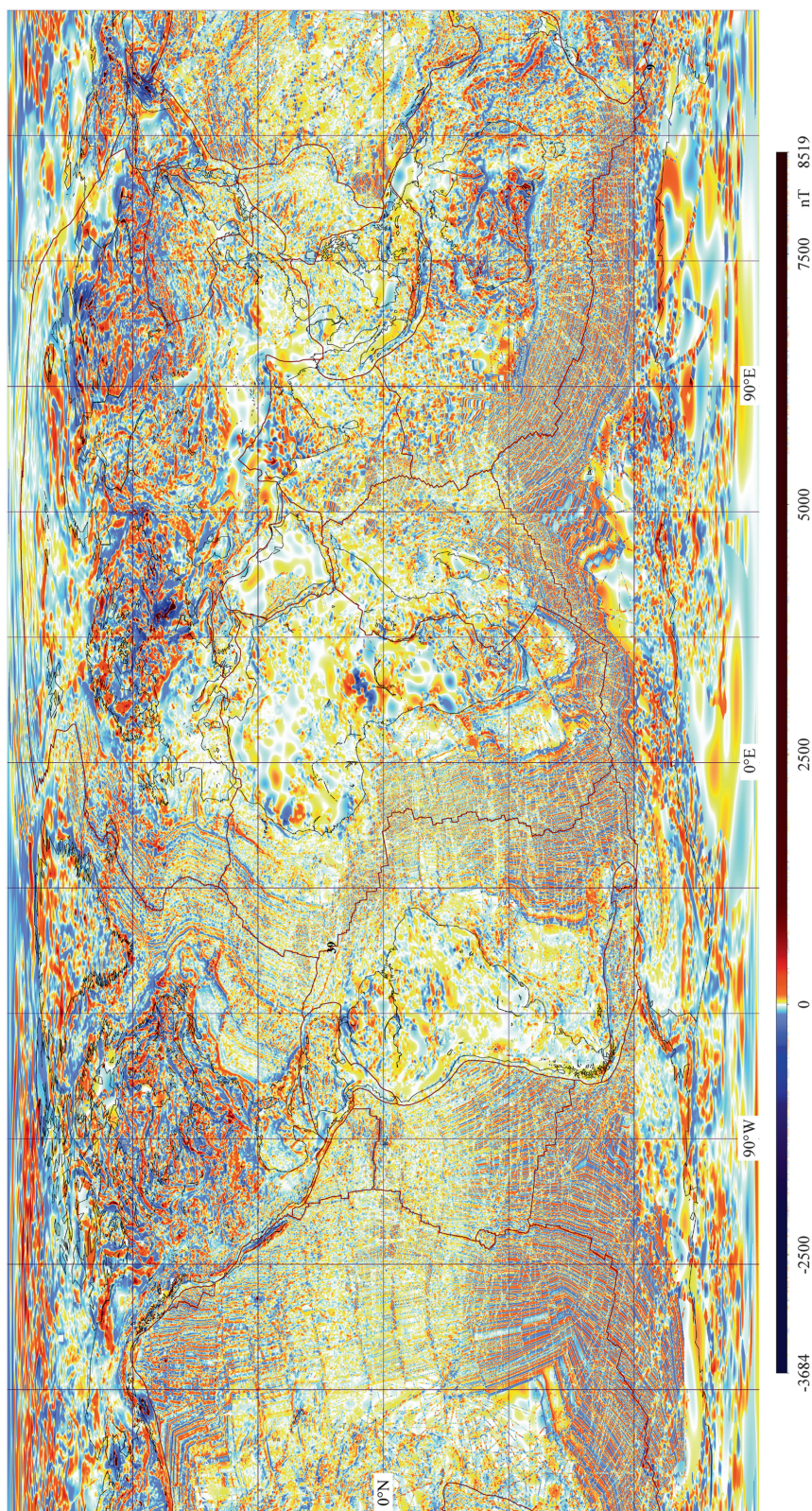
**Fig. 4.17** Magnetic field intensity anomaly,  $\Delta F$ , caused by the crustal magnetic field at a satellite altitude of 400 km above the Earth's surface, as given by the MF3 model (Maus et al. 2006)

on the basis of POGO (Polar Orbiting Geophysical Observatory) and MAGSAT (Magnetic Field Satellite) low-orbiting satellite measurements of magnetic intensity since the 1960s (Maus et al. 2006 and references therein). The first data were scalar total field magnitudes, whereas starting from MAGSAT the satellites were also equipped with vector magnetometers. One of the most recent data sets comes from the CHAMP (CHALLENGING Minisatellite Payload) mission, a German satellite mission for geologic and atmospheric research that started in July 2000. An example of crustal field model based on these data is shown in Fig. 4.17. This model provides long wavelength features of the crustal component of the internal field, which can be useful in studies of global scale tectonics.

Maps of crustal magnetic field on continents can be used in conjunction with gravity and geological maps to identify tectonic provinces, dikes, faults, and any other geologic feature having a magnetization contrast with the surrounding. In recent years, the World Digital Magnetic Anomaly.

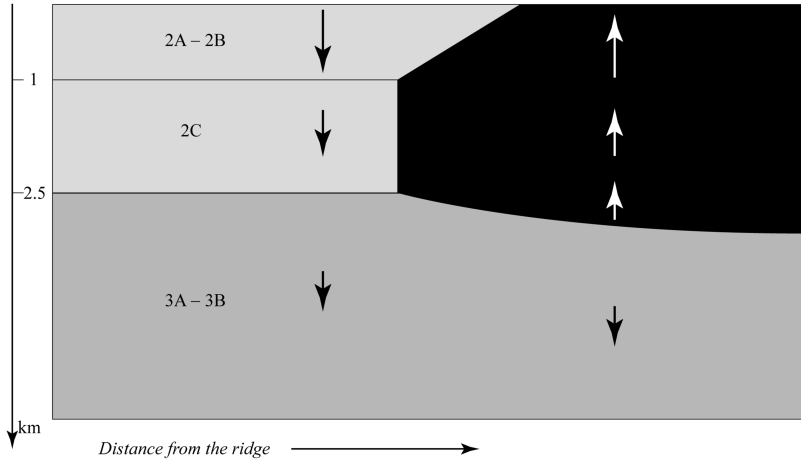
Map (WDMAM) has been compiled under the auspices of the International Association for Geomagnetism and Aeronomy (IAGA). This map,

which is shown in Fig. 4.18 combines ground, airborne, and marine magnetic data and includes all the wavelengths that could be useful for the geological and tectonic mapping of the Earth's crust. We have already stressed the fact that ocean floor magnetization represents a primary source of data in plate tectonics. In Sect. 1.3 we have shown that the oceanic crust has a layered structure (Fig. 1.6), the 2A pillow lavas being the uppermost of the igneous layers. These submarine basalts cool quickly from high temperatures in the seawater environment. Consequently, their titanomagnetite content is fine-grained and carries an intense TRM. This mineral has an average composition  $\text{Fe}_{2.4}\text{Ti}_{0.6}\text{O}_4$  and Curie point in the range 150–200 °C (McElhinny and McFadden 2000). We know that the geothermal gradient close to mid-ocean ridges is  $\sim 100$  °C/km, while for a 120 Ma old oceanic crust it is reduced to  $\sim 10$  °C/km (Chap. 12). Therefore, the depth to the Curie isotherm is very shallow near the spreading centers and increases with the distance from the ridge. Regarding the lower layers (2B, 2C, and 3), they cool slowly and have Curie points generally higher than 500 °C, but the magnetization is at least one order of magnitude less than the overlying pillow lavas. It is also



**Fig. 4.18** The World Digital Magnetic Anomaly Map 2007 (Korhonen et al. 2007)

**Fig. 4.19** Magnetic polarity boundaries of the oceanic crust, based on Tivey (1996). Arrows are directions of magnetization. The black and grey regions have respectively normal and reversed polarity



important to note that in so far as the new oceanic crust moves away from the ridge, the cooling rate of rocks at any given depth decreases, as we shall see in Chap. 14. Therefore, the geometry of the boundaries of blocks with opposite polarity is determined by the balance between cooling rate and spreading velocity, so that they are not vertical but curved towards or away from the ridge. In most cases, the magnetic boundaries of the rapidly cooling extrusive 2A layer dip toward the spreading axis. Then, the boundaries become nearly vertical in the 2C dikes layer and gently dip away from the spreading center in the gabbro section (3A–3B) (Fig. 4.19).

An important feature of the magnetization pattern in the oceanic crust is represented by the strong magnetization of the normal polarity axial blocks associated with the present day chron (C1n), which is generally two to three times that of adjacent blocks ( $10\text{--}15 \text{ Am}^{-1}$  instead of  $5\text{--}6 \text{ Am}^{-1}$ ). McElhinny and McFadden (2000) report a time constant of magnetization decay (the time for the magnetization to reduce to  $1/e$  of its initial value) of  $\sim 20$  kyrs. Towards the continental margins, the magnetization becomes less intense, with typical values of  $M = 3\text{--}4 \text{ Am}^{-1}$ . Furthermore, some of the anomalies observed close to the COBs may be associated with upper mantle serpentinization and not with sea floor spreading (see Sect. 1.3), a common situation along non-volcanic continental margins such as the western Iberian margin. In the next sections, we shall consider in detail the

mathematical properties of the geomagnetic field, a necessary step to create crustal field models that match the observed data.

## 4.7 The Geomagnetic Potential

According to Gauss' law, the Earth's magnetic field,  $\mathbf{B}$ , is solenoidal (Eq. 3.25). Outside the Earth's surface, it is also irrotational, because  $\mathbf{j} = \mathbf{0}$  in (3.27) almost everywhere. Therefore, there exists a scalar field  $V$  such that (3.28) holds. We say that the magnetic field is a *potential field* in the region outside the Earth. In this instance, the scalar geomagnetic potential  $V$  satisfies Laplace's equation (3.29). Now we are going to describe some important mathematical properties of the potential, which will be helpful to fully understand the meaning of geophysical models of crustal magnetization and mass density distributions. These properties are known as *Green identities*.

### Green's First Identity

Let  $\phi = \phi(\mathbf{r})$  and  $\psi = \psi(\mathbf{r})$  be two scalar fields, defined in a closed region  $\mathbf{R}$  with boundary  $S(\mathbf{R})$ . Then:

$$\oint_{S(\mathbf{R})} \phi(\mathbf{r}) \frac{\partial \psi}{\partial n} dS = \int_{\mathbf{R}} \nabla \phi \cdot \nabla \psi dV + \int_{\mathbf{R}} \phi(\mathbf{r}) \nabla^2 \psi dV \quad (4.61)$$

*Proof* Let us consider the vector field  $\mathbf{A}(\mathbf{r}) = \phi(\mathbf{r})\nabla\psi$  and its flux through  $S(\mathbf{R})$ . We remind that the quantity  $\nabla\psi \cdot d\mathbf{S}$  is the directional derivative of  $\psi$  along the normal to  $S$  (see Appendix 1), which is usually indicated by  $\partial\psi/\partial n$ .

Then, by the divergence theorem we have:

$$\begin{aligned} \int_{\mathbf{R}} \nabla \cdot (\phi \nabla \psi) dV &= \int_{\mathbf{R}} [\nabla \phi \cdot \nabla \psi + \phi \nabla^2 \psi] dV \\ \oint_{S(\mathbf{R})} \phi \nabla \psi \cdot d\mathbf{S} &= \oint_{S(\mathbf{R})} \phi(\mathbf{r}) \frac{\partial \psi}{\partial n} dS \\ &= \int_{\mathbf{R}} [\nabla \phi \cdot \nabla \psi + \phi \nabla^2 \psi] dV \end{aligned}$$

The Green first identity (4.61) immediately follows. This completes the proof. ■

A first interesting consequence of the Green first identity follows when  $\psi$  is harmonic, so that  $\nabla^2\psi = 0$  in  $\mathbf{R}$ . In this instance, setting  $\phi = 1$  in (4.61) gives:

$$\oint_{S(\mathbf{R})} \frac{\partial \psi}{\partial n} dS = 0 \quad (4.62)$$

This equation states that the normal derivative of a harmonic function averages to zero on the frontier of the region where it is harmonic. It also can be shown that if (4.62) holds, then  $\psi$  is harmonic in  $\mathbf{R}$ . Therefore, (4.62) is a necessary and sufficient condition for  $\psi$  to be harmonic in  $\mathbf{R}$ . Now let us assume that  $\mathbf{F}$  is the potential field associated with  $\psi$ , so that:  $\mathbf{F} = \nabla\psi$ . Then, if  $\mathbf{n}$  is a versor normal to  $S(\mathbf{R})$  we have:

$$\frac{\partial \psi}{\partial n} = \nabla \psi \cdot \mathbf{n} = \mathbf{F} \cdot \mathbf{n}$$

Therefore, Eq. (4.62) can be rewritten as follows:

$$\oint_{S(\mathbf{R})} \mathbf{F} \cdot d\mathbf{S} = 0 \quad (4.63)$$

This equation is often indicated as *Gauss' law*. It says that the normal component of a potential field averages to zero over the boundary of the

domain where its harmonic potential is defined. Alternatively, we can say that the flux of a conservative vector field  $\mathbf{F}$  through the closed boundary  $S(\mathbf{R})$ , where its harmonic potential is defined, is zero.

Applying the divergence theorem to (4.63) yields the following alternative form of Gauss' law:

$$\int_{\mathbf{R}} \nabla \cdot \mathbf{F} dV = 0 \quad (4.64)$$

Now let us assume that  $\psi$  is harmonic in  $\mathbf{R}$  and that  $\psi = 0$  on  $S(\mathbf{R})$ . Setting  $\phi = \psi$  in (4.61) yields:

$$\int_{\mathbf{R}} (\nabla \psi)^2 dV = 0$$

Therefore,  $\nabla\psi = 0$  in  $\mathbf{R}$ , thereby,  $\psi = \text{const}$ . The hypothesis that  $\psi = 0$  on  $S(\mathbf{R})$  and the continuity of  $\psi$  then imply that  $\psi = 0$  in  $\mathbf{R}$ . This proves the important property that if  $\psi$  is harmonic in  $\mathbf{R}$  and  $\psi = 0$  on  $S(\mathbf{R})$ , then  $\psi$  vanishes in the whole region  $\mathbf{R}$ . The following uniqueness theorem is another corollary of Green's first identity.

#### Corollary 4 (Stokes' Theorem)

*A harmonic function is uniquely determined by its boundary values.*

*Proof* Let  $\phi = \phi(\mathbf{r})$  and  $\psi = \psi(\mathbf{r})$  be two harmonic functions in a closed region  $\mathbf{R}$ . Let us also assume that  $\phi = \psi$  on the boundary  $S(\mathbf{R})$ . Clearly,  $\phi - \psi$  is harmonic in  $\mathbf{R}$  and it results  $\phi - \psi = 0$  on  $S(\mathbf{R})$ . Therefore, according to the previous corollary we must have  $\phi - \psi = 0$  in  $\mathbf{R}$ , that is,  $\phi = \psi$ . This proves that if two functions coincide on the boundary of  $\mathbf{R}$  then they also coincide in  $\mathbf{R}$ . Therefore, a harmonic function is uniquely determined by the boundary values. ■

Stokes' theorem implies that the *Dirichlet boundary-value problem*, that is, find a function that solves Laplace's equation in the interior of  $\mathbf{R}$  given the values on the boundary of this

region, admits a unique solution. The following corollary of Green's first identity is related to another boundary-value problem, the *Neumann boundary-value problem*, which is to find a function that solves Laplace's equation in the interior of  $\mathbf{R}$  given the values of the normal derivatives on the boundary of this region.

### Corollary 5

Any function that is harmonic in a region  $\mathbf{R}$  is determined, up to an additive constant, by the values of its normal derivative on the boundary  $S(\mathbf{R})$ .

*Proof* If  $f = f(\mathbf{r})$  is harmonic in a closed region  $\mathbf{R}$ , then the Green first identity yields:

$$\oint_{S(\mathbf{R})} f(\mathbf{r}) \frac{\partial f}{\partial n} dS = \int_{\mathbf{R}} (\nabla f)^2 dV$$

Let  $\phi = \phi(\mathbf{r})$  and  $\psi = \psi(\mathbf{r})$  be two harmonic functions in  $\mathbf{R}$ , such that  $f(\mathbf{r}) = \phi(\mathbf{r}) - \psi(\mathbf{r})$ , and let us assume that  $\partial\phi/\partial n = \partial\psi/\partial n$ . Then:

$$\begin{aligned} \int_{\mathbf{R}} [\nabla(\phi - \psi)]^2 dV &= \oint_{S(\mathbf{R})} (\phi - \psi) \\ &\times \left[ \frac{\partial\phi}{\partial n} - \frac{\partial\psi}{\partial n} \right] dS = 0 \end{aligned}$$

Therefore,  $\nabla(\phi - \psi) = 0$  in  $\mathbf{R}$  and  $\phi - \psi = \text{const}$ . This proves the theorem. ■

The last corollary is a uniqueness theorem for the *mixed boundary-value problem*.

### Corollary 6

Let  $\psi$  be a harmonic function in a region  $\mathbf{R}$ , and let  $\alpha, \beta$ , and  $g$  be continuous functions on  $S(\mathbf{R})$ , with  $\alpha/\beta > 0$ , such that:

$$\alpha\psi + \beta \frac{\partial\psi}{\partial n} = g \quad (4.65)$$

on  $S(\mathbf{R})$ . Then  $\psi$  is uniquely determined in  $\mathbf{R}$ .

*Proof* Let  $\phi = \phi(\mathbf{r})$  and  $\psi = \psi(\mathbf{r})$  be two harmonic functions in  $\mathbf{R}$ , such that (4.65) is satisfied on  $S(\mathbf{R})$ . Then:

$$\alpha(\phi - \psi) + \beta \left( \frac{\partial\phi}{\partial n} - \frac{\partial\psi}{\partial n} \right) = 0$$

$$\left( \frac{\partial\phi}{\partial n} - \frac{\partial\psi}{\partial n} \right) = -\frac{\alpha}{\beta}(\phi - \psi)$$

Thus, by the Green first identity we have:

$$\begin{aligned} \oint_{S(\mathbf{R})} (\phi - \psi) \left[ \frac{\partial\phi}{\partial n} - \frac{\partial\psi}{\partial n} \right] dS &= \\ - \oint_{S(\mathbf{R})} \frac{\alpha}{\beta} (\phi - \psi)^2 dS &= \int_{\mathbf{R}} [\nabla(\phi - \psi)]^2 dV \end{aligned}$$

This identity can be satisfied only if all terms are zero, because  $\alpha/\beta > 0$ . Therefore,

$$\oint_{S(\mathbf{R})} \frac{\alpha}{\beta} (\phi - \psi)^2 dS = \int_{\mathbf{R}} [\nabla(\phi - \psi)]^2 dV = 0$$

Consequently,  $\phi = \psi + \text{const}$  in  $\mathbf{R}$  and  $\phi = \psi$  on  $S(\mathbf{R})$ . By continuity, we must have  $\phi = \psi$  also in  $\mathbf{R}$  and the uniqueness is proved. ■

Another important set of properties for the potential arises from the second Green's identity, which can be obtained easily from (4.61).

### Green's Second Identity

Let  $\phi = \phi(\mathbf{r})$  and  $\psi = \psi(\mathbf{r})$  be scalar fields, defined in a closed region  $\mathbf{R}$  with boundary  $S(\mathbf{R})$ . Then:

$$\begin{aligned} \oint_{S(\mathbf{R})} \left[ \phi(\mathbf{r}) \frac{\partial\psi}{\partial n} - \psi(\mathbf{r}) \frac{\partial\phi}{\partial n} \right] dS \\ = \int_{\mathbf{R}} [\phi(\mathbf{r}) \nabla^2 \psi - \psi(\mathbf{r}) \nabla^2 \phi] dV \quad (4.66) \end{aligned}$$

*Proof* If we interchange  $\phi$  and  $\psi$  in (4.61) and subtract the result from this identity, the identity (4.66) immediately follows. ■

As a corollary, when both  $\phi$  and  $\psi$  are harmonic in  $\mathbf{R}$ , the second Green's identity becomes:

$$\oint_{S(\mathbf{R})} \left[ \phi(\mathbf{r}) \frac{\partial\psi}{\partial n} - \psi(\mathbf{r}) \frac{\partial\phi}{\partial n} \right] dS = 0 \quad (4.67)$$

The following theorem, known as the *Green's third identity*, is probably the most important result for potential functions, in particular for harmonic fields.

### Green's Third Identity

Let  $\psi = \psi(\mathbf{r})$  be a scalar field, defined in a closed region  $\mathbf{R}$  with boundary  $S(\mathbf{R})$ , and let  $P$  be a fixed point in  $\mathbf{R}$ . If  $r$  is the distance of any other point  $Q \in \mathbf{R}$  from  $P$ , then:

$$\begin{aligned} \psi(P) = & -\frac{1}{4\pi} \int_{\mathbf{R}} \frac{1}{r} \nabla^2 \psi dV + \frac{1}{4\pi} \oint_{S(\mathbf{R})} \frac{1}{r} \frac{\partial \psi}{\partial n} dS \\ & - \frac{1}{4\pi} \oint_{S(\mathbf{R})} \psi \frac{\partial}{\partial n} \left( \frac{1}{r} \right) dS \end{aligned} \quad (4.68)$$

*Proof* Let us consider the scalar field  $\phi = 1/r$ . In this instance, the Green's second identity assumes the form:

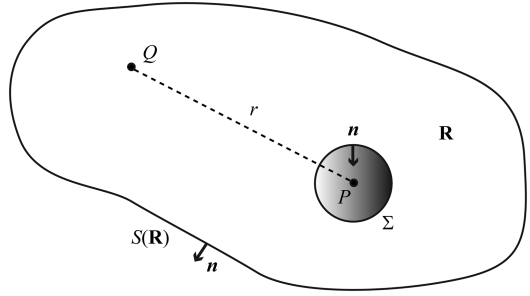
$$\begin{aligned} \oint_{S(\mathbf{R})} \left[ \frac{1}{r} \frac{\partial \psi}{\partial n} - \psi \frac{\partial}{\partial n} \left( \frac{1}{r} \right) \right] dS \\ = \int_{\mathbf{R}} \left[ \frac{1}{r} \nabla^2 \psi - \psi \nabla^2 \left( \frac{1}{r} \right) \right] dV \end{aligned}$$

We know that  $\nabla^2(1/r) = 0$  for any point  $Q \neq P$ . Therefore, let us consider a small sphere,  $\Sigma$ , about  $P$  (Fig. 4.20) and the set  $\mathbf{E} = \mathbf{R} - \Sigma$ .

In  $\mathbf{E}$ , the previous identity can be rewritten as follows:

$$\begin{aligned} \int_{\mathbf{E}} \frac{1}{r} \nabla^2 \psi dV &= \oint_{S(\mathbf{E})} \left[ \frac{1}{r} \frac{\partial \psi}{\partial n} - \psi \frac{\partial}{\partial n} \left( \frac{1}{r} \right) \right] dS \\ &= \oint_{S(\mathbf{R})} \left[ \frac{1}{r} \frac{\partial \psi}{\partial n} - \psi \frac{\partial}{\partial n} \left( \frac{1}{r} \right) \right] dS \\ &\quad + \oint_{\Sigma} \left[ \frac{1}{r} \frac{\partial \psi}{\partial n} - \psi \frac{\partial}{\partial n} \left( \frac{1}{r} \right) \right] dS \end{aligned} \quad (4.69)$$

In evaluating the integral over the surface of  $\Sigma$ , we must take into account that any surface



**Fig. 4.20** Geometry of problem for Green's third identity

element vector  $\mathbf{n}$  is directed *inwards*, because the sphere itself belongs to the complement of  $\mathbf{E}$ . Therefore, for any scalar field  $f$  defined on the surface of  $\Sigma$ , the directional derivative in the direction  $\mathbf{n}$  is the opposite of a directional derivative in the direction of  $Q$ , so that it results:

$$\frac{\partial f}{\partial n} = -\frac{\partial f}{\partial r}$$

Furthermore, when we integrate over the surface of  $\Sigma$ , the points  $Q$  have a fixed distance  $r = R$  from  $P$ ,  $R$  being the radius of  $\Sigma$ , and the spatial average of  $\psi$  over the surface of this sphere is:

$$\langle \psi \rangle_{\Sigma} = \frac{1}{4\pi R^2} \oint_{\Sigma} \psi dS$$

Clearly, if we take the limit as  $R \rightarrow 0$  of this expression, we obtain simply  $\psi(P)$ . At the same time, the integral at left-hand side of (4.69) will be extended to the whole region  $\mathbf{R}$ . The last integral at the right-hand side of (4.69) can be evaluated as follows:

$$\begin{aligned} \oint_{\Sigma} \left[ \frac{1}{r} \frac{\partial \psi}{\partial n} - \psi \frac{\partial}{\partial n} \left( \frac{1}{r} \right) \right] dS \\ = \oint_{\Sigma} \left[ -\frac{1}{r} \frac{\partial \psi}{\partial r} + \psi \frac{\partial}{\partial r} \left( \frac{1}{r} \right) \right] dS = \\ = \oint_{\Sigma} \left[ -\frac{1}{r} \frac{\partial \psi}{\partial r} - \frac{1}{r^2} \psi \right] dS = \\ = -4\pi \langle \psi \rangle_{\Sigma} - \frac{1}{R} \oint_{\Sigma} \frac{\partial \psi}{\partial r} dS \end{aligned}$$

Therefore, taking the limit as  $R \rightarrow 0$  of this expression, we obtain:

$$\int_{\mathbf{R}} \frac{1}{r} \nabla^2 \psi dV = \oint_{S(\mathbf{R})} \left[ \frac{1}{r} \frac{\partial \psi}{\partial n} - \psi \frac{\partial}{\partial n} \left( \frac{1}{r} \right) \right] dS - 4\pi \psi(P)$$

The Green's third identity immediately follows from this identity. ■

The Green's third identity implies that the potential at any point  $P$  has three components:

1. The potential associated with the volume  $\mathbf{R}$ , with density  $-(1/4\pi r)\nabla^2 \psi$ ;
2. A potential associated with the surface  $S(\mathbf{R})$ , with density  $(1/4\pi r)\partial\psi/\partial n$ ;
3. Another potential associated with the surface  $S(\mathbf{R})$ , with density  $-(1/4\pi)\psi\partial(1/r)/\partial n$ .

If  $\psi$  is harmonic, then the volume component in (4.68) is zero, and the identity simplifies to:

$$\psi(P) = \frac{1}{4\pi} \oint_{S(\mathbf{R})} \left[ \frac{1}{r} \frac{\partial \psi}{\partial n} - \psi \frac{\partial}{\partial n} \left( \frac{1}{r} \right) \right] dS \quad (4.70)$$

This important and surprising result is called *representation formula* (e.g., Blakely 1996). It says that the value of a harmonic function can be calculated from the values it takes over the boundary  $S(\mathbf{R})$  of the harmonicity region  $\mathbf{R}$  and from normal derivatives along the same boundary. However, if we consider any subset  $\mathbf{E} \subset \mathbf{R}$  such that  $P \in \mathbf{E}$ , this is clearly a region of harmonicity for  $\psi$ . Therefore, we can use the boundary values over  $S(\mathbf{E})$  to determine  $\psi(P)$  as well. Therefore, if the Dirichlet boundary value problem ensures the unicity of a harmonic function given the boundary values, the converse is not true, in the sense that we cannot determine uniquely the boundary values starting from a known value of  $\psi$  in  $\mathbf{R}$ . We shall see that Eq. (4.70) is an invaluable tool for the manipulation of potential field data. An important corollary of Green's third identity is the following *Gauss' theorem of the arithmetic mean*.

### Gauss' Theorem of the Arithmetic Mean

*The value of a harmonic function at a point is the average of the function over any spherical neighbor of harmonicity about the point.*

*Proof* Let  $\Sigma$  be a sphere centered on the point  $P \in \mathbf{R}$ . If  $R$  is the radius of  $\Sigma$ , then the representation formula (4.70) assumes the form:

$$\psi(P) = \frac{1}{4\pi R} \oint_{\Sigma} \frac{\partial \psi}{\partial n} dS + \frac{1}{4\pi R^2} \oint_{\Sigma} \psi dS$$

According to the first Green's identity, the first integral vanishes. Therefore:

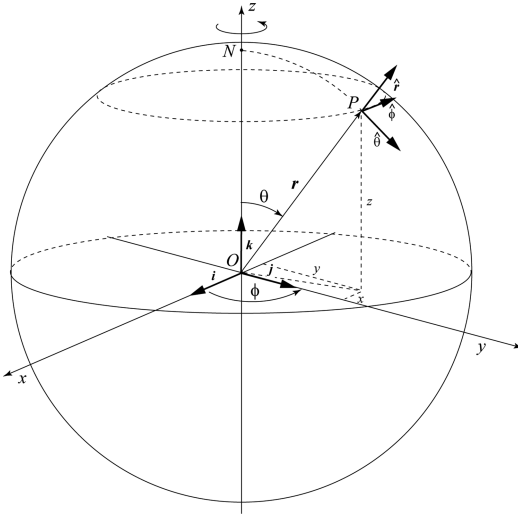
$$\psi(P) = \frac{1}{4\pi R^2} \oint_{\Sigma} \psi dS = \langle \psi \rangle_{\Sigma} \quad (4.71)$$

This identity proves the theorem. ■

Gauss' theorem reveals two key interconnected features of a harmonic function, namely, the fact that its value at a point coincides with the average over a spherical neighbor, and the lack of maxima and minima within the region of harmonicity. These properties were used in Sect. 3.3 to find a numerical solution to Laplace's equation. In the next section, we shall face the problem to find an analytic solution to this equation.

## 4.8 Spherical Harmonic Expansion of the Geomagnetic Field and the IGRF

In Sect. 3.3, we mentioned the fact that the scalar magnetic potential  $V$  satisfies Laplace's Eq. (3.29) in any region  $\mathbf{R}$  where the current density is zero (including magnetization currents). Now we want to find a solution to this equation for the geomagnetic field. To this purpose, it is convenient to use spherical coordinates  $(r, \theta, \phi)$  (see Sect. 2.3) instead of the usual Cartesian coordinates  $(x, y, z)$ , the laws of transformation between the two systems being given by Eqs. (2.27) and (2.28).



**Fig. 4.21** Base versors for the transformation from Cartesian to spherical derivatives

To transform Laplace's equation to spherical coordinates, it is useful to introduce three orthogonal versors,  $\hat{r}$ ,  $\hat{\theta}$ , and  $\hat{\phi}$ , which are directed respectively toward increasing distance from the origin, increasing colatitude (that is, southward), and increasing longitude (i.e., eastward) at point  $P \equiv (x, y, z)$  (Fig. 4.21). Our objective is to convert the Cartesian derivatives of Laplace's equation to spherical derivatives, which measure variations of potential with respect to small increments of  $r$ ,  $\theta$ , and  $\phi$ .

It is easy to show that the Cartesian components of the three versors are given by the following transformations:

$$\begin{cases} \hat{r} = \sin \theta \cos \phi \hat{i} + \sin \theta \sin \phi \hat{j} + \cos \theta \hat{k} \\ \hat{\theta} = \cos \theta \cos \phi \hat{i} + \cos \theta \sin \phi \hat{j} - \sin \theta \hat{k} \\ \hat{\phi} = -\sin \phi \hat{i} + \cos \phi \hat{j} \end{cases} \quad (4.72)$$

To transform the Laplacian operator  $\nabla^2$ , it is necessary to find first an expression for the gradient of the potential  $V$  in spherical coordinates. The spatial derivatives  $\partial V/\partial x$ ,  $\partial V/\partial y$ , and  $\partial V/\partial z$  can be transformed using simple chain rules and transformations (2.27) and (2.28). The resulting expression for the gradient is:

$$\nabla V = \frac{\partial V}{\partial r} \hat{r} + \frac{1}{r} \frac{\partial V}{\partial \theta} \hat{\theta} + \frac{1}{r \sin \theta} \frac{\partial V}{\partial \phi} \hat{\phi} \quad (4.73)$$

A formula for the Laplacian can be found taking into account that  $\nabla^2 V = \nabla \cdot \nabla V$ . We obtain the following expression:

$$\begin{aligned} \nabla^2 V &= \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial V}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \\ &\times \left( \sin \theta \frac{\partial V}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 V}{\partial \phi^2} \end{aligned} \quad (4.74)$$

Therefore, the spherical form of Laplace's equation can be written as follows:

$$\begin{aligned} \frac{\partial}{\partial r} \left( r^2 \frac{\partial V}{\partial r} \right) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial V}{\partial \theta} \right) \\ + \frac{1}{\sin^2 \theta} \frac{\partial^2 V}{\partial \phi^2} = 0 \end{aligned} \quad (4.75)$$

We can solve this equation by separation of variables. First, let us try the following separation of the potential into a radial component,  $R$ , and a component depending only from colatitude and longitude:

$$V(r, \theta, \phi) = R(r)Y(\theta, \phi) \quad (4.76)$$

Substituting into (4.75) gives:

$$\begin{aligned} Y \frac{\partial}{\partial r} \left( r^2 \frac{\partial R}{\partial r} \right) + \frac{R}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial Y}{\partial \theta} \right) \\ + \frac{R}{\sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} = 0 \end{aligned}$$

Hence, dividing both sides by  $YR$ :

$$\begin{aligned} \frac{1}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) &= -\frac{1}{Y \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial Y}{\partial \theta} \right) \\ &- \frac{1}{Y \sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} \end{aligned} \quad (4.77)$$

where we have substituted partial derivatives at the left-hand side by ordinary ones, because  $R$

depends only from the distance  $r$ . This equation has the form:  $f(r) = g(\theta, \phi)$ . If we fix latitude and longitude, the right-hand side of (4.77) does not change. Therefore, function  $f$  must be a constant:  $f(r) = k$ . For reasons that will be clear later, it is convenient to set  $k = n(n+1)$ . In this instance, taking the left-hand side of (4.77) gives:

$$\frac{1}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) = n(n+1)$$

This is an ordinary differential equation, which can be rewritten as follows:

$$r^2 \frac{d^2 R}{dr^2} + 2r \frac{dR}{dr} - n(n+1) R(r) = 0 \quad (4.78)$$

The solutions of (4.78) have the form:

$$R(r) = \begin{cases} r^n \\ r^{-(n+1)} \end{cases} \quad (4.79)$$

Therefore, we have two classes of solutions of Laplace's equations, one characterized by increasing values of the radial component for increasing distances from the Earth's center, and one that decreases with the distance from the Earth:

$$V(r, \theta, \phi) = \begin{cases} V_e(r, \theta, \phi) = r^n Y(\theta, \phi) \\ V_i(r, \theta, \phi) = r^{-(n+1)} Y(\theta, \phi) \end{cases} \quad (4.80)$$

The first class of solutions clearly corresponds to external sources for the geomagnetic potential (magnetosphere and ionosphere), whereas the second class arises from the internal sources (crustal and core). Now let us focus on the right-hand side of (4.77). We perform further separation of the variables by assuming that the component depending only from colatitude and longitude has the form:  $Y(\theta, \phi) = T(\theta)L(\phi)$ . Equating right-hand side of (4.77) to  $n(n+1)$  gives:

$$\frac{1}{Y \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{Y \sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} = -n(n+1)$$

Then,

$$\begin{aligned} \frac{1}{T \sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{dT}{d\theta} \right) + \frac{1}{L \sin^2 \theta} \frac{d^2 L}{d\phi^2} \\ = -n(n+1) \end{aligned}$$

where we have substituted total derivatives for single variable functions  $T$  and  $L$  with respect to  $\theta$  and  $\phi$ . Finally,

$$\begin{aligned} \frac{\sin \theta}{T} \frac{d}{d\theta} \left( \sin \theta \frac{dT}{d\theta} \right) + n(n+1) \sin^2 \theta \\ = -\frac{1}{L} \frac{d^2 L}{d\phi^2} \end{aligned} \quad (4.81)$$

This equation has the form  $f(\theta) = g(\phi)$ . If we fix a longitude  $\phi$ , then the right-hand side of (4.81) does not change. Therefore, function  $f$  must be a constant:  $f(\theta) = k$ . Setting  $k = m^2$  and considering the right-hand side of (4.81) gives:

$$\frac{d^2 L}{d\phi^2} = -m^2 L(\phi) \quad (4.82)$$

This is another ordinary differential equation, whose solutions have the form:

$$L(\phi) = \begin{cases} \sin m\phi \\ \cos m\phi \end{cases} \quad (4.83)$$

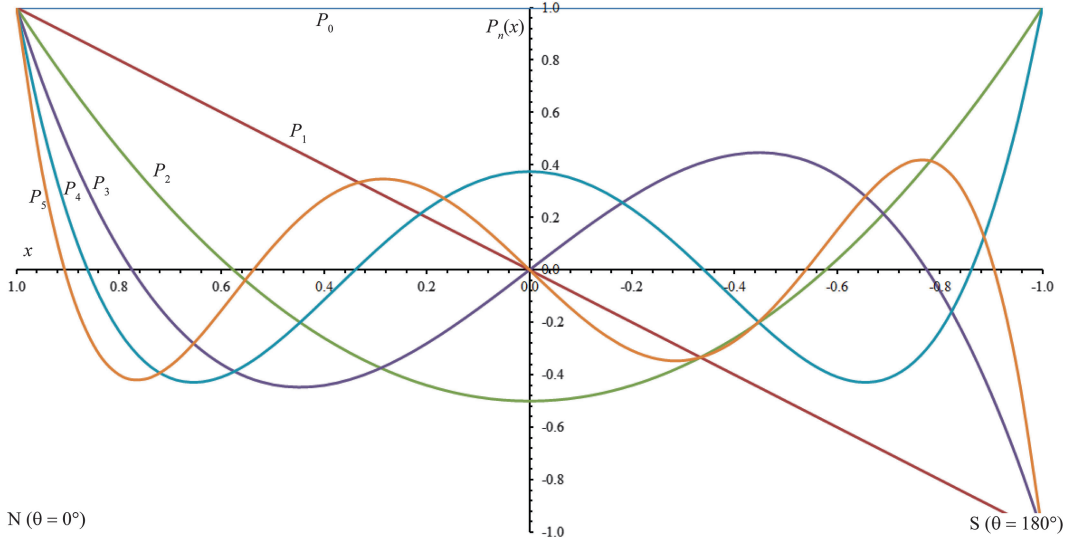
Finally, let us consider the left-hand side of (4.81). This equation is more complicate:

$$\frac{\sin \theta}{T} \frac{d}{d\theta} \left( \sin \theta \frac{dT}{d\theta} \right) + n(n+1) \sin^2 \theta = m^2$$

If we perform the substitution:  $x = \cos \theta$  in this equation, in a few steps we obtain:

$$\begin{aligned} (1-x^2) \frac{d^2 T}{dx^2} - 2x \frac{dT}{dx} \\ + \left[ n(n+1) - \frac{m^2}{1-x^2} \right] T(x) = 0 \end{aligned} \quad (4.84)$$

This is the well-known *Legendre equation*, whose solutions are the *Legendre associate polynomials*. In particular, it can be shown that the



**Fig. 4.22** Legendre polynomials for  $n \leq 5$ . Note that the  $x$  axis scale has been inverted to conform to increasing colatitudes (from North to South)

parameters  $n$  and  $m$ , which are called respectively *degree* and *order* of the solution, must be non negative integers such that  $m \leq n$ . The standard associate Legendre polynomials are indicated as  $P_{nm}(x)$ . These functions have the form:

$$T(x) = P_{nm}(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_n(x) \quad (4.85)$$

where the polynomials  $P_n(x)$ , which are called simply *Legendre polynomials*, are defined by the following Rodrigues' formula:

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n \quad (4.86)$$

These polynomials are solutions to the Legendre equation for  $m = 0$ :

$$(1-x^2) \frac{d^2 T}{dx^2} - 2x \frac{dT}{dx} + n(n+1) T(x) = 0 \quad (4.87)$$

The shape of the first six Legendre polynomials is plotted in Fig. 4.22. We note that Legendre functions of even degree are symmetric about the Equator ( $P_n(x) = P_n(-x)$ ), whereas they are antisymmetric for  $n$  odd (i.e.,  $P_n(x) =$

$-P_n(-x)$ ). Both the associate and standard Legendre polynomials have specific normalization properties and satisfy orthogonality conditions. The normalization of (4.85) and (4.86) is called *Ferrers normalization* (Winch et al. 2005). These functions satisfy the following orthogonality conditions:

$$\int_{-1}^{+1} P_{nm}(x) P_{sm}(x) dx = \frac{2}{2n+1} \frac{(n+m)!}{(n-m)!} \delta_{ns} \quad (4.88)$$

$$\int_{-1}^{+1} P_n(x) P_s(x) dx = \frac{2}{2n+1} \delta_{ns} \quad (4.89)$$

where  $\delta_{ns}$  is the Kronecker delta ( $\delta_{ns} = 1$  for  $n = s$  and zero otherwise). Setting  $n = s$  in (4.88) gives:

$$\int_{-1}^{+1} |P_{nm}(x)|^2 dx = \frac{2}{2n+1} \frac{(n+m)!}{(n-m)!}$$

Therefore, the root-mean-square magnitudes of the Legendre polynomials are subject to

considerable variability when  $m$  varies from 0 to  $n$ , ranging between  $[2/(2n+1)]^{1/2}$  and  $[2(2n)!/(2n+1)]^{1/2}$ . Consequently, if we build a general solution to Laplace's equation by superposition of Legendre polynomials, the coefficients of the series have a wide range of values, depending on  $n$  and  $m$ , to compensate the variability of the functions  $P_{nm}$ . However, an expansion in series of Legendre polynomials would be more informative if the magnitude of the expansion coefficients reflected the relative importance of the corresponding polynomial terms. Therefore, in geomagnetism the Ferrers normalization is substituted by a different criterion, which is called *Schmidt quasi-normalization*. The new polynomials have the form:

$$P_n^m(x) \equiv \sqrt{2 \frac{(n-m)!}{(n+m)!}} P_{nm}(x) \quad (4.90)$$

To obtain the general solution to Laplace's equation we must first combine particular solutions to Eqs. (4.78) and (4.82) with the normalized Legendre polynomials. Then, the general solution will be a linear combination of particular solutions for any value of  $n$  and  $m$ . A particular solution for the external field is:

$$V_e(r, \theta, \phi) = r^n \left\{ \frac{\sin m\phi}{\cos m\phi} \right\} P_n^m(\cos \theta) \quad (4.91)$$

Similarly, for the internal field it results:

$$V_i(r, \theta, \phi) = r^{-(n+1)} \left\{ \frac{\sin m\phi}{\cos m\phi} \right\} P_n^m(\cos \theta) \quad (4.92)$$

From here on, we shall focus on the potential associated with internal sources, which gives the main contribution to the observed field. This potential will be called simply  $V$ . When taking a linear combination of particular solutions (4.92), we must take into account that the solution is appropriate only where the potential is harmonic. Therefore, usually the potential is represented outside a reference sphere of minimum radius,  $a$ , that encloses all the sources. In geomagnetism,

the reference sphere is conventionally chosen to have a radius equal to the Earth's mean radius,  $R_e \cong 6,371$  km. Furthermore, the radial distances are expressed in units of  $a$ , and the whole series is multiplied by the radius  $a$ , in order to have the coefficients expressed in tesla:

$$V(r, \theta, \phi) = a \sum_{n=1}^{\infty} \left( \frac{a}{r} \right)^{n+1} \sum_{m=0}^n [g_n^m \cos m\phi + h_n^m \sin m\phi] P_n^m(\cos \theta); \quad r \geq a \quad (4.93)$$

The terms within the square brackets, which depend only from longitude  $\phi$ , resemble the usual Fourier's harmonic series. Multiplied by the Legendre polynomials, they are known as *surface spherical harmonics*:

$$Y_n^m(\theta, \phi) = \left\{ \frac{\cos m\phi}{\sin m\phi} \right\} P_n^m(\cos \theta) \quad (4.94)$$

These functions are the spherical equivalent of sines and cosines of the more familiar Fourier's harmonic series. It is easy to prove that their rms magnitude over a sphere of radius  $r$  is independent from the order  $m$ .

Let  $dS$  be a surface element on the sphere of radius  $r$ :

$$dS = r^2 \sin \theta d\theta d\phi \quad (4.95)$$

Then, the surface spherical harmonics satisfy the following orthogonality condition:

$$\begin{aligned} & \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi Y_n^m(\theta, \phi) Y_s^r(\theta, \phi) \sin \theta d\theta \\ &= \frac{1}{2n+1} \delta_{ns} \delta_{mr} \end{aligned} \quad (4.96)$$

Therefore, the magnitude of the coefficients  $g$  and  $h$  in the harmonic expansion (4.93) measures the strength of the corresponding terms in the series according to the degree. These coefficients are called *Gauss coefficients*, in recognition of the great contribution of this scientist to the development of the spherical harmonic representation.

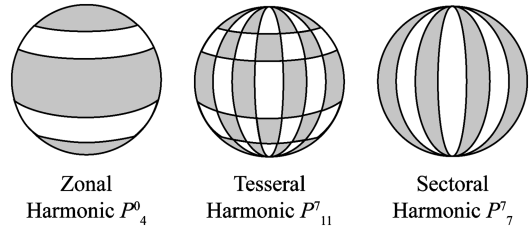
In the case of a time-averaged geomagnetic field (see Sect. 4.3), the series (4.93) is assumed to be symmetric with respect to the spin axis, so that the potential does not depend anymore from longitude. This is made by setting  $m = 0$  in (4.93), so that at the Earth's surface we have simply:

$$V(\theta) = a \sum_{n=1}^{\infty} g_n^0 P_n(\cos \theta) \quad (4.97)$$

A Legendre polynomial  $P_n$  has  $n$  zeroes between  $\theta = 0$  and  $\theta = \pi$ . These special surface harmonics, which depend only from colatitude, are called *zonal harmonics*. In the general case, a surface harmonic  $Y_n^m(\theta, \phi)$  vanishes along  $n-m$  parallels of latitude, corresponding to the zeroes of the  $m$ -th order derivative of  $P_n(\theta)$ , and it also has  $m$  zeroes at the poles, where the terms  $(1 - x^2)^{m/2} = \sin^m \theta$  vanish. Regarding the dependency from longitude, both  $\cos m\phi$  and  $\sin m\phi$  have  $2m$  zeroes between  $0$  and  $2\pi$ . Finally, for  $n = m$  the function  $Y_n^m(\theta, \phi)$  has only  $n$  zeroes at the poles and  $2n$  zeroes in longitude. The parallels of latitude and the meridians along which a surface harmonic  $Y_n^m(\theta, \phi) = 0$  divide the spherical surface of radius  $r$  into a series of *tesserae* where the values of the function have alternate signs. Therefore, when  $m > 0$  and  $n > m$  these functions are also called *tesseral harmonics*. Finally, for  $n = m$  the function  $Y_n^m(\theta, \phi)$  divides the spherical surface of radius  $r$  into a series of sectors bounded by meridians, thereby, we say that  $Y_n^m(\theta, \phi)$  is a *sectoral harmonic*. Figure 4.23 shows some examples of zonal, sectoral, and tesseral harmonics.

A key concept in spherical harmonic analysis is that of *harmonic wavelength*. In standard Fourier's analysis, the relative contribution of the sines and cosines to the series is determined by their amplitude and wavelength, and a given set of terms in the series can be related to a specific physical phenomenon. A similar feature characterizes spherical harmonics.

In particular, if we consider the surface harmonics as waves on the surface of a sphere of radius  $a$ , it is possible to define a wavelength  $\lambda$  as the distance between two successive peaks or zeroes of  $Y_n^m(\theta, \phi)$  and it is quite intuitive that higher degree harmonics have shorter wavelength



**Fig. 4.23** Positive (gray) and negative (white) patches for some surface harmonics

than the lower ones. Therefore, we can remove the high-degree terms of a spherical harmonic expansion to emphasize long-wavelength features of the field associated with core processes or, alternatively, we could cut off low-degree terms to enhance short-wavelength features of the crustal field. To find a precise relationship between harmonic degree and wavelength, let us consider a small patch of a sphere with radius  $a$  and a local Cartesian coordinate system  $(\xi, \psi, \zeta)$  oriented as in Fig. 4.21. On the sphere, the surface harmonic  $Y_n^m(\theta, \phi)$  satisfies the equation:

$$\begin{aligned} \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial Y_n^m}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 Y_n^m}{\partial \phi^2} \\ = -n(n+1) Y_n^m(\theta, \phi) \end{aligned} \quad (4.98)$$

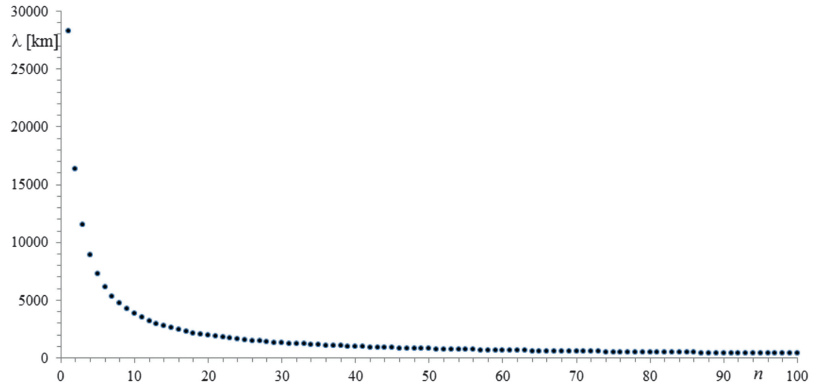
We want to transform this equation to local Cartesian coordinates  $(\xi, \psi, \zeta)$ . To this purpose, we note that for any function  $f = f(\theta, \phi)$ :

$$\begin{aligned} \frac{\partial f}{\partial \theta} &= \frac{\partial f}{\partial \xi} \frac{\partial \xi}{\partial \theta} = a \frac{\partial f}{\partial \xi}; \quad \frac{\partial^2 f}{\partial \theta^2} = a^2 \frac{\partial^2 f}{\partial \xi^2} \\ \frac{\partial f}{\partial \phi} &= \frac{\partial f}{\partial \psi} \frac{\partial \psi}{\partial \phi} = a \sin \theta \frac{\partial f}{\partial \psi}; \quad \frac{\partial^2 f}{\partial \phi^2} = a^2 \sin^2 \theta \frac{\partial^2 f}{\partial \psi^2} \end{aligned}$$

Furthermore, we observe that the variations of  $Y_n^m(\theta, \phi)$  with colatitude occur on a much shorter scale than  $\sin \theta$ . Therefore, the term  $\sin \theta$  at the left-hand side of (4.98) can be considered as approximately constant. Consequently, Eq. (4.98) can be simplified as follows:

$$\frac{\partial^2 Y_n^m}{\partial \theta^2} + \frac{1}{\sin^2 \theta} \frac{\partial^2 Y_n^m}{\partial \phi^2} \cong -n(n+1) Y_n^m(\theta, \phi) \quad (4.99)$$

**Fig. 4.24** Wavelength of spherical harmonics as a function of degree



Transforming this equation to local coordinates gives:

$$\frac{\partial^2 Y_n^m}{\partial \xi^2} + \frac{\partial^2 Y_n^m}{\partial \psi^2} = \nabla^2 Y_n^m \cong -\frac{n(n+1)}{a^2} Y_n^m(\theta, \phi) \quad (4.100)$$

Therefore, the surface harmonics  $Y_n^m(\theta, \phi)$  satisfy the scalar Helmholtz's equation and are eigenvectors of the two-dimensional Laplacian  $\nabla^2 = \partial^2/\partial \xi^2 + \partial^2/\partial \psi^2$  with eigenvalues  $-n(n+1)/a^2$ . The structure of this equation clearly suggests a solution that is a combination of sines and cosines. Therefore, to interpret the harmonics  $Y_n^m(\theta, \phi)$  as waves, we try the following solution:

$$Y_n^m(\xi, \psi) = e^{i\mathbf{k} \cdot \boldsymbol{\rho}} \quad (4.101)$$

where  $\boldsymbol{\rho} = (\xi, \psi)$  and  $\mathbf{k} = (k_\xi, k_\psi)$  is the *wave vector*. This function has wavelength  $\lambda$  given by:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{\sqrt{k_\xi^2 + k_\psi^2}} \quad (4.102)$$

Substituting the ansatz (4.101) into (4.100) gives:

$$k^2 = \frac{n(n+1)}{a^2} \quad (4.103)$$

Therefore,

$$\lambda = \frac{2\pi a}{\sqrt{n(n+1)}} \cong \frac{2\pi a}{n+1/2} \quad (4.104)$$

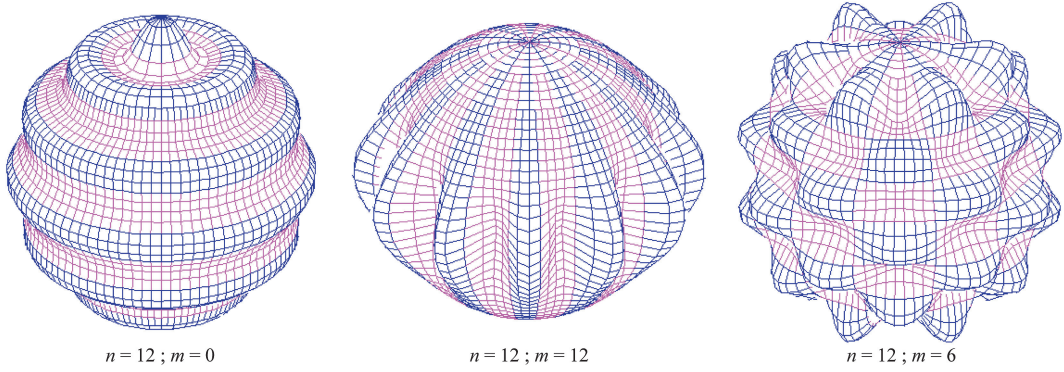
Equation 4.104 is known as *Jeans relation*. It links the wavelength of individual spherical harmonics to the corresponding degree, independently from the order. Figure 4.24 shows a plot of  $\lambda$  as a function of  $n$ . We note that in order to reach a resolution of  $\sim 1,000$  km it is necessary to have  $n = 40$ .

A 3-D view of the oscillations associated with surface harmonics is shown in Fig. 4.25. We stress again the importance of the result that *any* function that satisfies Laplace's equation can be represented by a sum of spherical harmonics (Eq. 4.93), which is the spherical analogue of the classic Fourier series. We shall see soon that this result furnishes a powerful tool for the parametric fit to observed data and the creation of global models of the geomagnetic field.

Let us consider again the magnetic pressure  $P_m = B^2/2\mu_0$  (Eq. 4.56). This quantity has dimensions  $\text{Pa} = \text{Nm}^{-2} = \text{Jm}^{-3}$ . Therefore, it represents an *energy density* for the magnetic field. In fact, any magnetic field  $\mathbf{B}$  is generated by a system of macroscopic or microscopic currents. These currents must be maintained by some emf and require a certain amount of work. It is possible to show that the total work that is necessary to set up the system of currents that generates a magnetic field  $\mathbf{B}$  in a region  $\mathbf{R}$  is given by:

$$E = \frac{1}{2\mu_0} \int_{\mathbf{R}} \mathbf{B}(\mathbf{r}) \cdot \mathbf{B}(\mathbf{r}) dV \quad (4.105)$$

This is the total energy stored by the magnetic field in the region  $\mathbf{R}$ . Now we want to calculate



**Fig. 4.25** Examples of zonal, sectoral, and tesseral surface harmonics, generated using the software utility *sph* by Phil McFadden, which can be downloaded at: [http://www.ngdc.noaa.gov/geomag/geom\\_util/sph.shtml](http://www.ngdc.noaa.gov/geomag/geom_util/sph.shtml)

the *average* magnetic energy density over a spherical surface of radius  $a$ . This is given by:

$$\begin{aligned} \frac{1}{2\mu_0} \langle B^2 \rangle_{S(a)} &= \frac{1}{8\pi a^2 \mu_0} \oint_{S(a)} \mathbf{B}(\mathbf{r}) \cdot \mathbf{B}(\mathbf{r}) dS \\ &= \frac{1}{8\pi a^2 \mu_0} \oint_{S(a)} \nabla V(\mathbf{r}) \cdot \nabla V(\mathbf{r}) dS \end{aligned} \quad (4.106)$$

To calculate the integral at the right-hand side of (4.106), we must rewrite the general solution (4.93) in terms of surface harmonics. We first set:

$$Y_n^m(\theta, \phi) = \begin{cases} P_n^m(\cos \theta) \cos m\phi; & m \geq 0 \\ P_n^{|m|}(\cos \theta) \sin |m|\phi; & m < 0 \end{cases} \quad (4.107)$$

$$v_n^m(\theta, \phi) = \begin{cases} g_n^m; & m \geq 0 \\ h_n^{|m|}; & m < 0 \end{cases} \quad (4.108)$$

Then, the potential can be rewritten as follows:

$$\begin{aligned} V(r, \theta, \phi) &= a \sum_{n=1}^{\infty} \sum_{m=-n}^{+n} \left(\frac{a}{r}\right)^{n+1} v_n^m Y_n^m(\theta, \phi); \quad r \geq a \end{aligned} \quad (4.109)$$

The functions  $Y_n^m(\theta, \phi)$  satisfy not only the conditions (4.96) but are also subject to additional conditions of orthogonality for the gradients on the unit sphere.

They are (Lowes 1966):

$$\begin{aligned} \frac{1}{4\pi} \oint_{S(1)} \nabla Y_n^m(\theta, \phi) \cdot \nabla Y_s^r(\theta, \phi) dS \\ = \frac{n(n+1)}{2n+1} \delta_{ns} \delta_{mr} \end{aligned} \quad (4.110)$$

We shall use these relations to evaluate the integral (4.106). The objective is to find an expression for the average magnetic energy density at distance  $a$  as a function of the harmonic coefficients. Let us first evaluate the derivatives of the gradient on a sphere of radius  $a$ . We can rewrite (4.73) as follows:

$$\begin{aligned} \nabla V &= \frac{\partial V}{\partial r} \hat{\mathbf{r}} + \frac{1}{r} \frac{\partial V}{\partial \theta} \hat{\boldsymbol{\theta}} + \frac{1}{r \sin \theta} \frac{\partial V}{\partial \phi} \hat{\boldsymbol{\phi}} \\ &\equiv \frac{\partial V}{\partial r} \hat{\mathbf{r}} + \nabla_s V \end{aligned} \quad (4.111)$$

Then, using (4.109) we have:

$$\begin{aligned} \frac{\partial V}{\partial r} \Big|_{r=a} &= -a(n+1) \sum_{n,m} a^{n+1} \frac{1}{r^{n+2}} \Big|_{r=a} \\ &\times v_n^m Y_n^m = -(n+1) \sum_{n,m} v_n^m Y_n^m \end{aligned} \quad (4.112)$$

$$\begin{aligned} \frac{1}{r} \frac{\partial V}{\partial \theta} \Big|_{r=a} &= \sum_{n,m} a^{n+1} \frac{1}{r^{n+1}} \Big|_{r=a} v_n^m \frac{\partial Y_n^m}{\partial \theta} \\ &= \sum_{n,m} v_n^m \frac{\partial Y_n^m}{\partial \theta} \end{aligned} \quad (4.113)$$

$$\left. \frac{1}{r \sin \theta} \frac{\partial V}{\partial \phi} \right|_{r=a} = \frac{1}{\sin \theta} \sum_{n,m} v_n^m \frac{\partial Y_n^m}{\partial \phi} \quad (4.114)$$

Therefore,

$$\begin{aligned} \nabla_s V|_{r=a} &= \sum_{n,m} v_n^m \left[ \frac{\partial Y_n^m}{\partial \theta} \hat{\theta} + \frac{1}{\sin \theta} \frac{\partial Y_n^m}{\partial \phi} \hat{\phi} \right] \\ &= \sum_{n,m} v_n^m \nabla_s Y_n^m|_{r=a} \end{aligned} \quad (4.115)$$

Substituting (4.112) and (4.115) into (4.106) gives:

$$\begin{aligned} \frac{1}{2\mu_0} \langle B^2 \rangle_{S(a)} &= \frac{1}{8\pi a^2 \mu_0} \sum_{n,m,s,r} (n+1)(s+1) v_n^m v_s^r \\ &\times \oint_{S(a)} Y_n^m Y_s^r dS + \frac{1}{8\pi a^2 \mu_0} \\ &\times \sum_{n,m,s,r} v_n^m v_s^r \oint_{S(a)} \nabla Y_n^m \cdot \nabla Y_s^r dS \end{aligned} \quad (4.116)$$

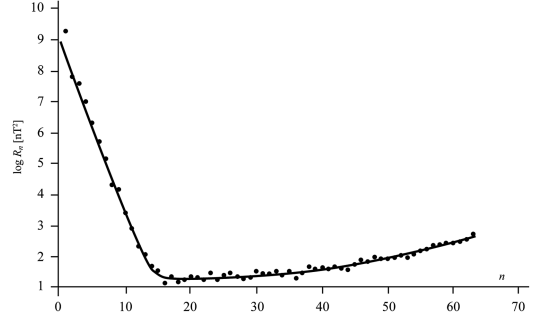
The two integrals at the right-hand side of (4.116) can be converted to integrals over the unit sphere by multiplying both terms by  $a^2$ . Then, applying the orthogonality conditions (4.96) and (4.110) we obtain:

$$\begin{aligned} \frac{1}{2\mu_0} \langle B^2 \rangle_{S(a)} &= \frac{1}{2\mu_0} \sum_{n,m} \frac{(n+1)^2}{2n+1} |v_n^m|^2 \\ &+ \frac{1}{2\mu_0} \sum_{n,m} \frac{n(n+1)}{2n+1} |v_n^m|^2 \end{aligned} \quad (4.117)$$

Finally,

$$\langle B^2 \rangle_{S(a)} = \sum_{n,m} (n+1) |v_n^m|^2 \quad (4.118)$$

This is an important result, which establishes how the average squared magnitude of the geomagnetic field over the reference surface depends from the various harmonics and wavelengths. Let us introduce the quantity:



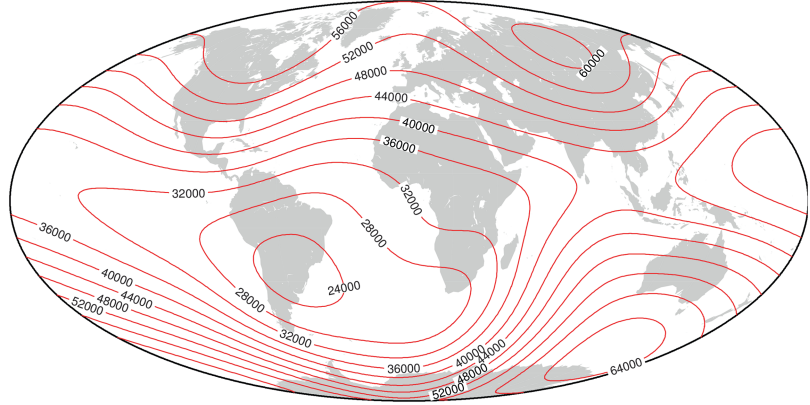
**Fig. 4.26** Power spectrum of geomagnetic field according to the field model of Cain et al. (1989)

$$\begin{aligned} R_n &= (n+1) \sum_{m=-n}^{+n} |v_n^m|^2 \\ &= (n+1) \sum_{m=0}^n \left[ (g_n^m)^2 + (h_n^m)^2 \right] \end{aligned} \quad (4.119)$$

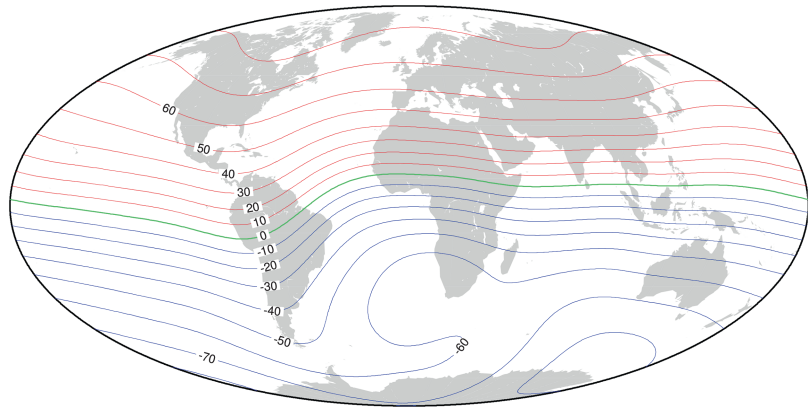
By analogy with time domain spectral analysis, a plot of  $R_n$  as a function of  $n$  can be called the *power spectrum* (Lowes 1974). Figure 4.26 shows the power spectrum for the field model of Cain et al. (1989). We note that the spectrum breaks into two parts, with a transitional region from  $n = 13$  to  $n = 16$ . An obvious interpretation of this result is that the two spectra are expressions of distinct sources. In fact, there exists very strong evidence that the terms from  $n = 1$  to  $n = 13$  are representative of the core field, whereas the crustal contribution would be limited to the terms with  $n > 13$ . However, this separation of core and crustal terms is not perfect, because large-scale features of the crustal field are also contained in the terms from  $n = 1$  to  $n = 13$ , just as short wavelength features of the core field are included in the  $n > 13$  series.

The *International Geomagnetic Reference Field* (IGRF) is a geomagnetic field model representative of core sources, produced and maintained by a team of modelers under the auspices of the *International Association of Geomagnetism and Aeronomy* (IAGA). It is based on a least squares parametric regression of observed data by a truncated version of the spherical harmonic expansion (4.93). The highest

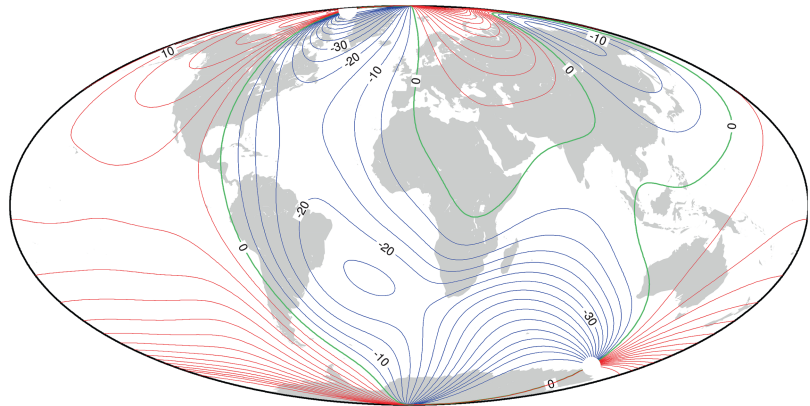
**Fig. 4.27** Total field intensity,  $F$ , in the IGRF 2010.0 field model (Finlay et al. 2010). Units are nT



**Fig. 4.28** Magnetic inclination,  $I$ , in the IGRF 2010.0 field model (Finlay et al. 2010). Units are deg



**Fig. 4.29** Magnetic declination,  $D$ , in the IGRF 2010.0 field model (Finlay et al. 2010). Units are deg



degree and order of the spherical harmonics,  $N$ , is chosen so that the coefficients of the model are reliably determined given the available coverage and quality of observations. The model can be used to predict the large scale features of the internal field. The model is updated at intervals of 5 years (*epochs*) to incorporate new data and is available for the time interval since 1900.0 A.D. To take into account of the secular variation, the

model also specifies a rate of variation of the Gauss coefficients within each epoch. In the 11th generation model, IGRF – 11,  $N$  was chosen to be 10 up to and including epoch 1995.0, and 13 for subsequent epochs, which included excellent satellite data provided by the Ørsted and CHAMP satellites (Finlay et al. 2010). The latest IGRF model is shown in Figs. 4.27, 4.28 and 4.29 (total field intensity, inclination, and declination). We

note the presence of a prominent intensity low,  $F \sim 23,000$  nT, in Argentina (Fig. 4.27), which is known as *South Atlantic anomaly* (Heirtzler 2002). This field low is associated with a region of intense radiation in the space near the Earth, which causes damage to spacecrafts in low Earth orbit and is a hazard for astronauts. The magnetic inclination map (Fig. 4.28) shows a roughly dipolar trend, especially in the tropical belt. Conversely, the declination plot of Fig. 4.29 shows a much more complex pattern. Recall that a pure tilted dipole generates a field with two *agonic lines* (contours of zero  $D$ ) and a single *dip equator* (where  $I = 0$ ). The contour lines of Fig. 4.29 show three agonic lines: one through the Americas, one that encompasses Europe, western Asia, and northeast Africa, and one that crosses the northwestern Pacific, southeast Asia, and western Australia.

In the next chapter, we shall see that the IGRF is an essential tool for the modelling of marine magnetic anomalies, because it allows to separate the crustal components of the observed field from the contribution of the Earth's core.

## Problems

1. Write a computer program to convert from geographic to geomagnetic coordinates and vice versa;
2. Determine an expression for the field components starting from the spherical harmonic expansion (4.93);
3. Download the MS Excel worksheet CK-GTS2004.xls, containing a combined time scale Cande and Kent (1995) and Gradstein et al. (2004). Plot the frequency distribution and the cumulative frequency distribution of the lengths of polarity intervals. What kind of distribution results?;
4. Perform a Fourier analysis of the time scale CK-GTS2004.xls, plotting the power density as a function of the frequency of inversions;
5. Describe quantitatively the motion of a charged particle in the auroral zone, where the magnetic field lines are close to being vertical and the field intensity increases as approaching the Earth;
6. How does the ring current influences the Earth's magnetic field, assuming that it is formed by protons concentrated in a thin band having  $1 \text{ km}^2$  cross-section at distance  $r = 4R_e$  with density  $n_p = 4 \text{ nA m}^{-2}$ ?

## References

- Blakely RJ (1996) Potential theory in gravity and magnetic applications. Cambridge University Press, Cambridge, UK, p 441
- Buffet BA (2000) Earth's core and the geodynamo. *Science* 288:2007–2012. doi:10.1126/science.288.5473.2007
- Butler RF (1992) Paleomagnetism: magnetic domains to geologic terranes. Blackwell Scientific Publications, Boston, p 238
- Cain JC, Wang Z, Schmitz DR, Meyer J (1989) The geomagnetic spectrum for 1980 and core-crustal separation. *Geophys J Int* 97(3):443–447
- Campbell WH (2003) Introduction to geomagnetic fields, 2nd edn. Cambridge University Press, Cambridge, p 337
- Campbell WH, Schiffmacher ER, Kroehl HW (1989) Global quiet day field variation model WDCA/SQ1. *Eos Trans AGU* 70:66–74
- Cande SC, Kent DV (1995) Revised calibration of the geomagnetic time scale for the late Cretaceous and Cenozoic. *J Geophys Res* 100(B4):6093–6095
- Courtillot V, Le Mouél JL (1988) Time variations of the earth's magnetic field-from daily to secular. *Annu Rev Earth Planet Sci* 16:389–476
- Creel KM (1977) Geomagnetic secular variations during the last 25 000 years: an interpretation of data obtained from rapidly deposited sediments. *Geophys J R Astron Soc* 48:91–109
- Dobson DP, Brodholt JP (2000) The electrical conductivity and thermal profile of the Earth's mid-mantle. *Geophys Res Lett* 27(15):2325–2328
- Elsasser WM (1939) On the origin of the Earth's magnetic field. *Phys Rev* 55:489–498
- Finlay CC et al (2010) International geomagnetic reference field: the eleventh generation. *Geophys J Int* 183(3):1216–1230. doi:10.1111/j.1365-246X.2010.04804.x
- Glatzmaier GA, Roberts PH (1995) A three-dimensional convective dynamo solution with rotating and finitely conducting inner core and mantle. *Phys Earth Planet Inter* 91(1):63–75

- Gradstein FM, Agterberg FP, Ogg JG, Hardenbol J, van Veen P, Thierry J, Huang Z (1994) A Mesozoic time scale. *J Geophys Res* 99(B12): 24 051–24 074
- Gradstein FM et al (2004) *Geologic time scale 2004*. Cambridge University Press, Cambridge
- Hathaway DH (2010) The solar cycle. *Living Rev Solar Phys* 7:65
- Heirtzler JR (2002) The future of the South Atlantic anomaly and implications for radiation damage in space. *J Atmos Solar-Terr Phys* 64(16): 1701–1708
- Heirtzler JR, Dickson GO, Herron EM, Pitman WC III, Le Pichon X (1968) Marine magnetic anomalies, geomagnetic field reversals, and motions of the ocean floor and continents. *J Geophys Res* 73: 2119–2136
- Hitchman AP, Lilley FEM, Campbell WH (1998) The quiet daily variation in the total magnetic field: global curves. *Geophys Res Lett* 25(11): 2007–2010
- Korhonen JV et al (2007) Magnetic anomaly map of the World (and associated DVD), scale: 1:50,000,000, 1st edn. Commission for the Geological Map of the World, Paris
- Lowes FJ (1966) Mean-square values on sphere of spherical harmonic vector fields. *J Geophys Res* 71(8):2179. doi:10.1029/JZ071i008p02179
- Lowes FJ (1974) Spatial power spectrum of the main geomagnetic field, and extrapolation to the core. *Geophys J Roy Astron Soc* 36(3):717–730
- Lowrie W, Alvarez W (1977) Upper Cretaceous–Paleocene magnetic stratigraphy at Gubbio, Italy, III. Upper Cretaceous magnetic stratigraphy. *Geol Soc Am Bull* 88:374–377
- Malin SRC, Bullard E (1981) The direction of the Earth's magnetic field at London, 1570–1975. *Philos Trans R Soc London, Ser A* 299(1450):357–423
- Mankinen EA, Dalrymple GB (1979) Revised geomagnetic polarity time scale for the interval 0–5 m.y. B.P. *J Geophys Res* 84(B2):615–626
- Maus S, Rother M, Hemant K, Stolle C, Lühr H, Kuvshinov A, Olsen N (2006) Earth's lithospheric magnetic field determined to spherical harmonic degree 90 from CHAMP satellite measurements. *Geophys J Int* 164(2):319–330
- McElhinny MW, McFadden PL (2000) *Paleomagnetism: continents and oceans*. Academic, San Diego, p 382
- McElhinny MW, McFadden PL, Merrill RT (1996) The time-averaged paleomagnetic field 0–5 Ma. *J Geophys Res* 101(B11):25007–25027
- Merrill RT, McFadden PL (2003) The geomagnetic axial dipole field assumption. *Phys Earth Planet Inter* 139(3):171–185
- Schunk RW, Nagy AF (2009) *Ionospheres*, 2nd edn. Cambridge University Press, Cambridge, p 628
- Tivey MA (1996) Vertical magnetic structure of ocean crust determined from near-bottom magnetic field measurements. *J Geophys Res* 101(B9):20275–20296. doi:10.1029/96JB01307
- Tsurutani BT, Gonzalez WD, Lakhina GS, Alex S (2003) The extreme magnetic storm of 1–2 September 1859. *J Geophys Res* 108(A7):1268. doi:10.1029/2002JA009504
- Vine FJ, Matthews DH (1963) Magnetic anomalies over oceanic ridges. *Nature* 199:947–949
- Wilmot-Smith AL, Priest ER, Hornig G (2005) Magnetic diffusion and the motion of field lines. *Geophys Astrophys Fluid Dyn* 99(2):177–197
- Winch DE, Ivers DJ, Turner JPR, Stening RJ (2005) Geomagnetism and Schmidt quasi-normalization. *Geophys J Int* 160(2):487–504
- Yokoyama N, Kamide Y (1997) Statistical nature of geomagnetic storms. *J Geophys Res* 102(A7): 14215–14222



## Abstract

This chapter is devoted to the techniques of collecting and analysing magnetic data on oceanic crust. These data represent the main source for the determination of global plate motions during the Cenozoic and part of the Mesozoic. I describe the methods to design magnetic surveys, process the raw data, and build isochron maps of oceanic regions.

## 5.1 Magnetic Anomalies

Marine magnetic anomalies represent a fundamental data component in plate kinematics studies, as they furnish most of the basic information that is necessary to unravel the evolution of the world's oceanic basins through the geological time. Vine and Matthews (1963) were the first to discover that the magnetic signature of sea-floor spreading is represented by long stripes having alternate polarity of crustal magnetization. Starting from this pioneer work, the analysis of the magnetic anomaly field across the world's oceans has been a formidable tool for the study of the spreading history along the oceanic ridges and the determination of global plate motions. In particular, if we assume that the highly magnetic 2A layer of the oceanic crust can be represented by an assembly of long prisms having uniform remnant magnetization, then the boundaries between adjacent blocks having opposite directions of magnetization (normal or reversed) are magnetic lineations that can be combined with fracture zone trends to form

magnetic isochrons. These lines are used in turn in the determination of the finite reconstruction (Euler) poles that describe the relative position of two plates in the geologic past (e.g., Klitgord and Schouten 1986; Matias et al. 2005).

In Chap. 4, we have seen that the observed magnetic field at a location  $\mathbf{r}$  near the Earth's surface at time  $t$ ,  $\mathbf{T} = \mathbf{T}(\mathbf{r}, t)$ , results from the superposition of three basic component fields. The main contribution comes from the core field,  $\mathbf{F} = \mathbf{F}(\mathbf{r}, t)$ , whose model is represented by an IGRF for the corresponding epoch. This field varies from being horizontal with magnitude  $\sim 30,000$  nT near the Equator to vertical and with magnitude  $\sim 60,000$  nT near the poles; the root mean square (rms) magnitude of the field vector over the Earth's surface is  $\sim 45,000$  nT. This field also varies in time, on a timescale of months and longer, as a consequence of the secular variation, which is on average  $\sim 80$  nT/year. The second contribution is the generally small external field  $\mathbf{S} = \mathbf{S}(\mathbf{r}, t)$ , associated with solar and ionosphere activity and with currents induced in the crust, which has variable magnitude. In favourable conditions, the  $S_q$  variations determine changes

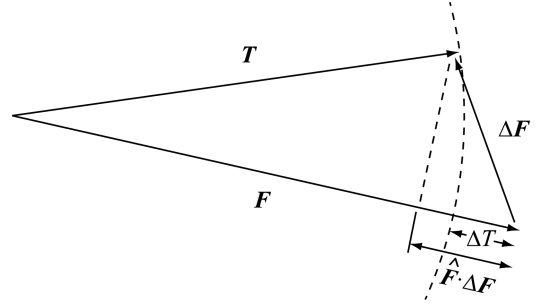
having primarily frequencies of 24, 12, 8, and 6 h and amplitudes of only a few tens nT. However, we have seen that the external contribution to the geomagnetic field can reach 1,000 nT during magnetic storms. Finally, the third important contribution to the total magnetic field is represented by an “anomalous” field  $\Delta\mathbf{F} = \Delta\mathbf{F}(\mathbf{r}, t)$  (intended as perturbation of the main core field) associated with the remnant and induced magnetizations of crustal rocks. This field can be considered as a time independent field when the main component of magnetization is the remnant magnetization, a condition which is generally met by oceanic basalts (see Table 3.1). Therefore, the total magnetic field vector that is observed at the Earth’s surface can be written as follows:

$$\mathbf{T}(\mathbf{r}, t) = \mathbf{F}(\mathbf{r}, t) + \mathbf{S}(\mathbf{r}, t) + \Delta\mathbf{F}(\mathbf{r}) \quad (5.1)$$

A *total field magnetic anomaly* is calculated from *scalar* field measurements by subtracting the reference core field, usually an IGRF, and eventually applying a diurnal correction, which removes those components of the measured field associated with solar and ionospheric activity. Let  $T = T(\mathbf{r}, t)$  be the observed magnitude of total field at location  $\mathbf{r}$  and time  $t$ , which can be obtained by a scalar magnetometer survey. Let  $F = F(\mathbf{r}, t)$  be the IGRF field at the same point and time. Finally, let us assume that an estimate of the external contribution to the magnitude of the observed field, that is a *diurnal correction*  $\Delta S = \Delta S(\mathbf{r}, t)$ , is available. Then the total field anomaly is defined as:

$$\Delta T(\mathbf{r}, t) = T(\mathbf{r}, t) - F(\mathbf{r}, t) - \Delta S(\mathbf{r}, t) \quad (5.2)$$

In the next section, we shall see that an estimation of the external components in Eq. (5.2) can be performed using nearby magnetic observatory data and/or a special design of the survey tracks. Unfortunately, most oceanic surveys are performed far away from magnetic observatories, and the ship-track design generally must satisfy the requirements of other kinds of geophysical measurement. Therefore, the calculation of marine magnetic anomalies is often performed by simple subtraction of the reference field from the



**Fig. 5.1** Relationship between main (*core*) field  $\mathbf{F}$ , observed field  $\mathbf{T}$ , and anomalous field  $\Delta\mathbf{F}$  in the definition of magnetic anomalies

observed data. Now we want to give a physical significance to the expression (5.2). To this purpose, we note that the field  $\Delta\mathbf{F}$  in (5.1) can be considered as a small perturbation to the main reference field, caused by the magnetization of crustal rocks. In fact, ignoring the external contribution, the average magnitude of the observed field is  $\sim 45,000$  nT, whereas crustal field magnitudes in the oceans generally do not exceed 500 nT.

Following Blakely (1996), we also observe that the total field anomaly  $\Delta T$  defined in (5.2) is not equivalent to the magnitude of the anomalous field,  $\Delta F$ , because  $\Delta T = |\mathbf{F} + \Delta\mathbf{F}| - |\mathbf{F}| \neq \Delta F$ , as illustrated in Fig. 5.1. However, for  $\Delta F \ll |\mathbf{F}|$  we can write:

$$\begin{aligned} \Delta T &= |\mathbf{F} + \Delta\mathbf{F}| - |\mathbf{F}| \\ &\cong \sqrt{\mathbf{F} \cdot \mathbf{F} + 2\mathbf{F} \cdot \Delta\mathbf{F}} - |\mathbf{F}| \\ &= F \sqrt{1 + \frac{2\mathbf{F} \cdot \Delta\mathbf{F}}{\mathbf{F} \cdot \mathbf{F}}} - F \cong \\ &\cong F \left[ 1 + \frac{\mathbf{F} \cdot \Delta\mathbf{F}}{\mathbf{F} \cdot \mathbf{F}} \right] - F = F \frac{\mathbf{F} \cdot \Delta\mathbf{F}}{\mathbf{F} \cdot \mathbf{F}} = \hat{\mathbf{F}} \cdot \Delta\mathbf{F} \end{aligned} \quad (5.3)$$

Therefore, a total field anomaly  $\Delta T$  approximately coincides with the *projection* of the anomalous field  $\Delta\mathbf{F}$  onto the reference field axis. In other words,  $\Delta T$  approximates the component of the field generated by the crustal sources in the direction of the regional field. Typical total field oceanic anomalies range from a few nT to thousands of nT, with an rms value of 200–300 nT. Therefore, the condition  $|\mathbf{F}| \gg \Delta F$  is usually met. Note that in general

$\Delta T$  is *not* a function of the position only, even if we consider  $\Delta \mathbf{F}$  as a time independent vector quantity, because it is obtained by projecting  $\Delta \mathbf{F}$  onto a time-varying field direction.

Magnetic anomalies are scalar fields that can be considered as harmonic functions under some simplifying assumptions. In general, the expression (5.2) does not warrant the harmonicity of  $\Delta T$  in any region. However, if  $\hat{\mathbf{F}}$  is approximately constant over the survey area, then taking the Laplacian of (5.3) gives:

$$\nabla^2 \Delta T = \hat{\mathbf{F}} \cdot \nabla^2 \Delta \mathbf{F} = 0 \quad (5.4)$$

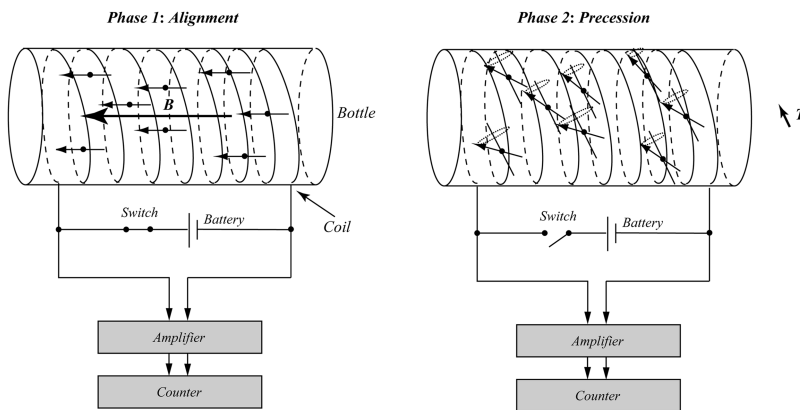
In fact, the components of the crustal field  $\Delta \mathbf{F}$  are harmonic functions. Therefore, under the assumptions that  $\Delta F \ll |\mathbf{F}|$  and that  $\hat{\mathbf{F}}$  is approximately constant over the survey area, a magnetic anomaly field can be considered as a harmonic function.

## 5.2 Acquisition and Pre-processing of Raw Magnetic Data

The acquisition of magnetic data that can be used in studies of crustal magnetization, in particular in plate tectonics, is done mainly through *scalar*

*magnetometers*, measurements devices that furnish the magnitude of the ambient magnetic field. Today there are essentially three kinds of scalar magnetometers: *proton precession magnetometers*, *Overhauser effect magnetometers*, and *caesium vapor magnetometers*. Proton precession magnetometers are commonly used in land-based magnetic surveys. The operation of these measuring devices is based on the Larmor precession of magnetic moments in an external field (see Sect. 3.2). However, differently from the example of Fig. 3.5, in this case the elementary magnetic moments are that of hydrogen nuclei (i.e., protons). Therefore, basically a proton precession magnetometer is formed by a cylindrical container (*bottle*) filled with a liquid rich in hydrogen atoms. The bottle is surrounded by a coil, as shown in Fig. 5.2.

The liquid in the bottle can be distilled water, kerosene, alcohol, or any other hydrocarbon fluid rich in hydrogen atoms. The measurement proceeds in two steps. During phase 1, a DC current is delivered by a battery through the circuit, generating a relatively strong magnetic field  $\mathbf{B}$  in the solenoid. This field causes the dipole moments to align with the solenoid axis in the direction of  $\mathbf{B}$ . The objective of this polarization is to start a coherent precession of the protons about the external field  $\mathbf{T}$  and

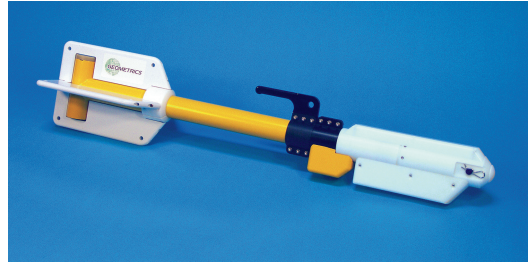


**Fig. 5.2** Operation of proton precession magnetometers. During phase 1, the current flowing through the coil generates a strong magnetic field  $\mathbf{B}$  in the bottle, which determines the alignment of atomic spins. When the circuit is open (phase 2), the magnetic moments start precessing about the ambient geomagnetic field  $\mathbf{T}$ . The Larmor pre-

cession frequency (Eq.3.20) depends from atomic parameters and the field strength  $T$ . This precession induces a small alternating current in the coil, with frequency equal to the Larmor frequency of precession, which is measured through the frequency counter. Finally, the field intensity  $T$  is calculated scaling the observed Larmor frequency

generate a measurable alternate current at the Larmor frequency. Therefore, after a small time interval ( $\sim 1$  s), in phase 2 the DC power supply is switched off. The torque exerted by the external field determines now a precession of the magnetic moments about the field direction, with a Larmor frequency that is proportional to the strength of the total field ( $\sim 2$  kHz). This produces a weak rotating magnetic field that induces a small alternate current through the circuit. This current decays exponentially with a time constant of a few seconds. Therefore, it is promptly amplified and sent to a digital counter, which furnishes a signal frequency proportional to the geomagnetic field intensity. The main disadvantage of proton precession magnetometers is represented by their sensitivity to the orientation of the external field. For example, if the polarizing field  $\mathbf{B}$  has the same direction as  $\mathbf{T}$ , then no precession will occur and the induced current will be zero. In general, the induced signal strength is proportional to the sine of the angle between the Earth's magnetic field and the axis of the solenoid. An improvement with respect to the basic design illustrated in Fig. 5.2 can be obtained using toroidal cores instead of linear solenoids. The accuracy of this kind of magnetometers can reach 0.5 nT.

Overhauser effect magnetometers are an improved class of proton precession magnetometers. The bottle liquid of these sensors contains a chemical additive formed by free radicals, which allows a different method of polarization. In this approach, the spins of unbound electron of the free radicals are polarized through a low-power radiofrequency electromagnetic field. Then, the polarization of electrons is spontaneously transferred to the protons in the liquid via a nuclear magnetic resonance phenomenon known as Overhauser effect. This method allows to reduce the required power supply by one order of magnitude and to increase the sensitivity by two orders of magnitude. An Overhauser magnetometer is capable to perform readings with a 0.01–0.02 nT standard deviation while sampling once per second. This kind of magnetometer is widely used in marine geophysics and is installed at several



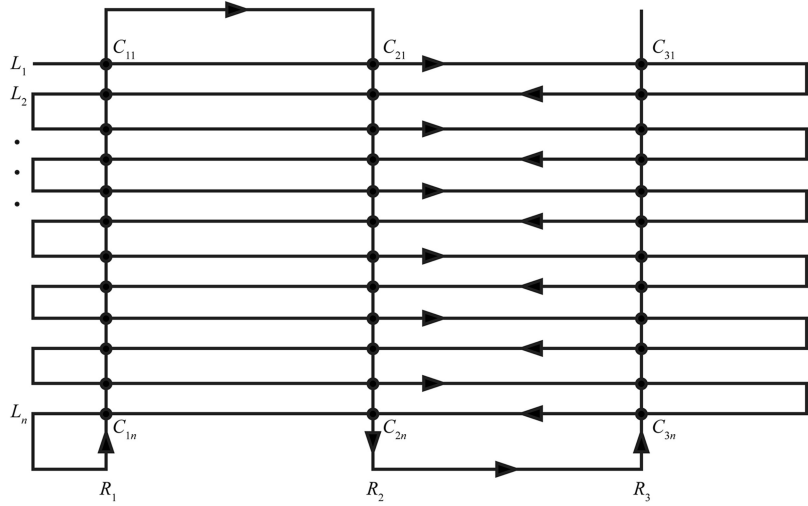
**Fig. 5.3** The G-882 marine magnetometer (Picture courtesy of Geometrics)

magnetic observatories. It was also used on the Ørsted and CHAMP satellites.

The state-of-art technology in scalar magnetometers is undoubtedly represented by alkali atoms (e.g., helium, caesium) vapor sensors. In these devices, polarized laser light is transmitted through a glass cell containing a vapor of the alkaline substance. Again, the idea is to create an initial alignment of spins, which then precess about the external field axis. The sensitivity of these devices can be as high as 0.002 nT at 1 s sample rate. Figure 5.3 shows an example of caesium vapor magnetometer used in marine geophysics.

The importance of oceanic magnetic surveys for plate kinematics has been repeatedly emphasized in the previous chapters. However, plate tectonic modelling also relies on structural information collected along continental margins and continental plate boundaries. In these areas, even small-scale land surveys can help the identification of geological structures. Ground-based magnetic surveys usually require one or two proton precession or Overhauser magnetometers. Data are acquired walking across a rectangular grid pattern and taking readings at more or less equally spaced grid nodes (called *stations*). The average spacing of stations depends from the wavelength of the anomalies that must be mapped and ranges from less than 1–10 m. To remove the diurnal variations from the signal, it is possible to use a second sensor, which is called *base-station magnetometer*. This is kept fixed at a nearby location, in order to record geomagnetic field variations associated

**Fig. 5.4** Example of marine or aeromagnetic survey pattern (black line), formed by 12 survey lines,  $L_i$ , and 3 tie lines,  $R_k$ . Crossover points,  $C_{ik}$ , are shown as black dots



exclusively with ionospheric and magnetospheric activity. In fact, at the time scale of a magnetic survey (a few days) secular variations can be neglected. Alternatively, if a magnetic observatory exists at reasonable distance (within a few tens km), it is possible to use the total field intensity magnetogram of that station to perform the diurnal correction. These data can be easily downloaded from the *Intermagnet* network at <http://www.intermagnet.org>. The procedure is simple. Let  $T_0 = T_0(t)$  be a base-station magnetogram with 1 s sampling rate and  $T_k = T(r_k, t_k)$  the observed total field intensities at locations  $r_k$  at times  $t_k$  ( $k = 1, 2, \dots, n$ ). Then, the sequence corrected for external field variations can be obtained applying the transformation:

$$T_k \rightarrow T_k - [T_0(t_k) - T_0(t_1)] \equiv T_k - \Delta T_0(t_k);$$

$$k = 1, 2, \dots, n \quad (5.5)$$

This procedure can be also applied to aeromagnetic surveys on continental regions. It only requires a series of base-station magnetometers placed at fixed locations in the survey area. In absence of these additional magnetometers, or in the case of oceanic surveys far away from continental margins, a different procedure must be used. In this instance, the track line is formed by a series of *survey lines*, which are oriented

according to geological considerations, and some transverse *tie lines* (Fig. 5.4).

The idea is to have duplicate measurements at some stations, which are called *crossover points*. A crossover point is placed at the intersection between a survey line and a tie line. The difference between the two readings gives information about the diurnal variations of the geomagnetic field, as we shall see in the next section. In this instance, the correction for external field daily variations requires a special algorithm, and the pre-processing of raw data will consist essentially into the removal of *spikes*, either manually or through a filtering procedure, and into the *parallax correction*, which adjusts the data coordinates to take into account for the sensor offset relative to the ship (or to the airplane).

Although most modern marine magnetic surveys are made using the criteria described above, many “historical” surveys across the World’s oceans, made during the 1960s and 1970s, are simply long ship tracks that cross the oceanic basins from one continental margin to the opposite margin. In some cases these data were taken during magnetic storms, and there is no way to remove the contribution of the external fields. Therefore, it is necessary to be cautious when using these data in plate kinematics. The minimum pre-processing work consists into a recalculation of the magnetic anomalies, starting

from the observed total intensity data, on the basis of a *definitive* version of the IGRF for the survey epoch (that is, a *DGRF*).

### 5.3 Levelling Techniques

The procedure of normalization of ship-track or aeromagnetic data according to the information given by a set of crossover points is called *levelling*. The objective of this technique is to reduce systematically the *intersection errors* (difference between two readings) at crossover points. This is made through an iterative algorithm that distributes the errors between tie lines and survey lines (Luyendyk 1997). Although in origin the method was developed as an alternative to the use of base station data (Yarger et al. 1978; Nabighian et al. 2005), today it is commonly used even after the application of base-station diurnal corrections, in conjunction with a survey track design that includes tie lines. In fact, in absence of levelling magnetic anomaly maps often show small long wavelength artifacts aligned with the survey lines that can be misleading during the phase of structural interpretation. The application of a levelling procedure is important for the production of reliable magnetic anomaly maps, especially in regions of severe space weather conditions at high latitudes, where diurnal variations may be as high as 100 nT (Mauring et al. 2002).

The basic idea behind levelling is that the crossover errors,  $\varepsilon_{ki}$ , form a time sequence of data that can be used to estimate the diurnal drift function through a regression procedure. Although several algorithms have been designed for the levelling of magnetic data, here we are going to describe two simple methods that are widespread in the applied geophysics community: (1) *Polynomial fitting* of crossover errors, and (2) *Median filtering*. A procedure of polynomial fitting starts with a *zero-order* levelling step. Let  $\langle \varepsilon \rangle_k$  be the average crossover error along tie line  $R_k$  (see Fig. 5.4):

$$\langle \varepsilon \rangle_k = \frac{1}{n} \sum_{i=1}^n \varepsilon_{ki} \quad (5.6)$$

where  $n$  is the number of survey lines. Then, for each point  $\mathbf{r}_j$  on the tie line  $R_k$ , if  $T$  is the observed magnetic field intensity, then the following transformation is performed:

$$T(\mathbf{r}_j) \rightarrow T(\mathbf{r}_j) - \langle \varepsilon \rangle_k \quad (5.7)$$

This transformation is applied to the observed field values of all tie lines. Then, the crossover errors are updated using the new tie line magnetic intensities. The procedure is now repeated for the survey lines  $L_i$ . If there are  $m$  tie lines, then we set:

$$\langle \varepsilon \rangle_i = \frac{1}{m} \sum_{k=1}^m \varepsilon_{ki} \quad (5.8)$$

In this case, for each point  $\mathbf{r}_s$  on the survey line  $L_i$ , the following transformation is performed:

$$T(\mathbf{r}_s) \rightarrow T(\mathbf{r}_s) - \langle \varepsilon \rangle_i \quad (5.9)$$

These transformations do not introduce distortions in the data set, because they simply offset the tie lines or the survey lines. However, the resulting data set is not perfectly levelled and an output magnetic anomaly map would show narrow bands aligned as the survey lines. Therefore, usually a high-order levelling procedure is applied. This consists into a least squares regression of the crossover errors as a time series functional of the elapsed time through low order polynomials (the typical order is 1–3). The regression can be applied to tie lines, survey lines, or to an entire flight. The resulting polynomials, which are called *drift curves*, are then subtracted from the corresponding data set to minimize the crossover errors. In the simplest case, a levelling procedure requires three steps (Mauring et al. 2002): (1) levelling the tie lines; (2) update the crossover errors, and (3) levelling the survey lines. However, more complex procedures can be used to obtain better results (e.g., see the procedure described in Appendix 1 of Luyendyk 1997), and the regression can be also performed using non-polynomial estimators (Sander and Mrazek 1982). A major problem of polynomial regression is that for a given order the degree of smoothing

depends from the number of crossover errors to be fitted. For example, fitting a second order function to 5 points will generally produce a lower smoothing than fitting the same function to 20 points. Therefore, a direct manual assessment of the required polynomial degree could be necessary. Alternatively, it should be possible to use spline regression techniques or other forms of piecewise low-order polynomials to generate drift curves.

Levelling by median filtering of crossover errors is a valid alternative technique to polynomial fitting (e.g., Mauring et al. 2002). This approach is especially convenient for the removal of random noise and does not require pre-processing procedures such as despiking. A median filter is a non-linear filter based on the following algorithm:

**Algorithm 5.1: (Median Filtering Algorithm)**

Input: A time-ordered sequence of crossover errors  $\{\varepsilon_{k1}, \varepsilon_{k2}, \dots, \varepsilon_{kn}\}$ ; filter size  $r$  ( $r$  odd)

Output: A median-filtered sequence  $\{\bar{\varepsilon}_{k1}, \bar{\varepsilon}_{k2}, \dots, \bar{\varepsilon}_{kn}\}$ ;

```
{
1)  $i \leftarrow 1$ ;
2) if  $i \leq \lfloor r/2 \rfloor$  then set  $S \leftarrow \{\varepsilon_{k1}, \varepsilon_{k1}, \varepsilon_{k1}, \dots, \varepsilon_{ki}, \varepsilon_{ki}, \varepsilon_{ki+1}, \dots, \varepsilon_{k,i+r}\}$  (fill missing values with  $\varepsilon_{k1}$ ); jump #6;
3) if  $i > n - \lfloor r/2 \rfloor$  then set  $S \leftarrow \{\varepsilon_{ki}, \varepsilon_{ki+1}, \dots, \varepsilon_{kn}, \varepsilon_{kn}, \varepsilon_{kn}, \dots, \varepsilon_{kn}\}$  (fill missing values with  $\varepsilon_{kn}$ );
4) jump #6;
5)  $S \leftarrow \{\varepsilon_{ki}, \varepsilon_{k,i+1}, \dots, \varepsilon_{k,i+r}\}$ ;
6)  $\bar{\varepsilon}_{k,i+r/2} \leftarrow \text{median}(S)$ ;
7)  $i \leftarrow i + 1$ ;
8) if  $i \leq n - r + 1$  then jump #2;
}
```

In this algorithm, the function *median()* performs a sorting of the input sequence, then it returns the median of the resulting distribution of values. The algorithm uses a running window of size  $r$  over the input data set, which is moved from position  $i = 1$  to position  $i = n - r + 1$ . Given the group of crossover errors in the window at position  $i$ , it simply replaces the middle value of the array by the median

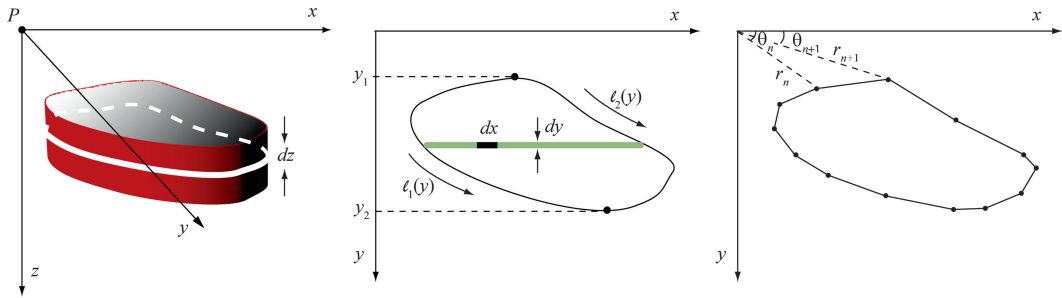
value of the group. If  $i \leq \lfloor r/2 \rfloor$ , then the set  $S$  cannot be filled with  $r$  values whose central position has index  $i + r/2$ , thereby, the set is filled with a series of  $\varepsilon_{k1}$ . For example, if  $r = 5$  and the data set is represented by the sequence:  $\{21, 5, -8, 12, 4, -3, 10, 14, \dots\}$ , then for  $i = 1$  we set  $S \leftarrow \{21, 21, 21, 5, -8\}$ , because the set  $S$  must contain five values. A similar procedure is used at the end of the sequence, for  $i > n - \lfloor r/2 \rfloor$ . In this case (at step #3), the set will be completed by a series of  $\varepsilon_{kn}$ . This algorithm is particularly useful for removing spikes, although it preserves sharp edges in the input data set.

The levelling procedure described above generally produces good results, but small “corrugations” aligned as the survey lines may be still observed on the output magnetic anomaly maps. In this instance, an additional *microlevelling* filtering algorithm is applied, which will remove the residual levelling errors (e.g., Minty 1991; Mauring and Kihle 2006).

## 5.4 Modelling of Marine Magnetic Anomalies

In general, dating the ocean floor, constructing isochron maps that describe the sea floor spreading history of oceanic basins (Sect. 2.5), and discovering the plate kinematics of continents during the geological time, have a common starting point in the modelling of oceanic crust magnetization through the identification of marine magnetic anomalies. Such identification requires in turn calculation of the “anomalous” field  $\Delta F = \Delta F(r)$  associated with an assumed distribution of magnetized blocks of oceanic crust. Then, total field magnetic anomalies are computed for the survey time and compared with the observed anomalies. A best match is found by trial and errors varying the spreading rate function, hence the width of crustal blocks having normal or reversed magnetization according to a geomagnetic polarity time scale (see Sect. 4.4).

Let us consider first the problem of calculating the gravitational or magnetic field generated by



**Fig. 5.5** Stack of laminae in the Talwani and Ewing's (1960) method of calculation. See text for explanation

a distribution of mass or (respectively) magnetization. The method described here was first proposed by Talwani and Ewing (1960) and subsequently modified by Won and Bevis (1987) to improve the computational efficiency. In this approach, a body is represented by a stack of infinitely thin laminae and the boundary of each lamina is approximated by a polygon (Fig. 5.5). The observation point is placed at the origin of the reference frame. It is convenient to consider first the calculation of the gravity field generated by the mass distribution.

By Newton's law of gravitation, the potential  $V$  of a mass distribution is given by:

$$V = G \int_{\mathbf{R}} \frac{\rho(\mathbf{r})}{r} dx dy dz \quad (5.10)$$

where  $\mathbf{R}$  is the region containing the mass distribution,  $\rho = \rho(\mathbf{r})$  is the local mass density, and  $G$  is the gravitational constant. The gravity associated with this potential is:

$$\mathbf{g} = \nabla V = G \int_{\mathbf{R}} \frac{\rho(\mathbf{r}) \hat{\mathbf{r}}}{r^2} dx dy dz \quad (5.11)$$

We are generally only interested to the vertical component of gravity, because gravity meters just measure this quantity. If we indicate this component by  $\gamma$ , then:

$$\gamma = \frac{\partial V}{\partial z} = G \int_{\mathbf{R}} \frac{\rho(x, y, z) z}{(x^2 + y^2 + z^2)^{3/2}} dx dy dz \quad (5.12)$$

This equation has the general form:

$$\gamma = \int_{\mathbf{R}} \rho(x, y, z) \psi(x, y, z) dx dy dz \quad (5.13)$$

where the function:

$$\psi(x, y, z) = G \frac{z}{(x^2 + y^2 + z^2)^{3/2}} \quad (5.14)$$

is called a *Green's function*. We now assume that the density is constant within the region  $\mathbf{R}$ . In this instance, Eq. (5.12) reduces to:

$$\begin{aligned} \gamma &= G \rho \int_{z_1}^{z_2} \left[ \int_{S(z)} \frac{dx dy}{(x^2 + y^2 + z^2)^{3/2}} \right] z dz \\ &\equiv G \rho \int_{z_1}^{z_2} \Gamma(z) z dz \end{aligned} \quad (5.15)$$

where:

$$\Gamma(z) = \int_{S(z)} \frac{dx dy}{(x^2 + y^2 + z^2)^{3/2}} \quad (5.16)$$

The integral  $\Gamma(z)$  represents a surface integral over a single horizontal lamina of the body. As shown in Fig. 5.5, it can be converted into a line integral around the perimeter of the lamina. In fact, let  $(x_1, y_1)$  and  $(x_2, y_2)$  be respectively the points having absolute minimum and maximum of  $y$  within the region  $S(z)$ . We also assume that the region  $S(z)$  has not relative maxima or minima for variable  $y$ . In this instance, the boundary of

$S(z)$  can be represented by two functions  $x = \ell_1(y)$  and  $x = \ell_2(y)$  connecting  $(x_1, y_1)$  to  $(x_2, y_2)$ . For any function  $f(x, y)$  on the surface of the lamina it results:

$$\begin{aligned} \int_{S(z)} f(x, y) dx dy &= \int_{y_1}^{y_2} dy \int_{\ell_1(y)}^{\ell_2(y)} f(x, y) dx \\ &= \int_{y_1}^{y_2} dy [F(\ell_2(y), y) - F(\ell_1(y), y)] \\ &= \oint_{B(z)} F(x, y) dy \end{aligned} \quad (5.17)$$

where  $B(z)$  is the boundary of  $S(z)$ . Therefore, the quantity  $\Gamma(z)$  in (5.16) assumes the following expression:

$$\Gamma(z) = \oint_{B(z)} \frac{x}{(y^2 + z^2) \sqrt{x^2 + y^2 + z^2}} dy \quad (5.18)$$

This integral can be calculated by approximating the perimeter  $B(z)$  of the lamina through a polygon having vertices  $(\xi_1, \zeta_1), (\xi_2, \zeta_2), \dots, (\xi_n, \zeta_n)$ , as shown in Fig. 5.5. This is equivalent to approximate the functions  $\ell_1(y)$  and  $\ell_2(y)$  by piecewise first-order polynomials. Therefore, the line integral (5.18) will be converted into a sum of simple integrals:

$$\begin{aligned} \Gamma(z) &\cong \sum_{n=1}^N \int_{\zeta_n}^{\zeta_{n+1}} \frac{x}{(y^2 + z^2) \sqrt{x^2 + y^2 + z^2}} dy; \\ \zeta_{N+1} &\equiv \zeta_1 \end{aligned} \quad (5.19)$$

The variable  $x$  in this equation can be easily expressed in terms of  $y$ , because the path is composed by straight line segments:

$$x = \alpha_n y + \beta_n \quad (5.20)$$

where:

$$\alpha_n = \frac{\xi_{n+1} - \xi_n}{\zeta_{n+1} - \zeta_n}; \beta_n = \xi_n - \alpha_n \zeta_n \quad (5.21)$$

Finally, substitution in (5.19) provides:

$$\begin{aligned} \Gamma(z) &\cong \sum_{n=1}^N \int_{\zeta_n}^{\zeta_{n+1}} \frac{(\alpha_n y + \beta_n)}{(y^2 + z^2) \sqrt{(\alpha_n^2 + 1)y^2 + 2\alpha_n \beta_n y + \beta_n^2 + z^2}} dy; \\ \zeta_{N+1} &\equiv \zeta_1 \end{aligned} \quad (5.22)$$

The solution of these integrals gives:

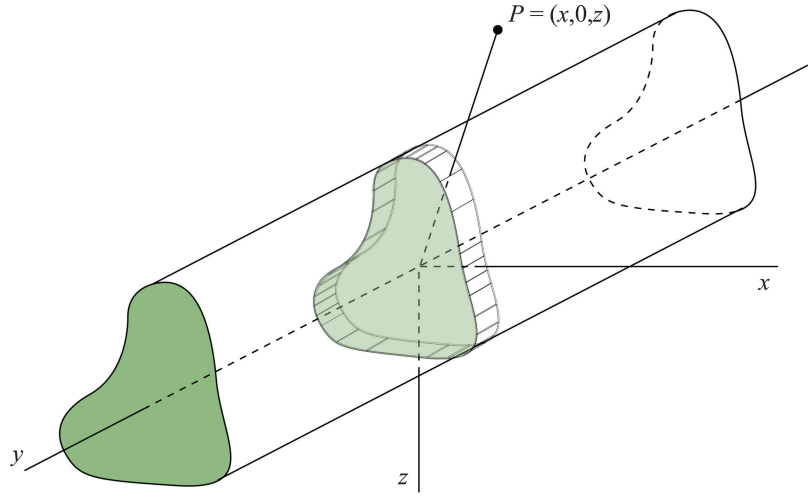
$$\begin{aligned} \Gamma(z) &\cong \sum_{n=1}^N [\arctan \Omega_n(\xi_{n+1}, \zeta_{n+1}, z) \\ &\quad - \arctan \Omega_n(\xi_n, \zeta_n, z)] \end{aligned} \quad (5.23)$$

where:

$$\begin{aligned} \Omega_n(x, y, z) &= \frac{z(\beta_n y - \alpha_n z^2)}{x[(1 + \alpha_n^2)z^2 + \beta_n^2] - (\alpha_n^2 z^2 + \beta_n^2) \sqrt{x^2 + y^2 + z^2}} \end{aligned} \quad (5.24)$$

Substituting the solution (5.23) into (5.15) provides the vertical component of gravity at the origin. In general, integration over  $z$  can be performed using standard numerical techniques and should not constitute a problem. The basic idea of converting a surface integral into a line integral around the surface boundary also represents the starting point of the method proposed by Talwani et al. (1959) for calculating the gravity anomalies of two-dimensional bodies. A geological structure having a linear trend, for example a long horizontal cylinder, generates linear magnetic or gravity anomalies and can be modelled by sources, respectively magnetic or gravitational, that are invariant along the direction parallel to the long side. In this case the  $y$  axis is often chosen parallel to the invariant direction (Fig. 5.6), leaving calculations to be performed only with respect to the  $x$  and  $z$  dimensions. We say that the corresponding problem is two-dimensional. This class of *forward-modelling* problems can be solved by approximating the cross-section of the body by

**Fig. 5.6** Geometry of a two-dimensional problem



an  $N$ -sided polygon, in a way similar to that illustrated in Fig. 5.5. As the density of a two-dimensional source does not vary along the

$y$  dimension, we can set:  $\rho = \rho(x, z)$ . The gravitational potential of a linear body such that illustrated in Fig. 5.6 can be written as:

$$\begin{aligned}
 V(x, 0, z) &= G \int_{\mathbf{R}} \frac{\rho(x', z')}{\sqrt{(x - x')^2 + y'^2 + (z - z')^2}} dx' dy' dz' \\
 &= G \int_S \rho(x', z') \left[ \int_{-\infty}^{+\infty} \frac{dy'}{\sqrt{(x - x')^2 + y'^2 + (z - z')^2}} \right] dx' dz' \\
 &= G \int_S \rho(x', z') \left[ \lim_{a \rightarrow \infty} \int_{-a}^{+a} \frac{dy'}{\sqrt{r^2 + y'^2}} \right] dx' dz' \\
 &= G \int_S \rho(x', z') \left\{ \lim_{a \rightarrow \infty} \left[ \log(a + \sqrt{r^2 + a^2}) - \log(-a + \sqrt{r^2 + a^2}) \right] \right\} dx' dz' \\
 &= G \int_S \rho(x', z') \left\{ \lim_{a \rightarrow \infty} \log \frac{a + \sqrt{r^2 + a^2}}{-a + \sqrt{r^2 + a^2}} \right\} dx' dz' \tag{5.25}
 \end{aligned}$$

where  $r \equiv \sqrt{(x - x')^2 + (z - z')^2}$  and  $S$  is the cross-section of the volume  $\mathbf{R}$  orthogonal to the  $y$  axis. Clearly, as  $a \rightarrow \infty$  the limit in (5.25) diverges, and the potential approaches infinity. This problem is overcome by changing the

definition of the potential for infinitely extended bodies. In fact, we note that if a potential  $V$  satisfies (3.28) (or an equivalent equation for the gravitational field), then  $V + c$  also satisfies this equation for any constant  $c$ . Therefore, the

potential is defined up to an additive constant. Generally, this constant is chosen so that  $V \rightarrow 0$  as  $r \rightarrow \infty$ . In the case of an infinite body, we choose the arbitrary additive constant so that  $V = 0$  at a unit distance from the body ( $r = 1$ ). Therefore, the solution (5.25) is changed as follows:

$$\begin{aligned} V(x, 0, z) &= G \int_S \rho(x', z') \\ &\times \left\{ \lim_{a \rightarrow \infty} \left[ \log \frac{a + \sqrt{r^2 + a^2}}{-a + \sqrt{r^2 + a^2}} \right. \right. \\ &\quad \left. \left. - \log \frac{a + \sqrt{1 + a^2}}{-a + \sqrt{1 + a^2}} \right] \right\} dx' dz' \\ &= 2G \int_S \rho(x', y') \log \frac{1}{r} dx' dz' \end{aligned} \quad (5.26)$$

If we move the observation point to the origin of the reference frame and assume a constant density  $\rho$ , then the vertical component of gravity will be given by:

$$\gamma = -\frac{\partial V}{\partial z} = 2G\rho \int_S \frac{z' dx' dz'}{x'^2 + z'^2} \quad (5.27)$$

Integration over  $x'$  yields:

$$\begin{aligned} \gamma &= 2G\rho \int_{z_1}^{z_2} dz' \left[ \arctan \frac{\ell_2(z')}{z'} - \arctan \frac{\ell_1(z')}{z'} \right] \\ &= 2G\rho \oint_{B(S)} \arctan \frac{x'}{z'} dz' \end{aligned} \quad (5.28)$$

where  $\ell_1$  and  $\ell_2$  are function of  $z'$  and  $B(S)$  is the boundary of  $S$ . As before, we now approximate the perimeter of  $S$  with an  $N$ -sided polygon having vertices  $(\xi_1, \zeta_1), (\xi_2, \zeta_2), \dots, (\xi_n, \zeta_n)$ . In this case, solution (5.28) becomes:

$$\gamma = 2G\rho \sum_{n=1}^N \int_{\zeta_n}^{\zeta_{n+1}} \arctan \frac{x'}{z'} dz'; \zeta_{N+1} = \zeta_1 \quad (5.29)$$

The expression for  $x'$  in terms of  $z'$  is similar to (5.20). We have:

$$x' = \alpha_n z' + \beta_n \quad (5.30)$$

where  $\alpha_n$  and  $\beta_n$  are given by (5.21). Finally, substitution in (5.29) provides:

$$\gamma = 2G\rho \sum_{n=1}^N \left\{ \frac{\pi}{2} (\zeta_{n+1} - \zeta_n) + \left( \zeta_n \arctan \frac{\zeta_n}{\xi_n} - \zeta_{n+1} \arctan \frac{\zeta_{n+1}}{\xi_{n+1}} \right) + \right. \\ \left. + \frac{\beta_n}{1 + \alpha_n^2} \left[ \log \frac{\sqrt{\xi_{n+1}^2 + \zeta_{n+1}^2}}{\sqrt{\xi_n^2 + \zeta_n^2}} - \alpha_n \left( \arctan \frac{\zeta_{n+1}}{\xi_{n+1}} - \arctan \frac{\zeta_n}{\xi_n} \right) \right] \right\}; \xi_{N+1} = \xi_1; \zeta_{N+1} = \zeta_1 \quad (5.31)$$

The first two terms in parentheses of summation give zero after summation. Therefore:

$$\gamma = 2G\rho \sum_{n=1}^N \left\{ \frac{\beta_n}{1 + \alpha_n^2} \left[ \log \frac{\sqrt{\xi_{n+1}^2 + \zeta_{n+1}^2}}{\sqrt{\xi_n^2 + \zeta_n^2}} - \alpha_n \left( \arctan \frac{\zeta_{n+1}}{\xi_{n+1}} - \arctan \frac{\zeta_n}{\xi_n} \right) \right] \right\}; \xi_{N+1} = \xi_1; \zeta_{N+1} = \zeta_1 \quad (5.32)$$

This solution implies that the gravity of a body in a two-dimensional problem only depends upon the coordinates of the vertices of a poly-

gon that approximates its cross-section. Vertex coordinates  $(\xi_n, \zeta_n)$  in (5.32) can be replaced by quantities  $r_n$  and  $\theta_n$  illustrated in Fig. 5.5. In fact,

$$r_n = \sqrt{\xi_n^2 + \zeta_n^2}; \theta_n = \arctan \frac{\zeta_n}{\xi_n} \quad (5.33)$$

In this instance, the solution assumes the following simple form:

$$\begin{aligned} \gamma &= 2G\rho \sum_{n=1}^N \left\{ \frac{\beta_n}{1 + \alpha_n^2} \left[ \log \frac{r_{n+1}}{r_n} \right. \right. \\ &\quad \left. \left. - \alpha_n (\theta_{n+1} - \theta_n) \right] \right\}; \\ \xi_{N+1} &= \xi_1; \zeta_{N+1} = \zeta_1 \end{aligned} \quad (5.34)$$

Therefore, the normal gravity of a two-dimensional problem depends upon the distances of the polygon vertices from the observation point and from the angles of the radii  $r_n$  with respect to the horizontal. In computer modelling software, the angles  $\theta_n$  are calculated through Eq. (5.33) by calling the `atan2()` C language library function. This call may lead to improper evaluation of these quantities when the observation point is located between  $\zeta_n$  and  $\zeta_{n+1}$ . Therefore, the following tests are performed:

if ( $\text{sgn}(\zeta_n) \neq \text{sgn}(\zeta_{n+1})$ ) then

$$\left\{ \begin{array}{l} \text{if } (\xi_n \zeta_{n+1} < \xi_{n+1} \zeta_n \text{ and } \zeta_{n+1} \geq 0) \text{ then } \theta_n \\ \quad \leftarrow \theta_n + 2\pi; \\ \text{else if } (\xi_n \zeta_{n+1} > \xi_{n+1} \zeta_n \text{ and } \zeta_n \geq 0) \text{ then } \theta_{n+1} \\ \quad \leftarrow \theta_{n+1} + 2\pi; \\ \text{else if } (\xi_n \zeta_{n+1} = \xi_{n+1} \zeta_n) \text{ then } \gamma \leftarrow 0; \end{array} \right\}$$

if ( $\xi_n = \zeta_n = 0$  or  $\xi_{n+1} = \zeta_{n+1} = 0$ ) then  $\gamma \leftarrow 0$ ;

if ( $\xi_n = \xi_{n+1}$ ) then

$$\begin{aligned} &\frac{\beta_n}{1 + \alpha_n^2} \left[ \log \frac{r_{n+1}}{r_n} - \alpha_n (\theta_{n+1} - \theta_n) \right] \\ &\quad \leftarrow \xi_n \log \frac{r_{n+1}}{r_n} \end{aligned}$$

The typical conventions for the calculation of normal gravity through (5.34) require clockwise polygons, and a downward directed  $z$  axis, as shown in Fig. 5.6.

Let us come back now to the problem of calculating the magnetic anomaly generated by a magnetized 2-dimensional body. This anomaly can be easily computed by the *Poisson's relation* using the previous expressions for the gravity potential. Let us consider a body with uniform magnetization  $\mathbf{M}$  and mass density  $\rho$ . A small element of the body can be considered as an elementary magnetic dipole having magnetic moment  $\mathbf{m} = \mathbf{M} dxdydz$ . If the observation point is placed at the origin, then a dipole at location  $\mathbf{r}$  generates a small magnetic field  $d\mathbf{B}$  which is approximately given by:

$$d\mathbf{B} \cong \frac{\mu_0}{4\pi} \left( 3 \frac{\mathbf{M} \cdot \mathbf{r}}{r^5} \mathbf{r} - \frac{\mathbf{M}}{r^3} \right) dxdydz \quad (5.35)$$

where we have used (3.21). Therefore, by (3.28) the potential  $dV$  is given by:

$$\begin{aligned} dV &\cong \frac{\mu_0}{4\pi} \frac{\mathbf{M} \cdot \mathbf{r}}{r^3} dxdydz \\ &= -\frac{\mu_0}{4\pi} \mathbf{M} \cdot \nabla \left( \frac{1}{r} \right) dxdydz \end{aligned} \quad (5.36)$$

Integrating (5.36) over the region  $\mathbf{R}$  occupied by the body gives the total magnetic potential  $V$ :

$$\begin{aligned} V &\cong \frac{\mu_0}{4\pi} \int_{\mathbf{R}} \frac{\mathbf{M} \cdot \mathbf{r}}{r^3} dxdydz \\ &= -\frac{\mu_0}{4\pi} \mathbf{M} \cdot \nabla \int_{\mathbf{R}} \left( \frac{1}{r} \right) dxdydz \end{aligned} \quad (5.37)$$

This formula is similar to expression (5.10) for the gravitational force of a mass distribution if the body density is constant. In fact, in this instance (5.10) gives:

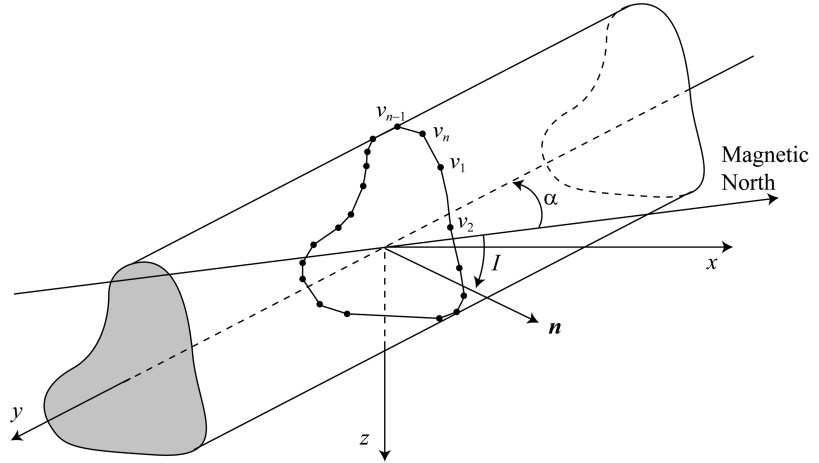
$$V = G\rho \int_{\mathbf{R}} \frac{1}{r} dxdydz \quad (5.38)$$

Then,

$$\mathbf{g} = -\nabla V = -G\rho \nabla \int_{\mathbf{R}} \left( \frac{1}{r} \right) dxdydz \quad (5.39)$$

Therefore, the magnetic potential of a uniformly magnetized body having constant density can be written as:

**Fig. 5.7** Conventions and geometry of a two-dimensional problem. The polygon approximating the body cross-section has vertices  $v_i$ , which are ordered clockwise



$$V = \frac{\mu_0}{4\pi G\rho} \mathbf{M} \cdot \mathbf{g} = \frac{\mu_0\chi}{4\pi G\rho} \mathbf{H} \cdot \mathbf{g} \quad (5.40)$$

where  $\chi$  is the magnetic susceptibility and  $\mathbf{H}$  is the inducing geomagnetic or paleomagnetic field. This expression is called *Poisson's relation*. It states that the magnetic potential of a uniformly magnetized body having constant density is proportional to the component of the gravity field in the direction of magnetization. Therefore, taking the gradient of Eq. (5.40) we obtain that the anomalous field of a body in a 2-dimensional problem can be written as:

$$\Delta \mathbf{F} = \nabla V = \frac{\mu_0 M}{4\pi G\rho} \frac{\partial \mathbf{g}}{\partial n} = \frac{\mu_0 \chi H}{4\pi G\rho} \frac{\partial \mathbf{g}}{\partial n} \quad (5.41)$$

where  $n$  is the direction of induced or remnant magnetization. This solution implies that unlike the gravity anomaly, the magnetic anomaly also depends on the strike of the body, as this affects the direction of magnetization. Let us assume that the body magnetization is purely NRM. Let  $I$  and  $\alpha$  be respectively the mean paleomagnetic field inclination in the survey area, and the strike of the body measured counterclockwise from the paleomagnetic North to the negative  $y$ -axis (Fig. 5.7).

The unit vector associated with the magnetization direction can be written as:

$$\begin{cases} n_x = \cos I \sin \alpha \\ n_y = \cos I \cos \alpha \\ n_z = \sin I \end{cases} \quad (5.42)$$

Therefore, for any scalar field  $\phi$  that is constant in the  $y$  direction, the directional derivative of  $\phi$  along direction  $n$  results to be:

$$\frac{\partial \phi}{\partial n} = \nabla \phi \cdot \mathbf{n} = \frac{\partial \phi}{\partial x} \cos I \sin \alpha + \frac{\partial \phi}{\partial z} \sin I \quad (5.43)$$

It is easy to determine the vertical and horizontal components of the anomalous field combining expressions (5.41) and (5.43). We have:

$$\begin{aligned} \Delta F_x &= \frac{\mu_0 M}{4\pi G\rho} \frac{\partial g_x}{\partial n} \\ &= \frac{\mu_0 M}{4\pi G\rho} \left( \frac{\partial g_x}{\partial x} \cos I \sin \alpha + \frac{\partial g_x}{\partial z} \sin I \right) \end{aligned} \quad (5.44)$$

$$\begin{aligned} \Delta F_z &= \frac{\mu_0 M}{4\pi G\rho} \frac{\partial g_z}{\partial n} \\ &= \frac{\mu_0 M}{4\pi G\rho} \left( \frac{\partial g_z}{\partial x} \cos I \sin \alpha + \frac{\partial g_z}{\partial z} \sin I \right) \end{aligned} \quad (5.45)$$

Quantities  $g_z$  and  $g_x$  can be calculated respectively using (5.34) (because  $g_z \equiv \gamma$ ) and a similar formula for  $g_x$ . These formulae can be written as follows:

$$g_x = 2G\rho \sum_{n=1}^N X_n; g_z = 2G\rho \sum_{n=1}^N Z_n \quad (5.46)$$

where,

$$X_n = \frac{\beta_n}{1 + \alpha_n^2} \left[ \alpha_n \log \frac{r_{n+1}}{r_n} + (\theta_{n+1} - \theta_n) \right] \quad (5.47)$$

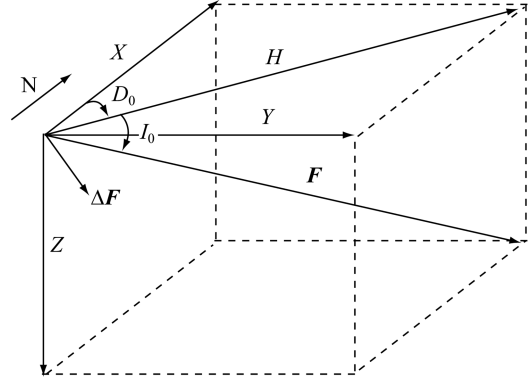
$$Z_n = \frac{\beta_n}{1 + \alpha_n^2} \left[ \log \frac{r_{n+1}}{r_n} - \alpha_n (\theta_{n+1} - \theta_n) \right] \quad (5.48)$$

Therefore, substituting these expressions into (5.44) and (5.45) we have:

$$\Delta F_x = \frac{\mu_0 M}{2\pi} \sum_{n=1}^N \left( \frac{\partial X_n}{\partial x} \cos I \sin \alpha + \frac{\partial X_n}{\partial z} \sin I \right) \quad (5.49)$$

$$\Delta F_z = \frac{\mu_0 M}{2\pi} \sum_{n=1}^N \left( \frac{\partial Z_n}{\partial x} \cos I \sin \alpha + \frac{\partial Z_n}{\partial z} \sin I \right) \quad (5.50)$$

In Sect. 5.1 we have shown that a total field anomaly can be calculated by projecting the anomalous field vector  $\Delta \mathbf{F}$  onto the IGRF field axis (Eq. 5.3). Let  $D_0$  and  $I_0$  be respectively the declination and inclination of the reference field (Fig. 5.8). The components of the anomalous field vector  $\Delta \mathbf{F}$  that are calculated through (5.49) and (5.50) are expressed in the local  $(x,y,z)$  frame of a prism. In order to combine the contributions of several blocks through the superposition principle, we must represent the anomalous vector components in a common reference frame. Then, the expected total field anomaly,  $\Delta T$ , associated with the crustal field  $\Delta \mathbf{F} = \Delta \mathbf{F}(\mathbf{r})$ , can be calculated by projecting the vector  $\Delta \mathbf{F}$  onto the axis of the present-day reference field  $\mathbf{F}$ , which has declination,  $D_0$ , and inclination,  $I_0$ , in the  $(X,Y,Z)$  local reference frame of Fig. 5.8. Therefore, it is convenient to express the components of the anomalous magnetic field vectors generated by each block in the standard  $(X,Y,Z)$  coordinate system. Let  $\beta$  ( $0^\circ \leq \beta < 360^\circ$ ) be the local strike of the  $y$  axis, measured clockwise from the North. As it is shown below, this quantity is determined by the local trend of the flow line. Then, the components of the vector  $\Delta \mathbf{F}$  in the local frame of reference



**Fig. 5.8** Crustal field  $\Delta \mathbf{F}$  and components of the IGRF field  $\mathbf{F}$  in a local frame of reference. Declination  $D_0$  is the angle between the North direction and the horizontal projection of the field, measured clockwise. The angle between projection  $H$  and the field vector  $\mathbf{F}$  is the inclination  $I_0$ , positive if  $\mathbf{F}$  is directed downward

can be easily obtained by the following simple transformation:

$$\begin{cases} \Delta F_X = \Delta F_x \sin \beta \\ \Delta F_Y = -\Delta F_x \cos \beta \\ \Delta F_Z = \Delta F_z \end{cases} \quad (5.51)$$

It should be noted that an anomalous field vector  $\Delta \mathbf{F}$  does not have  $y$ -component in the  $(x,y,z)$  reference frame of a magnetized prism, whereas it has a non-zero  $Y$ -component with respect to the standard  $(X,Y,Z)$  geomagnetic coordinate system. Therefore, the total field anomaly,  $\Delta T$ , at an observation point  $\mathbf{r}$  is given by:

$$\begin{aligned} \Delta T = \Delta \mathbf{F} \cdot \hat{\mathbf{F}} &= \hat{F}_X \Delta F_X + \hat{F}_Y \Delta F_Y \\ &+ \hat{F}_Z \Delta F_Z = \cos I_0 (\Delta F_X \cos D_0 \\ &+ \Delta F_Y \sin D_0) + \Delta F_Z \sin I_0 \end{aligned} \quad (5.52)$$

Evaluation of derivatives in formulae (5.49) and (5.50) is simple:

$$\begin{aligned} \frac{\partial Z_n}{\partial z} &= \frac{(\xi_{n+1} - \xi_n)^2}{R^2} \\ &\times \left[ \frac{\zeta_{n+1} - \zeta_n}{\xi_{n+1} - \xi_n} \log \frac{r_{n+1}}{r_n} - (\theta_{n+1} - \theta_n) \right] - P \end{aligned} \quad (5.53)$$

Furthermore,

$$\frac{\partial Z_n}{\partial x} = -\frac{(\xi_{n+1} - \xi_n)(\zeta_{n+1} - \zeta_n)}{R^2} \times \left[ \log \frac{r_{n+1}}{r_n} - (\theta_{n+1} - \theta_n) \right] + Q \quad (5.54)$$

$$\frac{\partial X_n}{\partial z} = \frac{(\xi_{n+1} - \xi_n)^2}{R^2} \times \left[ \log \frac{r_{n+1}}{r_n} + \frac{\zeta_{n+1} - \zeta_n}{\xi_{n+1} - \xi_n} (\theta_{n+1} - \theta_n) \right] + Q \quad (5.55)$$

Finally,

$$\frac{\partial X_n}{\partial x} = -\frac{(\xi_{n+1} - \xi_n)(\zeta_{n+1} - \zeta_n)}{R^2} \times \left[ \log \frac{r_{n+1}}{r_n} + \frac{\zeta_{n+1} - \zeta_n}{\xi_{n+1} - \xi_n} (\theta_{n+1} - \theta_n) \right] + P \quad (5.56)$$

where,

$$R^2 = (\xi_{n+1} - \xi_n)^2 + (\zeta_{n+1} - \zeta_n)^2 \quad (5.57)$$

$$P = -\frac{\xi_n \zeta_{n+1} - \xi_{n+1} \zeta_n}{R^2} \times \left[ \frac{\xi_n (\xi_{n+1} - \xi_n) - \zeta_n (\zeta_{n+1} - \zeta_n)}{r_n^2} - \frac{\xi_{n+1} (\xi_{n+1} - \xi_n) - \zeta_{n+1} (\zeta_{n+1} - \zeta_n)}{r_{n+1}^2} \right] \quad (5.58)$$

$$Q = -\frac{\xi_n \zeta_{n+1} - \xi_{n+1} \zeta_n}{R^2} \times \left[ \frac{\xi_n (\zeta_{n+1} - \zeta_n) + \zeta_n (\xi_{n+1} - \xi_n)}{r_n^2} - \frac{\xi_{n+1} (\zeta_{n+1} - \zeta_n) + \zeta_{n+1} (\xi_{n+1} - \xi_n)}{r_{n+1}^2} \right] \quad (5.59)$$

Also in this case the evaluation of angles  $\theta_n$  through the `atan2()` C library function must take into account of three special situations.

The first two conditions are the same of the gravity case, whereas the third one is now,

if  $(\xi_n = \xi_{n+1})$  then {

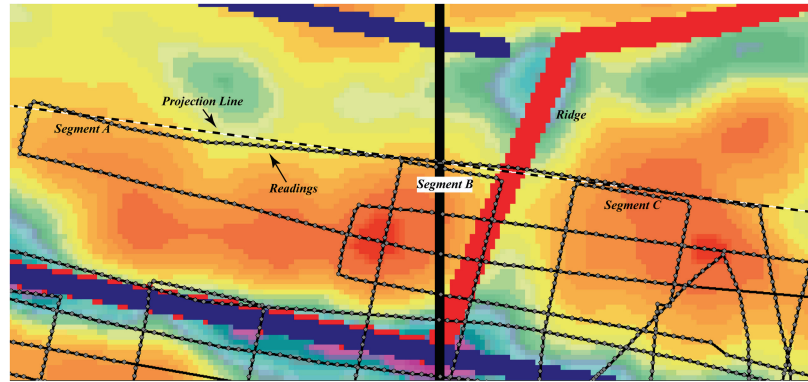
$$\begin{aligned} \frac{\partial Z_n}{\partial z} &= -P; \quad \frac{\partial Z_n}{\partial x} = -\frac{(\zeta_{n+1} - \zeta_n)^2}{R^2} \log \frac{r_{n+1}}{r_n} \\ &+ Q; \quad \frac{\partial X_n}{\partial z} = Q; \quad \frac{\partial X_n}{\partial x} = -\frac{(\zeta_{n+1} - \zeta_n)^2}{R^2} \\ &\times (\theta_{n+1} - \theta_n) + P \} \end{aligned}$$

The results obtained in this section allow creating and testing magnetization models of the oceanic crust. In fact, observation suggests that the pattern of magnetization in oceanic basins generally consists of assemblages of long prisms having uniform remnant magnetization, normal or reversed, running parallel to the spreading ridges (Fig. 4.18). These prisms can be approximated by 2-D bodies, for which the previous solutions allow to determine the resulting crustal field and the corresponding anomaly at any point. The size of these blocks in the  $x$  direction is not arbitrary, but depends from the spreading velocity and the duration of chrons in a geomagnetic polarity time scale. In the next section we shall see the procedures that must be followed to analyse and interpret the observed magnetic signal.

## 5.5 Forward Modelling Procedures

The analysis and interpretation of marine total field anomalies can be performed using either *forward modelling* or *inversion techniques* (e.g., Blakely 1996). A forward modelling algorithm assumes a “reasonable” initial magnetization pattern of the oceanic crust, which is built starting from a geomagnetic polarity time scale and a spreading rate function of time  $t$ ,  $v = v(t)$ . Then, a specialized computer program generates model anomalies,  $\Delta T'$ , that can be compared with the observed data  $\Delta T$ . Generally, the magnetization pattern is modified in a trial and error procedure through successive adjustments of the spreading rate function  $v = v(t)$ , until a good fit of the model anomalies to the observed data is

**Fig. 5.9** A complex ship track in the central Atlantic. In this example, data along the track segment A are projected, together with data from segments B and C, onto a projection line (dashed line) having a fixed strike with respect to the ridge axis. The background image shows gravity anomalies (Sandwell and Smith 1997)



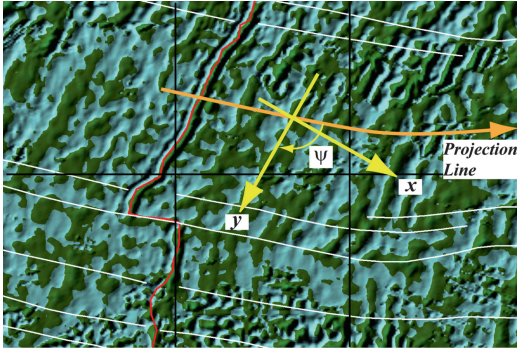
obtained. Alternatively, inversion techniques use the observed anomalies  $\Delta T$  to estimate both geometry and properties of the magnetized bodies (Bott 1967; Backus and Gilbert 1968; Bott and Hutton 1970; Parker and Huestis 1974; Parker 1974). In this instance, the processing method generally requires summation of Fourier transforms of bathymetry and magnetization functions (Parker 1972).

In general, forward modelling techniques have been more commonly employed in marine geophysics studies, whereas inverse modelling is predominant in exploration geophysics. The reason is that inverse models always require a series of simplifying assumptions that may not adequately fit the complexity of the sea floor spreading process. For instance, in the approach of Bott and Hutton (1970), Parker and Huestis (1974), and Parker (1974) the magnetization intensity may only vary horizontally along a traverse. This prevents the possibility to generate magnetization models where the prisms have dipping polarity boundaries as suggested by Tivey (1996). Another limitation of the inverse modelling is the requirement that the direction of magnetization can only change by  $\pm 180^\circ$ . This limitation does not significantly affect short profiles encompassing a few million years of sea floor spreading, but could introduce significant distortion in the shape of the model anomalies when the time interval is longer than a few tens million years.

Now we are going to describe the specific procedures that are used in the forward modelling of marine magnetic anomalies. In the following, we shall assume that the input data set is repre-

sented by a series of ship tracks or aeromagnetic flight lines from an oceanic basin. For example, the National Geophysical Data Center (NGDC) disseminates such data through the GEODAS data base. Ship tracks that can be used in plate kinematics should form an angle between  $40^\circ$  and  $140^\circ$  with the ridge axis, because outside this range it would be hard to identify correctly the anomalies. Therefore, the first step is to select the tracks (or track segments) that can be used in the analysis. An example of ship-track composed by several segments and tie lines is shown in Fig. 5.9.

The second step is to project the data from one or more segments (survey lines) onto a *projection line*, which can be aligned or not with the local direction of spreading. In general, the line of projection will have a unique strike with respect to the magnetized prisms, whereas a track line could swing irregularly about a definite direction. Furthermore, in some cases it is possible to project different neighboring survey lines onto a unique line of projection, in order to generate an averaged magnetic profile, as shown in Fig. 5.9. A computer program like *Magan* (Schettino 2012) uses the local strike of the projection line to define the *profile obliquity angle*,  $\psi$ , with respect to the magnetized prisms, and build a local reference frame according to the conventions used in Fig. 5.6. As an example, in Fig. 5.10 the strike of the magnetized prisms is clearly evidenced by the pattern of crustal magnetic anomalies, extracted from the global grid of Korhonen et al. (2007). Because of spreading obliquity (see Sect. 2.4), in this example we would have an obliquity angle



**Fig. 5.10** Example of determination of the profile obliquity angle  $\psi$

$\psi$  that is greater than  $90^\circ$  even in the case of a projection line oriented as the fracture zones,.

Once the data have been projected, it is necessary to assign the position of the *origin* along the magnetic profile, which will be the point with offset zero in age-distance plots. This point should be placed tentatively along the spreading ridge (or the extinct ridge) as seen on gravity anomaly maps. However, it will be adjusted later to match the observed central anomaly. In order to start a forward modelling procedure, we must assume an initial spreading velocity  $v$ , which will be used to build the initial configuration of the magnetized prisms according to a selected geomagnetic polarity time scale (see Sect. 4.4). Very often, the prisms are draped on bathymetry, with constant height  $H$  equal to the assumed magnetized layer thickness, but it is also possible to build models that are based on estimates of the real depth to the basement (excluding the sedimentary layer).

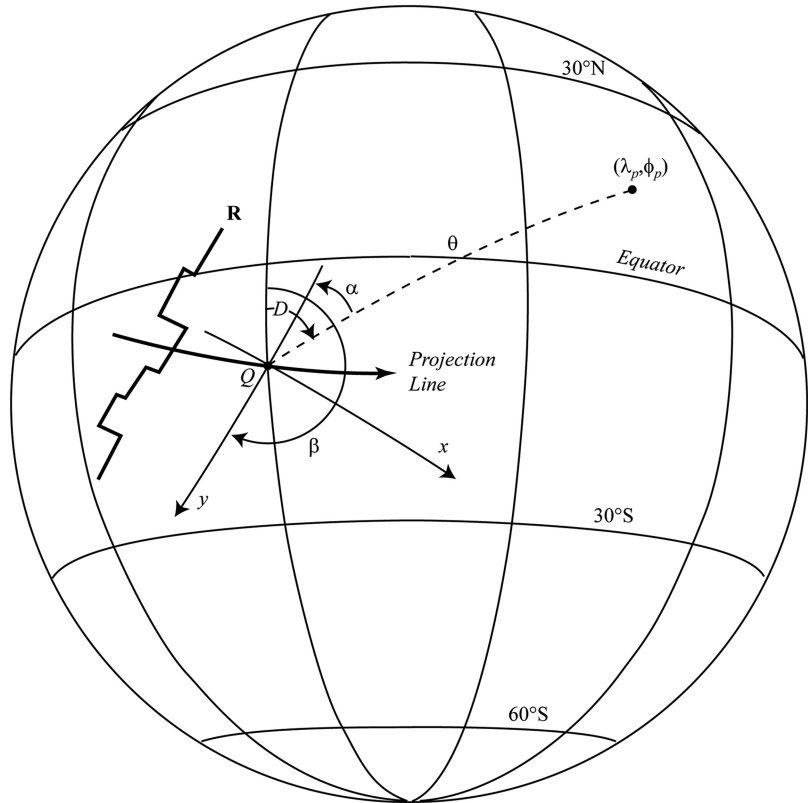
For each chron in the time scale, a polygon that approximates the cross-section of the crustal block that formed during this time interval is built, with horizontal width  $w_k$  proportional to the chron duration:

$$w_k = \frac{1}{2}v\Delta T_k \quad (5.60)$$

where  $v$  is the default full spreading rate and  $\Delta T_k$  is the duration (in Myrs) of the  $k$ -th chron. The polygons are assumed to have uniform magnetization directed as the average paleomagnetic

field vector during the corresponding chrons. In plate kinematics, the magnetization direction of a prism cannot be chosen as coincident with the present day reference field  $\mathbf{F}$ , not even when the data ages encompass the last 2–3 Myrs. In fact, assuming that the rock magnetization is entirely of NRM type, even in the case of rocks that formed during the last polarity chron, the average magnetization direction would be aligned with the *time-averaged geomagnetic field* for the last 0.78 Myrs, which is a GAD field. Therefore, in this instance the paleomagnetic direction in (5.49) and (5.50) would be  $I = 90^\circ$ ,  $D = 0^\circ$  and *not* that of the local IGRF field (i.e.,  $I_0$  and  $D_0$ ). These parameters can also be used for rocks of Pliocene – Pleistocene age, but in general older crust requires a different approach. In the next chapter, we shall see that for a tectonic plate that has been moving around the globe, the NRM directions of its rocks of various ages can be described by a temporal sequence of paleomagnetic fields whose dipole axes migrate in a regular fashion away from the present day Earth's spin axis according to an age progression. This *apparent polar wandering*, which is a consequence of plate motions, must be taken into account in plate kinematics modelling, because it determines a corresponding change of paleomagnetic directions in so far as we move away from a spreading centre. Let  $(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n)$  be a sequence of paleopole positions for one of the two conjugate plates about a spreading ridge. This sequence furnishes the apparent orientation of the spin axis (i.e., the apparent location of the geographic North Pole) during each chron in the time scale, as seen from the reference frame of this plate. In the next chapter, we shall study in detail these *apparent polar wander paths* (APW Paths). For the moment, it is sufficient to say that we can easily compute the paleomagnetic inclination and declination  $(I_k, D_k)$  of the NRM vector at any point along the projection line starting from these paleopoles. A similar procedure can be used to determine the inclination and declination  $(I'_k, D'_k)$  for any point on the tract of projection line placed on the opposite side of the spreading ridge. To this purpose, it could be necessary to have a sequence of paleopoles  $(\mathbf{p}'_1, \mathbf{p}'_2, \dots, \mathbf{p}'_n)$  also for

**Fig. 5.11** Local coordinate system for a magnetized prism at location  $Q$  along a projection line.  $\mathbf{R}$  is the present ridge axis.  $(\lambda_p, \phi_p)$  is the palaeopole location as seen for the prism.  $\theta$  is the paleocolatitude of point  $Q$



the conjugate plate. However, we shall see soon that knowledge of just one APW path is sufficient to calculate the field components.

A declination value can be used, together with the local strike,  $\beta$ , of the  $y$  axis (main axis of the magnetized prism), to calculate the parameter  $\alpha$  that appears in (5.49) and (5.50). Assuming that  $\beta$  is measured clockwise from the North (Fig. 5.11), with  $0^\circ \leq \beta < 360^\circ$ , it can be easily shown that in any case  $\alpha$  is given by:

$$\alpha = D - \beta + 180^\circ \quad (5.61)$$

Let us consider now the oceanic crust that is created along a ridge segment during a polarity chron. At the end of the chron, this is a crustal block formed by two adjacent prisms with identical GAD magnetization and strike  $\alpha$ , which are placed along the opposite flanks of the ridge segment. Therefore, at this time the two prisms have the same inclination,  $I' = I$ , which depends

from the paleolatitude of the spreading segment according to the dipole Eq. (4.49). Clearly, such identity will hold also in the future, because the separated blocks can only rotate about vertical axes and be displaced in so far as sea floor spreading proceeds. In fact, we know that the two magnetized prisms will move apart in opposite directions by a sequence of rotations about stage poles (see Chap. 2). Unlike the inclination, during this process the *direction* of magnetization will change, and it will change differently for the two blocks, so that after some million years the declination  $D$  of the two magnetization vectors will be significantly different. However, it is easily realized that the strike  $\alpha$  of the two prisms with respect to their directions of magnetization remains invariant, so that if  $\beta'$  is the modern strike of the conjugate block and  $D'$  is its declination, we have:

$$D' = \alpha + \beta' - 180^\circ \quad (5.62)$$

Therefore, in the calculation of the field components through (5.49) and (5.50), it is not necessary to know the corresponding sequence of paleopoles ( $\mathbf{p}'_1, \mathbf{p}'_2, \dots, \mathbf{p}'_n$ ) for the conjugate plate. *A conjugate pair of magnetized prisms on the two sides of a spreading ridge has a unique inclination,  $I$ , and a unique paleostrike  $\alpha$ .*

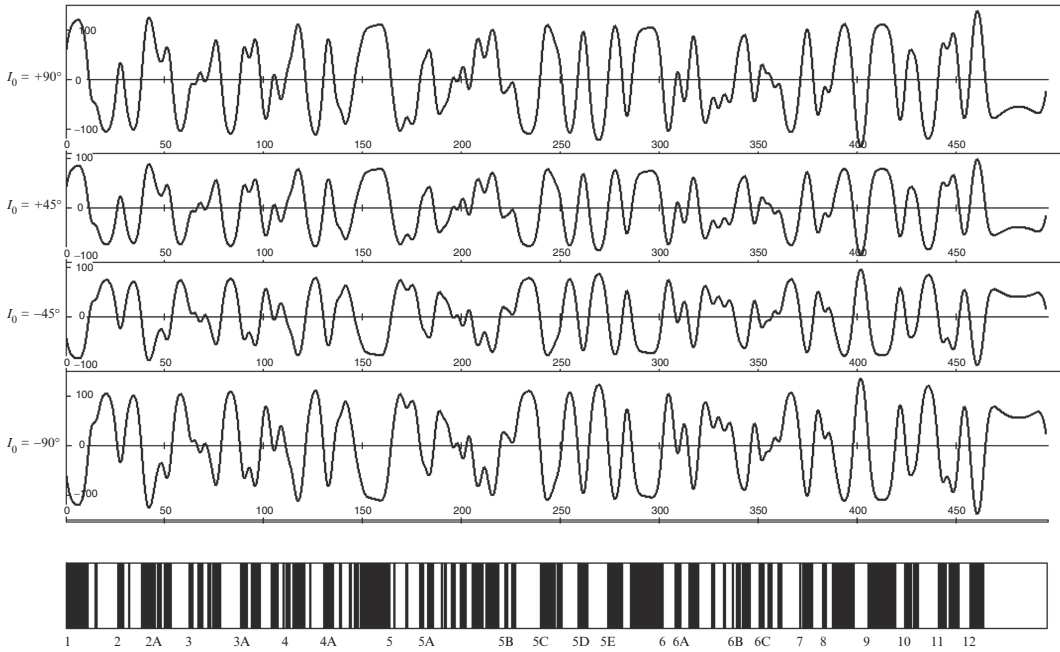
The expected magnetic anomaly profile is built calculating, for a series of equally spaced locations  $\mathbf{r}$  along the projection line, the contribution,  $\Delta \mathbf{F}_k(\mathbf{r})$ , to the total crustal field from each magnetized prism included in the model. The individual contribution of a block is determined through expressions (5.49) and (5.50). Then, a vector summation of these terms will give the total anomalous field associated with the magnetization model. Finally, the local anomaly is computed by projecting the total field  $\Delta \mathbf{F}(\mathbf{r})$  onto the reference field axis (Eq. 5.3). This procedure is repeated for each point  $\mathbf{r}$  along the projection line. The resulting magnetic anomaly profile can be compared with the observed data to evaluate if the assumed velocity is appropriate.

In general, the forward modelling of marine magnetic anomalies requires successive adjustments of the spreading velocity function  $v = v(T)$ , and eventually of the magnetization intensity  $M = M(T)$ , until an acceptable visual match between model and observed anomalies is obtained. In this trial and error procedure, the investigator first identifies the characteristic wiggles associated with the major anomalies (e.g., 2, 2A, 3, 4, 5, ...) on the observed profile, then he/she tries to change the spreading rate  $v$  of groups of chrons to improve the match. According to (5.60), a change of the spreading velocity  $v_k$  during the  $k$ -th chron determines a variation of the horizontal width  $w_k$  of the corresponding block, because the time interval  $\Delta T_k$  is fixed by the geomagnetic polarity time scale. A good rule of thumb is to match a well-known anomaly close to the profile end through an average constant velocity  $v$ . Then, we select an “easy” anomaly within the sequence and try to match the corresponding wiggle changing the average velocity of the lower half to some value  $v'$ . This operation will require an adjustment of the velocity of the upper half sequence to a new value

$v''$ . Such “divide-et-impera” algorithm can be repeated iteratively until we obtain a satisfactory fit of the major anomalies. In general, it is not recommended to use different velocities for the sub-chrons of a major chron, because the characteristic shape of the corresponding anomaly depends precisely upon the relative duration of the various sub-chrons. Therefore, changing arbitrarily the width of the blocks in the model could lead to a misinterpretation of the anomalies. It is important to note that the magnetization models resulting from procedures of forward or inverse modelling are never unique, because there are infinitely many block models that generate the same magnetic signal. When interpreting marine magnetic anomaly profiles, it is necessary to take into account that the shape of the major anomalies mainly depends from the following factors:

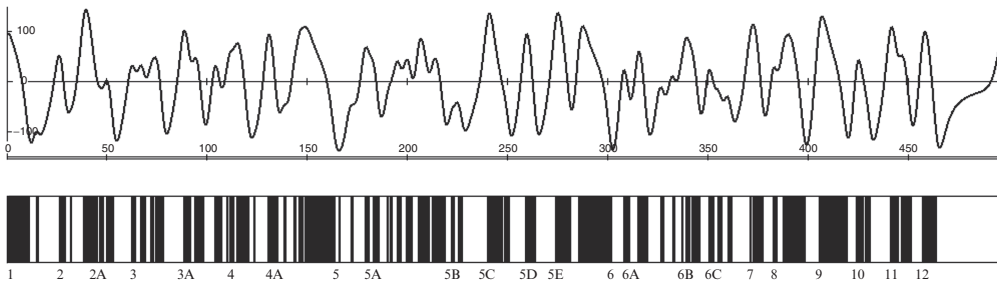
1. The bathymetric profile, which determines the geometry of magnetized prisms;
2. The presence of sea mounts and other volcanic features;
3. The present day latitude of the prisms,
4. The paleostrike,  $\alpha$ , of the magnetized blocks;
5. The paleolatitude of the blocks, which determines their inclination;
6. The present day strike,  $\beta$ , of the magnetized prisms;
7. The profile obliquity  $\psi$ ;
8. The presence of ridge jumps

For example, Fig. 5.12 illustrates the shape of the magnetic anomalies 1–12 along a N-S profile at various latitudes (hence, for different values of the reference field inclination  $I_0$ ). We note that the profiles that would be observed in the southern hemisphere are specular with respect to those observed in the northern hemisphere. The effect of another important factor influencing the shape of the magnetic anomalies, which is the paleo-strike of the spreading ridge, is illustrated in Fig. 5.13. While the inclination of the reference field essentially modifies the *amplitude* of the anomalies (within the same hemisphere), the paleo-strike,  $\alpha$ , of the magnetized blocks has a strong effect on the *shape*. Ridge jumps and strong spreading asymmetry are other sources of complexity in the analysis of marine



**Fig. 5.12** Theoretical magnetic anomaly profiles, generated at a ridge with constant full spreading rate  $v = 30 \text{ mm year}^{-1}$ , originally located at  $40^\circ\text{N}$  with a paleo-strike  $\alpha = 180^\circ$ . It is assumed that the present day strike of the prisms is  $\beta = 0^\circ$  (so that the ridge has not changed

orientation with respect to the geographic North). The magnetic signal changes according to the present day latitude of the prisms, as evidenced by the variation with the inclination  $I_0$  of the ambient geomagnetic field



**Fig. 5.13** Theoretical magnetic anomaly profile, generated at a ridge with constant full spreading rate  $v = 30 \text{ mm year}^{-1}$ , originally located at  $40^\circ\text{N}$  with a paleo-strike  $\alpha = 135^\circ$ . It is assumed that the present day strike of

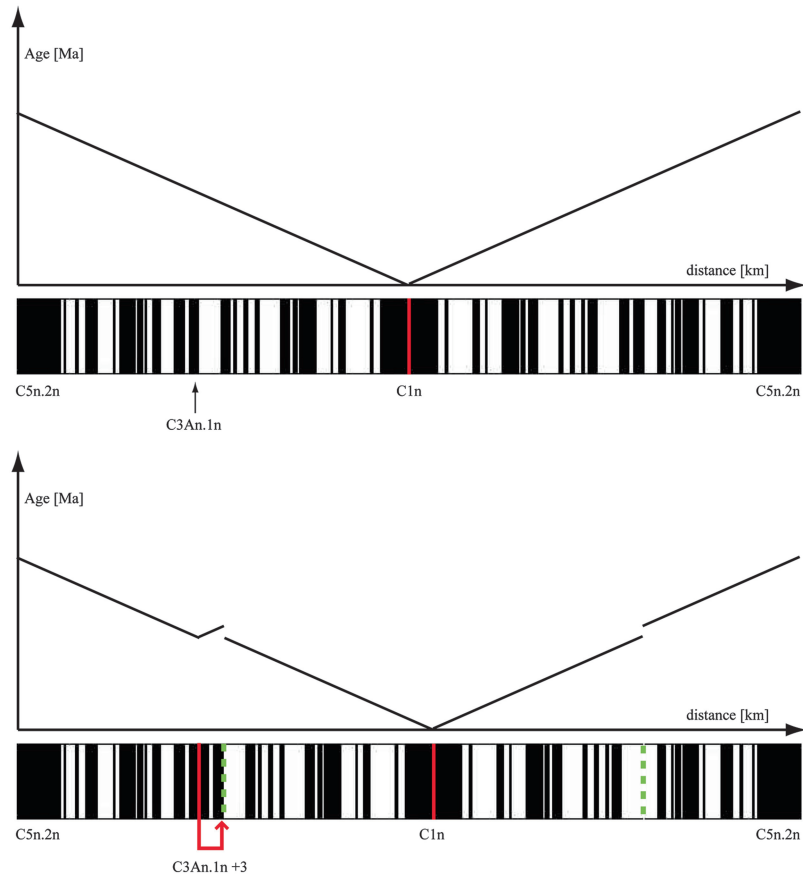
the prisms is  $\beta = 35^\circ$  (so that the ridge has not changed orientation with respect to the geographic North). The ambient geomagnetic field inclination is  $I_0 = 45^\circ$

magnetic anomalies. The effect of ridge jumps is to introduce gaps and duplications in the sequence of chrons. Figure 5.14 shows an example of magnetization model distorted by the presence of a ridge jump, along with the corresponding discontinuities in the age progression plot as a function of the distance from the ridge. The detection of minor ridge jumps in a real magnetic profile

is not a simple task, especially when the profile does not encompass both flanks of the spreading axis.

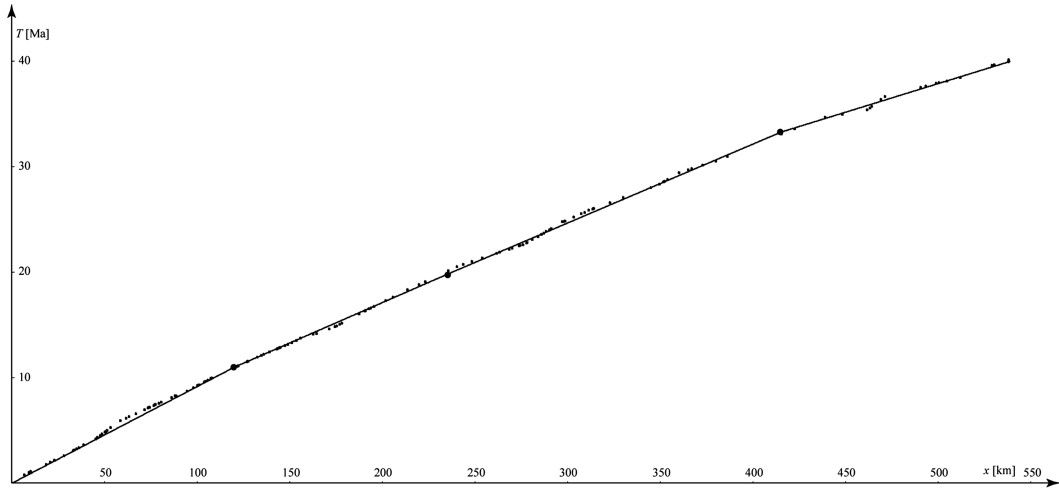
Therefore, it is good practice to be cautious when introducing a ridge jump in the block model: very often a reinterpretation of the anomalies allows to obtain anyway an acceptable fit of the data.

**Fig. 5.14** Discontinuities and duplications of a magnetization model caused by ridge jumps. In this example, an eastward jump by three blocks occurred at the end of chron C3An.1n. In general, eastward (or northward) ridge jumps add blocks to the western (southern) side and remove a corresponding number of prisms from the eastern (northern) side. The *green dashed lines* are discontinuities. The age-distance plots show the predicted crustal age discontinuities (After Schettino 2012)



The output of a forward modelling procedure is a spreading velocity function  $v = v(T)$ . This function can be used in turn to determine the offset of the anomalies along the profile, which is the first step in the construction of *isochron maps*. However, it is important to note that the best-fitting spreading velocities of a magnetization model do not necessarily represent true plate velocities during the corresponding chrons. For example, when a magnetic profile contains data from one side only, the resulting velocities may be anomalously high or low as a consequence of local spreading asymmetry or undetected ridge jumps. In Chap. 2, we have mentioned the fact that the spreading asymmetry may change not only between two neighbor ridge segments but also within the same segment (Sect. 2.4). Furthermore, a model where the block associated with chron  $C_k$  has apparent velocity  $v_k$  on the

eastern flank of a ridge cannot be distinguished from another representation in which the same block has been generated with velocity  $v_k/2$ , but has been subsequently doubled by a westward ridge jump. Therefore, in the analysis of magnetic anomaly profiles that encompass one side only of a spreading ridge, it is *not* possible to infer the real plate velocities from the apparent spreading rates associated with the block model. In general, the correct determination of the relative plate velocities requires a statistical treatment of many velocity models from the same area, and an analytical procedure of correlation of the magnetic profiles, in order to determine stage boundaries and average spreading rates. However, even the analysis of a single magnetic profile can furnish an rough estimate of the true spreading rate after an appropriate statistical treatment. In fact, let us consider the function  $x = x(T)$ , which gives



**Fig. 5.15** Age-Distance plot illustrating the relation between the offset,  $x$ , of each block included in a magnetization model and the corresponding chron upper boundary age  $T$  (small black dots). These points furnish the crustal

age as a function of the distance from the ridge axis. The black line shows the linear spline regression fit using a three-stages model. Large black dots are the knots of the regression curve

the distance  $x$  from a spreading centre segment, where we can find oceanic crust of age  $T$ :

$$x(T) = \frac{1}{2} \int_0^T v(t) dt \quad (5.63)$$

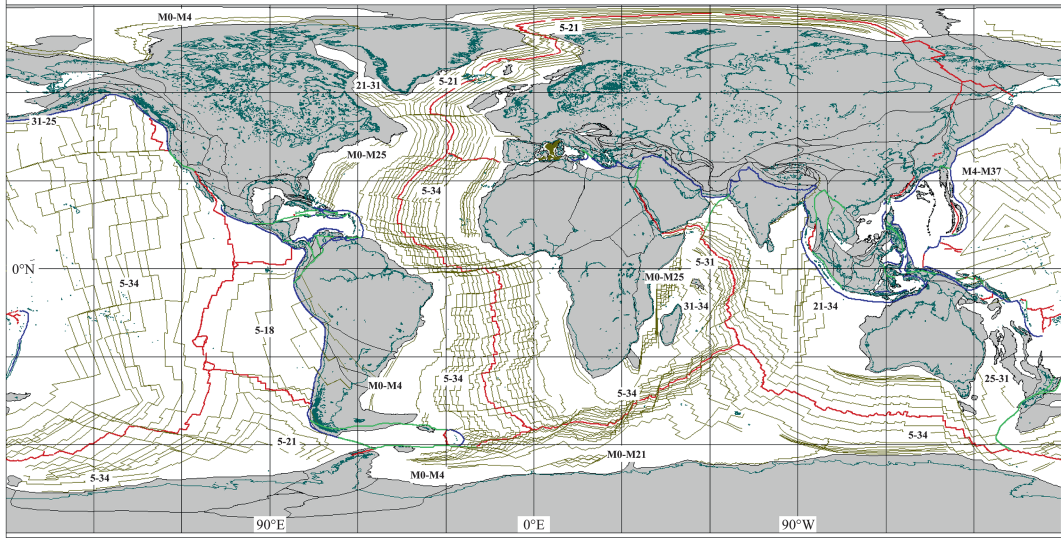
This age-distance function can be easily built on the basis of the velocity function  $v = v(T)$ . A key observation in plate kinematics is that the function (5.63) can be always approximated fairly well by a sequence of straight lines, that is by a first-order piecewise polynomial, in spite of the apparent full spreading rates variability through the geological time. As an example, Fig. 5.15 shows the age-distance plot associated with a magnetic profile in the central Atlantic. This plot suggests a change of spreading rate at anomaly 5, 6, and 13 times, associated with a change of slope of the regression lines.

Therefore, even the statistical analysis of a single magnetic profile can furnish an estimate of the true spreading rate over long time intervals. In fact, we have mentioned in Sect. 2.7 that a basic principle of plate kinematics establishes that the Euler vector describing the relative motion between two plates is approximately constant for long time intervals (of the order of tens Myrs) that

we called *stages*. This principle implies in turn that the spreading rate along a ridge segment is approximately constant during a stage, which is effectively what we observe through the analysis of marine magnetic anomalies. In general, a rough estimate of the true full spreading rates can be obtained from the apparent velocity function  $v = v(T)$  through a linear spline regression fit (Schettino 2012). A better estimate requires the more complex procedure that will be described in the next section.

## 5.6 Construction of Isochron Maps

Sea floor spreading *isochrons* are lines formed by a combination of points with the same age and fracture zone segments in an oceanic basin. They can be considered as determinations of the spreading ridge geometry in the geologic past. If the ocean is not yet subducting, we always have *two* conjugate isochrons for each possible time, placed on the opposite sides of the spreading center at a more or less equal distance. Although in principle we can build isochrons of any age less than the age of onset of sea floor spreading, it is common practice to construct isochron maps



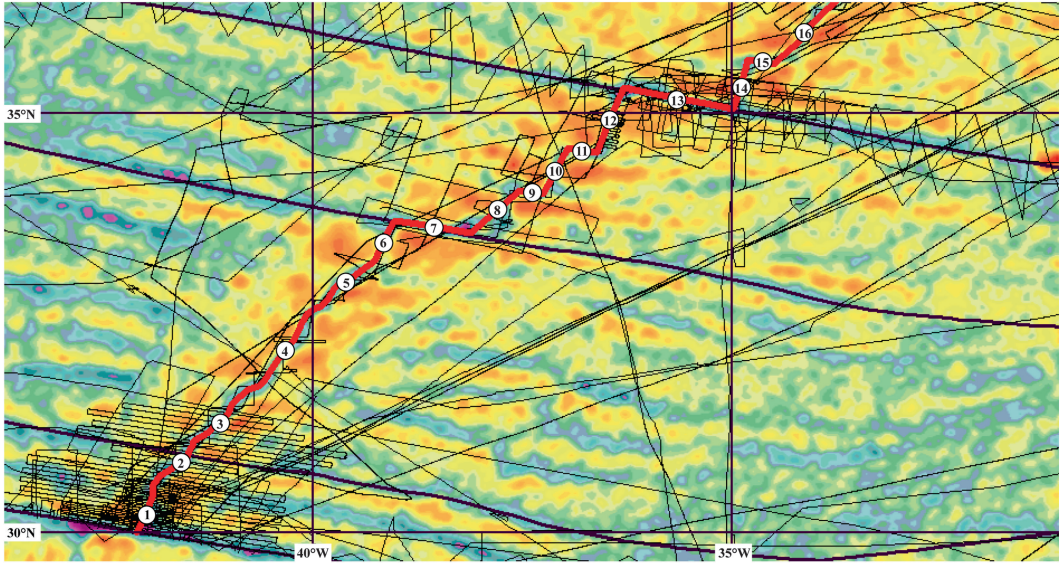
**Fig. 5.16** Global isochron chart of Royer et al. (1992). Labels are anomaly names of the corresponding isochrons. Modern plate boundaries and Mesozoic–Cenozoic tectonic elements are also displayed

that only include isochrons at stage boundaries. For instance, the first digital global compilation of isochrons from the World's oceans (Royer et al. 1992), which led to the well-known age map of the sea floor of Müller et al. (1997), included isochrons for only 14 anomalies: 5, 6, 13, 18, 21, 25, 31, 34, M0, M4, M10, M16, M21, and M25. In that model, the ages of these anomalies were identified as global *synchronous* stage boundaries (see Sect. 2.7), at which sharp changes of the stage poles occurred. We shall come back to this point in the next chapter. Figure 5.16 shows the original isochron chart of Royer et al. (1992), combined with the global compilation of tectonic elements of Schettino and Scotese (2005), some additional isochrons for the western Mediterranean (Schettino and Turco 2006), and a couple of synthetic (i.e., theoretical) isochrons for the Canada Basin area (based on the model of Rowley and Lottes 1988). This map illustrates the major tectonic features associated with the evolution of the oceanic basins since the middle Jurassic, including ridge jumps, ridge extinctions, changes of the stage pole, subduction of spreading centers, and the location where the Pacific plate formed as a small oceanic plate bounded by three ridges. However, the isochrons

also represent the geometrical expression of a statistical procedure that allows to determine finite reconstruction poles starting from locations of identified magnetic anomalies and fracture zones.

Now we are going to describe this procedure in detail. The starting point for the construction of isochrons is represented by a combination of ship track magnetic anomalies and fracture zones. To build a reliable map, it is necessary to have at least two magnetic profiles crossing each spreading ridge segment, and a digitized data set of fracture zones. An example of variable data coverage is illustrated in Fig. 5.17.

The second step consists into the analysis of the magnetic profiles, according to the procedure illustrated in the previous section. This step provides, for each profile, a sequence of apparent spreading velocities, which can be converted into a series of locations along the projection line. These locations correspond to the starting offset of each block in the magnetization model and are called *crossing points* or simply *crossings*. They specify where a certain anomaly can be found. Therefore, the set of all crossings corresponding to a given anomaly represents the basic



**Fig. 5.17** An example of data coverage for the construction of isochron maps. *Black lines* are ship tracks, *blue lines* are fracture zones. The spreading center is divided into 16 ridge segments. The coverage is good for segments

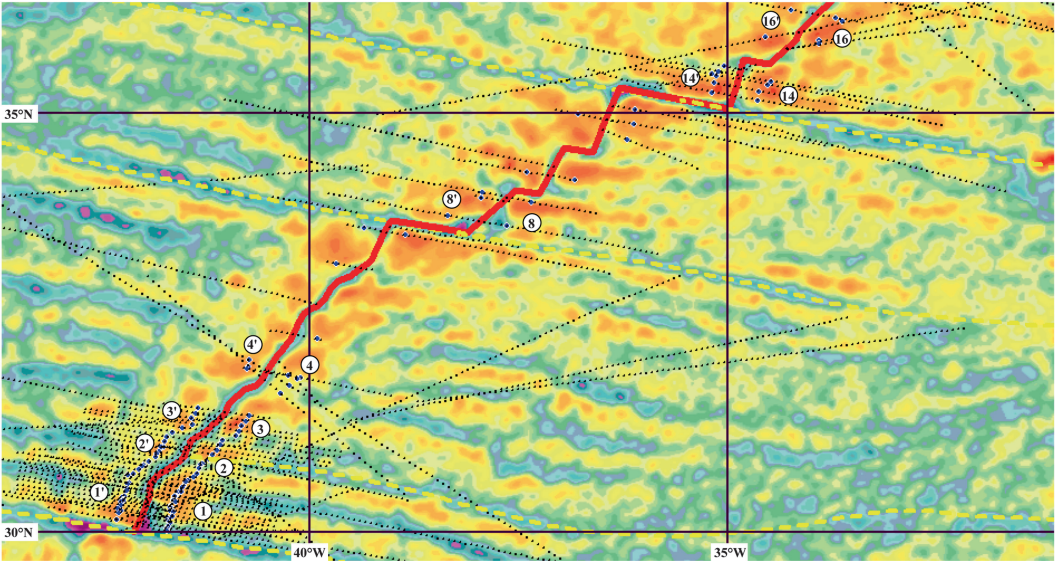
1–4, 8, 14, 16, and for fracture zone segments, while it is incomplete for segment 12 and scarce for segments 5, 6, and 10

building block for the construction of a sea floor spreading isochron. To this purpose, we usually group the crossings according to the side and to the ridge segment. Further subdivision is made when a ridge segment has variable strike. The reason is that we are going to search the finite reconstruction pole and rotation angle that best fits the set of all *conjugate groups of crossing and fracture zone lineations*, namely, the set of all groups of crossings or fracture zone points that are placed on the opposite flanks of a ridge and are approximately aligned along great circle arcs. As an example, Fig. 5.18 shows the conjugate groups of crossings for anomaly 2A relative to the data set of Fig. 5.17.

Therefore, at the next step we shall perform a statistical fitting of the magnetic lineations and the intervening fracture zone segments. This procedure will be described in the next section. It furnishes the Euler pole of closure of the ocean floor younger than the selected anomaly. Then, we proceed creating two additional sets of crossings by application of the resulting finite reconstruction matrix (see Sect. 2.7) to the eastern groups, and the conjugate matrix to the

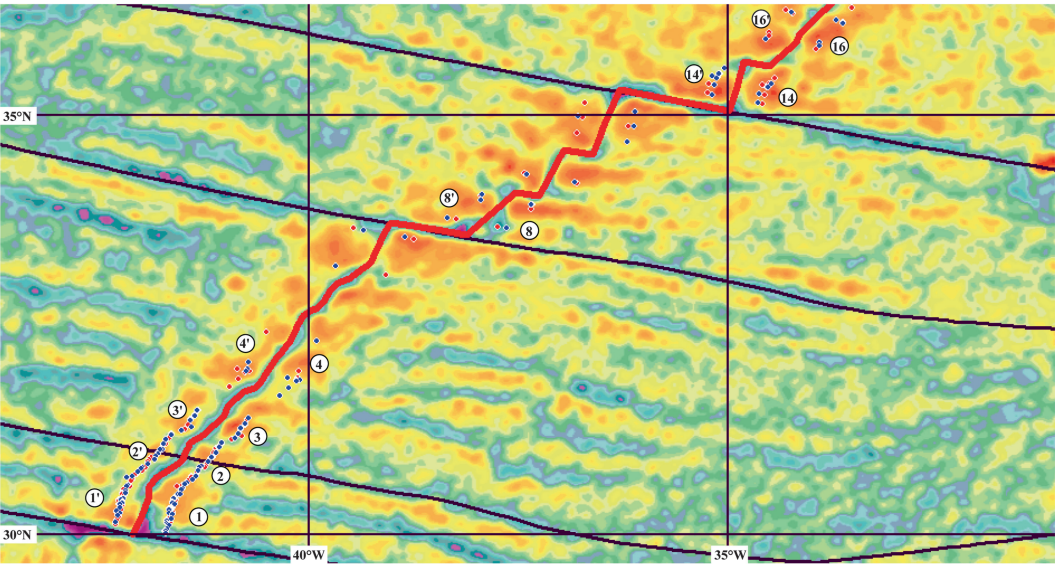
western groups. The result is a more dense set of crossing points that delineates better the magnetic lineations corresponding to the selected anomaly. Furthermore, this technique allows to perform a visual inspection of the goodness of fit. An example of this approach is illustrated in Fig. 5.19. The last step requires tracing the representative magnetic lineations on one of the two flanks (the choice is generally arbitrary). These lineations are prosecuted to intersect the fracture zones, where they are linked together by intervening segments of the digitized fracture zones. The result is a digital isochron. To obtain the conjugate isochron, we do not repeat the procedure on the opposite flank. Instead, we simply rotate the isochron using the reconstruction matrix determined during the previous step. This technique ensures that the two isochrons will match perfectly when making plate reconstructions.

The final result of the procedure illustrated above is represented by a pair of conjugate isochrons like those shown in Fig. 5.20. Repeating the steps for older anomalies gives the complete set of isochrons describing the evolution of an oceanic basin.

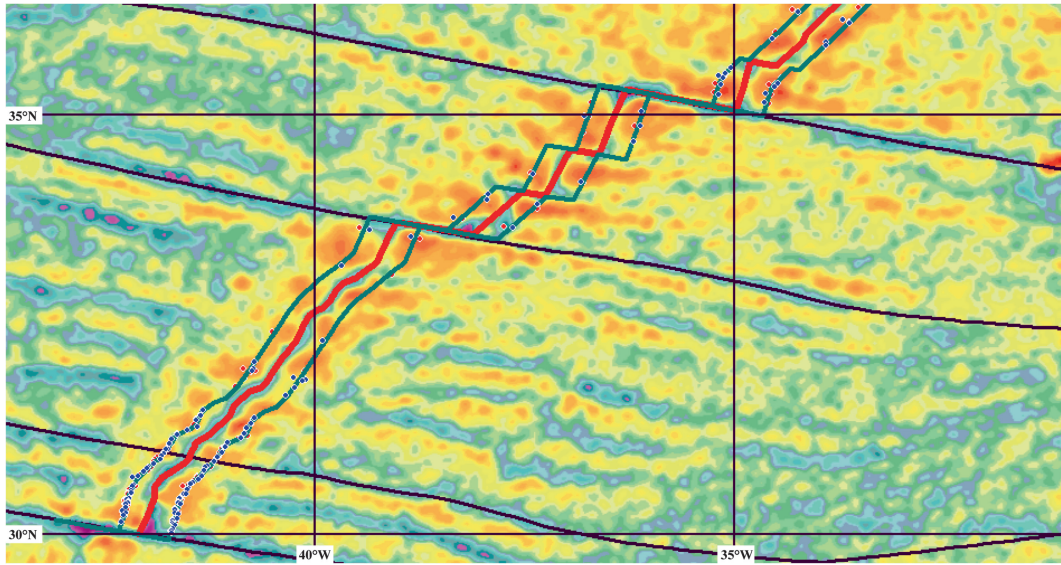


**Fig. 5.18** Crossing points for anomaly 2A (blue dots) in the example of Fig. 5.17. Projection lines, where the magnetic data of Fig. 5.17 have been projected, are shown as *dotted lines*. There are seven conjugate groups of crossings and five conjugate groups of fracture zone points that can be used to determine the finite reconstruction

pole and rotation angle at anomaly 2A time (2.58 Ma). Note that segments with only one crossing on one side cannot be used in the fitting procedure. Therefore, the resulting isochron geometry along these segments will be unconstrained



**Fig. 5.19** Complementary data set of crossing points (red dots) for the example of Fig. 5.18



**Fig. 5.20** Sea floor spreading isochrons 2A (green lines) generated by interpolation of complemented crossings for the example of Fig. 5.19

## 5.7 Determining Finite Rotations

Now we are going to describe the procedure for determining the best fit Euler rotation that matches two conjugate sets of crossings and fracture zone points, along with the associated uncertainty parameters. Both the crossings obtained through the analysis of magnetic anomalies and the points that can be sampled along a fracture zone are affected by errors. Apart from the case of mis-interpretation of magnetic anomalies, we have errors associated with navigation (up to 10–15 km), errors associated with the mapping of fracture zone (5–20 km), and an uncertainty relative to the sampling of fracture zone points within the zone of gravity anomaly low that characterizes these features (up to 30 km). For an in-depth discussion of these errors the reader is referred to the paper of Kirkwood et al. (1999). The uncertainty in position for crossings and fracture zone points determines in turn an uncertainty in the rotation parameters, which depends from the length of the two isochrons and their distance from the best fit Euler pole (Stock and Molnar 1983).

The most widely used algorithm for determining the best Euler rotation that matches two data

sets of points sampled along conjugate isochrons was proposed by Hellinger (1981). It can be described as follows. Let us assume that the conjugate isochrons can be divided into  $N$  lines that are representative of past spreading ridge segments and transform faults, as illustrated in Fig. 5.17. We also assume that these lines have roughly the geometry of great circle arcs. If a spreading ridge segment cannot be approximated by a great circle arc, it is subdivided in smaller segments that satisfy this assumption. Let  $\mathbf{p}_{ij}$  ( $i = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, n_i$ ) be the position vector of a data point on the  $i$ -th segment of one isochron. Similarly, let  $\mathbf{q}_{ik}$  ( $i = 1, 2, \dots, N$ ;  $k = 1, 2, \dots, m_i$ ) be the position vector of a data point on the  $i$ -th segment of the conjugate isochron. If  $\mathbf{R}$  is a test rotation matrix, close to the best fit rotation, then the points  $\mathbf{R}\mathbf{q}_{ik}$  should approximately match the corresponding points  $\mathbf{p}_{ij}$ . In this instance, both the  $\mathbf{R}\mathbf{q}_{ik}$  and the  $\mathbf{p}_{ij}$  should be aligned about a great circle arc, with a confidence interval not exceeding the size of the estimated errors of the data. To calculate the misfit, let us consider the unit vector  $\mathbf{w}_i$  normal to the  $i$ -th great circle arc. The scalar product  $\mathbf{w}_i \cdot \mathbf{p}_{ij}$  represents the angular distance between  $\mathbf{w}_i$  and  $\mathbf{p}_{ij}$ . Therefore, it determines the distance  $c_{ij}$  of  $\mathbf{p}_{ij}$  from the  $i$ -th great

circle arc. When this distance is small we can write:

$$\begin{aligned} c_{ij} &= \frac{\pi}{2} - \cos^{-1}(\mathbf{w}_i \cdot \mathbf{p}_{ij}) \\ &= \sin^{-1}(\mathbf{w}_i \cdot \mathbf{p}_{ij}) \cong \mathbf{w}_i \cdot \mathbf{p}_{ij} \end{aligned} \quad (5.64)$$

Similarly, the scalar product between  $\mathbf{R}\mathbf{q}_{ik}$  and  $\mathbf{w}$  will give the approximate distance  $d_{ij}$  of  $\mathbf{R}\mathbf{q}_{ik}$  from the *same* great circle arc. Now let us assume that these points have respectively standard deviations  $\sigma_{ij}$  and  $\tilde{\sigma}_{ik}$ . Then, the maximum-likelihood estimator of the misfit can be written as follows:

$$\chi^2 = \sum_{i=1}^N \left[ \sum_{j=1}^{n_i} \frac{(\mathbf{w}_i \cdot \mathbf{p}_{ij})^2}{\sigma_{ij}^2} + \sum_{k=1}^{m_i} \frac{(\mathbf{w}_i \cdot \mathbf{R}\mathbf{q}_{ik})^2}{\tilde{\sigma}_{ik}^2} \right] \quad (5.65)$$

Keeping fixed the Euler pole associated with  $\mathbf{R}$ , the angle of rotation that minimizes  $\chi^2$  is found iteratively searching in a neighborhood of the rotation angle  $\Omega$  associated with  $\mathbf{R}$ . From (2.18) we see that this angle can be calculated easily from the trace of  $\mathbf{R}$ :

$$\Omega = \cos^{-1} \left( \frac{\text{Tr}(\mathbf{R}) - 1}{2} \right) \quad (5.66)$$

At the next step, Hellinger's algorithm requires calculation of  $\chi^2$  and the corresponding best fit rotation angle for eight Euler poles that lie on the border of a spherical rectangle having edges of  $0.5^\circ$  and centered at the initial pole. The Euler pole that gives the minimum misfit is selected as the new starting point, if  $\chi^2$  is less than the misfit of the initial pole. Otherwise, the size of the rectangle is halved. The previous steps are repeated until the size of the rectangle drops below the acceptable precision. An improved version of this algorithm can be built considering that a fracture zone segment is better approximated by a small circle arc about the current Euler pole (Matias et al. 2005). Let  $\mathbf{e}$  be the Euler pole associated with  $\mathbf{R}$ . If the  $i$ -th set of points is a fracture zone segment, then the average angular distance from  $\mathbf{e}$  is:

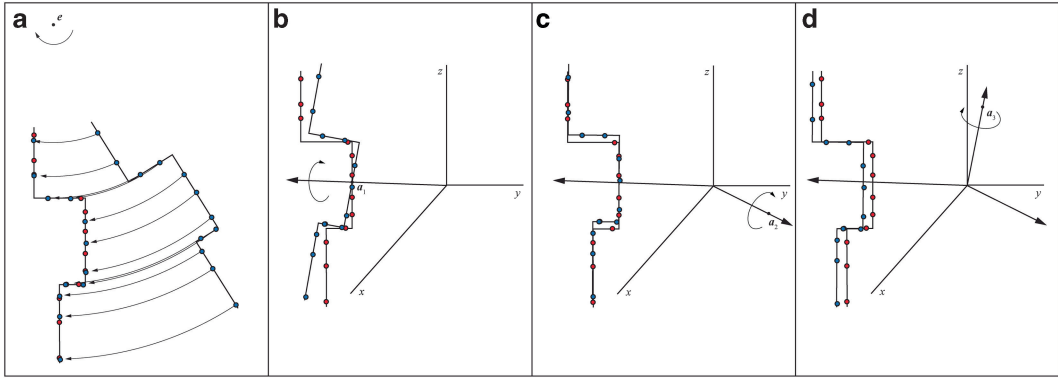
$$\begin{aligned} \langle \theta_i \rangle &= \frac{1}{n_i} \sum_{j=1}^{n_i} \cos^{-1}(\mathbf{p}_{ij} \cdot \mathbf{e}) \\ &+ \frac{1}{m_i} \sum_{k=1}^{m_i} \cos^{-1}(\mathbf{R}\mathbf{q}_{ik} \cdot \mathbf{e}) \end{aligned} \quad (5.67)$$

Therefore, taking the small circle arc with distance  $\langle \theta_i \rangle$  as the estimated best fit fracture zone segment, then the misfit is expressed by the following functional:

$$\begin{aligned} \chi_i^2 &= \frac{1}{n_i} \sum_{j=1}^{n_i} [\langle \theta_i \rangle - \cos^{-1}(\mathbf{p}_{ij} \cdot \mathbf{e})]^2 \\ &+ \frac{1}{m_i} \sum_{k=1}^{m_i} [\langle \theta_i \rangle - \cos^{-1}(\mathbf{R}\mathbf{q}_{ik} \cdot \mathbf{e})]^2 \end{aligned} \quad (5.68)$$

In the approach of Hellinger (1981), any rotation matrix  $\mathbf{R}$  (or Euler pole  $\mathbf{e}$  and rotation angle  $\Omega$ ) such that the total misfit does not exceed the average variance of the data is acceptable. The set of all Euler poles and rotation angles that are acceptable on the basis of this criterion furnish the confidence limits of the reconstruction. A more rigorous study of the statistical properties of the Hellinger solution has been performed by T. Chang and described in a series of papers (e.g., Chang 1993). Here we are going to illustrate the more intuitive, although heuristic, approach of Stock and Molnar (1983). These authors considered the problem of determining how the distribution of the data influences the confidence limits of the Euler pole location and the rotation angle. To this purpose, they perturbed the best fitting rotated data set  $\{\mathbf{R}\mathbf{q}_{ik}\}$  through small additional rotations that were called *partial uncertainty rotations* (PURs). These PURs correspond to three standard ways to distort a best fitting configuration introducing a known amount of misfit (Fig. 5.21).

The PURs can be combined to estimate the uncertainty associated with the best fit Euler rotation. Let  $\varepsilon$  and  $L$  be respectively the maximum allowed angular misfit of the data points along the



**Fig. 5.21** Partial uncertainty rotations (b–d) about an ideal best fit reconstruction (a). In (b) a skewed fit is generated by a small rotation about the isochron “center”  $a_1$ . Another way to introduce a distortion is through a small rotation about the versor  $a_2$ , which is orthogonal to  $a_1$  (c). This rotation generates a mismatch of the fracture

zone segments, while leaving the crossings on the correct great circles. Finally, we can distort the fit by a small rotation about axis  $a_3$  (d), which is orthogonal to both  $a_1$  and  $a_2$ . In this case, a mismatch of the crossings is generated while leaving the fracture zone points on the correct segments

reconstructed pair of isochrons and the total angular length of the spreading segments. It is easy to prove that the length  $L$  determines the rotation angle,  $\delta\Omega_1$ , for the skewed fit PUR. In fact, let us consider a spherical triangle with two sides of length  $L/2$  and one side of length  $\varepsilon$ . Applying the spherical cosine law to this triangle, we have that  $\cos \varepsilon = \cos^2(L/2) + \sin^2(L/2)\cos(\delta\Omega_1)$ .

Therefore, in order to have a maximum mismatch  $\varepsilon$  at distance  $L/2$  from the isochron center, the angle  $\delta\Omega_1$  must be given by:

$$\delta\Omega_1 = \cos^{-1} \left( \frac{\cos \varepsilon - \cos^2 L/2}{\sin^2 L/2} \right) \quad (5.69)$$

In the case of PURs associated with mismatched fracture zones or crossings, the distance of the isochron center from the corresponding pole is  $90^\circ$ . Therefore, the rotation angles  $\delta\Omega_2$  and  $\delta\Omega_3$  that introduce a maximum mismatch  $\varepsilon$  are:

$$\delta\Omega_2 = \delta\Omega_3 = \varepsilon \quad (5.70)$$

We see that the PUR angle  $\delta\Omega_1$  not only depends from the maximum allowed mismatch  $\varepsilon$ , but it also depends from the isochron size  $L$ . For each of the three PURs, the rotation angle  $\delta\Omega_k$  can be either positive or negative. Therefore, we have six end-member rotations that define a

roughly ellipsoidal uncertainty region about the best fit triple  $(\lambda_e, \phi_e, \Omega)$ ,  $(\lambda_e, \phi_e)$  being the Euler pole coordinates and  $\Omega$  being the best fit rotation angle. The end-member rotations associated with skewed fits and mismatched fracture zones will determine the uncertainty of the Euler pole location  $(\lambda_e, \phi_e)$ , while the end-member rotations associated with mismatched crossings will determine the uncertainty in the rotation angle  $\Omega$ . In summary, we can use Hellinger’s algorithm to determine the best fit rotation that matches two conjugate data set of crossings and fracture zone points. Then, assuming an a priori uncertainty of the data, we can calculate a confidence region for the best fit rotation  $\mathbf{R}$  multiplying this matrix by the six end-member rotations associated with the PURs.

## 5.8 Data Transformations

One of the most useful tools for the analysis of potential field data is the *Fourier transform*. In the context of potential field geophysics, this tool is used to map real functions of spatial variables  $(x, y, z)$  into complex functions of a wave vector  $\mathbf{k}$ . The reason to perform this transformation is that many complex operations, such as the so-called upward continuation, are simple linear transformations in the space of wave vectors. Let  $f = f(x)$

be a real continuous non-periodic function of variable  $x$ , such that  $f(x) \rightarrow 0$  as  $x \rightarrow \pm\infty$ .

The Fourier transform of  $f$  is a complex function of a real parameter  $k$ , defined as follows:

$$F(k) = \int_{-\infty}^{+\infty} f(x) e^{-ikx} dx \quad (5.71)$$

The variable  $k$  in (5.71) is called *wavenumber* and has units of  $\text{m}^{-1}$ . It is related to the *wavelength*  $\lambda$  by the following relation (see Sect. 4.8):

$$k = \frac{2\pi}{\lambda} \quad (5.72)$$

Being a complex function, on the basis of Euler's formula  $F$  can be written in the form:

$$F(k) = |F(k)| e^{i\Theta(k)} \quad (5.73)$$

where the functions  $|F(k)|$  and  $\Theta(k)$  are called respectively *amplitude* and *phase*. The key feature of Fourier transforms is that always exists an *inverse Fourier transform* that allows to go back from the Fourier wavenumber domain to the space domain:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(k) e^{ikx} dk \quad (5.74)$$

Therefore, the basic idea in Fourier-domain modelling is to simplify complex operations by application of the Fourier transform, then going back to the space domain through an inverse transformation. In the case of multivariate functions, the Fourier transform defines a complex function of a wavevector  $\mathbf{k}$ , thereby, the transformation and its inverse assume the form:

$$F(\mathbf{k}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}} dx dy dz \quad (5.75)$$

$$f(\mathbf{r}) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} dk_x dk_y dk_z \quad (5.76)$$

The Fourier transformation  $\mathcal{F}[f]$  of a real function has the basic properties listed below:

1. *Symmetry*.  $\text{Re}(F)$  is a symmetric function,  $\text{Im}(F)$  is antisymmetric:

$$\begin{aligned} \text{Re}(F(-k)) &= \text{Re}(F(k)); \text{Im}(F(-k)) \\ &= -\text{Im}(F(k)) \end{aligned} \quad (5.77)$$

2. *Linearity*. For any pair of constants  $a$  and  $b$ , if  $f$  and  $g$  are real continuous non-periodic functions of  $x$ , then:

$$\mathcal{F}[af + bg] = a\mathcal{F}[f] + b\mathcal{F}[g] \quad (5.78)$$

3. *Scaling*. For an arbitrary constant  $a \neq 0$ , if  $g(x) = f(ax)$  and  $G = \mathcal{F}[g]$ , then:

$$G(k) = \frac{1}{|a|} F\left(\frac{k}{a}\right) \quad (5.79)$$

4. *Shifting*. If  $g(x) = f(x - x_0)$  is a shift of function  $f$ , then its transform adds a linear phase factor to the Fourier transform of  $f$ , leaving the amplitude spectrum unaffected:

$$G(k) = F(k) e^{-ikx_0} \quad (5.80)$$

5. *Convolution*. Let  $f$  and  $g$  two real functions with Fourier transforms  $F$  and  $G$ , respectively. The following integral function is called *convolution* of  $f$  and  $g$ :

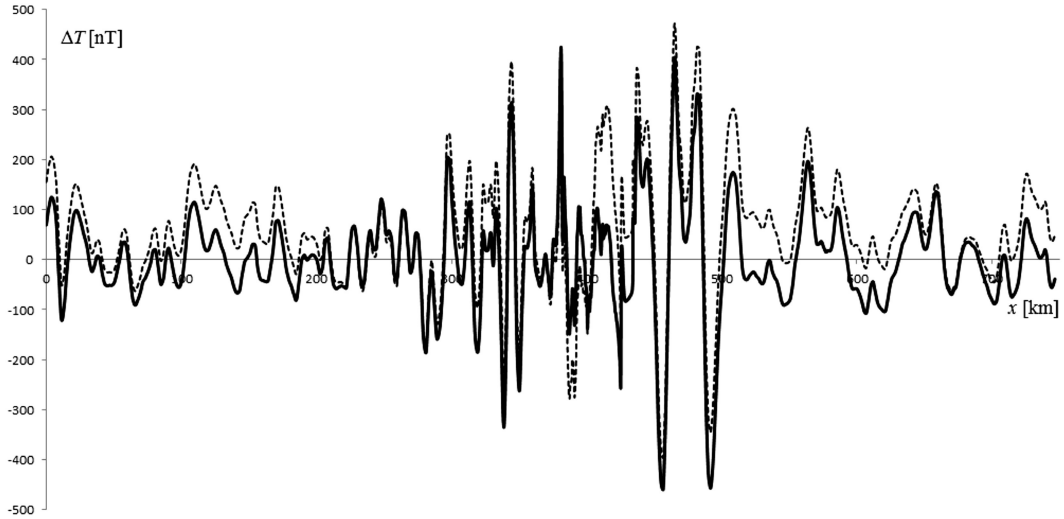
$$h(x) = \int_{-\infty}^{+\infty} f(\xi) g(x - \xi) d\xi \equiv f * g \quad (5.81)$$

Then, the convolution property states that:

$$H(k) = F(k)G(k) \quad (5.82)$$

6. *Derivative*. The operation of differentiation in the space domain is transformed into a multiplication in the wavenumber domain:

$$\mathcal{F}\left[\frac{d^n f}{dx^n}\right] = (ik)^n F(k) \quad (5.83)$$



**Fig. 5.22** Fourier filtering of magnetic anomaly data in the phase preceding forward modelling. In this example, wavenumbers below  $0.01 \text{ km}^{-1}$  have been filtered away.

The original signal (*dashed line*) had an almost constant offset  $\Delta T_0 \cong 58.3 \text{ nT}$ . The filtered magnetic anomalies have zero average and are more appropriate for modelling

7. *Parseval Identity*. The “energy” of a real function is invariant under Fourier transformation:

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |F(k)|^2 dk \quad (5.84)$$

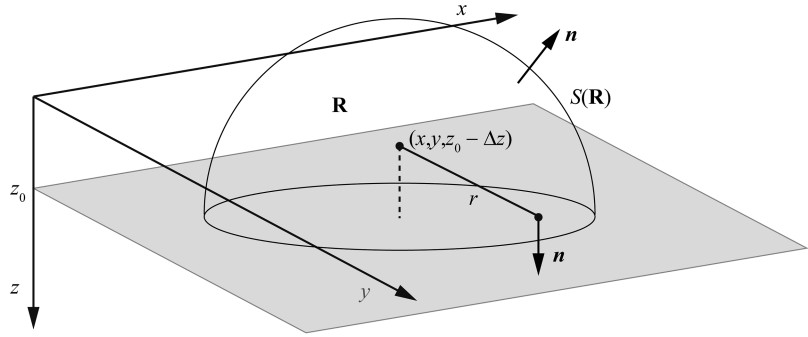
One of the simplest applications of Fourier’s transform is the filtering of magnetic anomaly data before initiating the forward modelling step. Figure 5.22 illustrates an example where the original data had a common offset of a few tens nT. The signal was first transformed to the Fourier domain, then the low wavenumbers (long wavelengths) that were responsible for the signal displacement were removed from the spectrum. Finally, the filtered Fourier domain representation was converted back to the space domain through an inverse transform. The resulting signal is clearly more suitable for the subsequent interpretation phase. Soon we will introduce a more complex application of the Fourier transform.

Let us consider now the problem of transforming the potential field data observed at some altitude to a different surface, more distant

from the magnetic sources. This transformation is called *upward continuation* and is useful when aeromagnetic data observed at different altitudes must be merged, or when an investigator wants to attenuate the short-wavelength components of the signal and enhance the complementary range, which is a form of data filtering. For example, one could wish to enhance the anomalies associated with deep sources, which have longer wavelengths, with respect to the short-wavelength anomalies generated by near-surface sources. The theoretical basis for upward continuation is the Green’s third identity (4.70). Let us assume to know the potential  $V$  everywhere on the plane  $z = z_0$ . We want to calculate the potential at some point with greater elevation  $(x, y, z_0 - \Delta z)$ . To this purpose, let us choose a harmonicity region  $\mathbf{R}$  as in Fig. 5.23, and assume that all magnetic sources are located at altitude  $z > z_0$ .

In this case, we can separate the integral (4.70) in two parts, one that performs integration over the hemisphere surface, and one that operates on a circle in the plane  $z = z_0$  (Fig. 5.23). Therefore,

**Fig. 5.23** Geometry for the upward continuation transformation



if we let the hemisphere radius to increase indefinitely, the integral over the hemisphere surface goes to zero, so that the Green third identity assumes the form:

$$\begin{aligned} V(x, y, z_0 - \Delta z) &= \frac{1}{4\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left[ \frac{1}{r} \frac{\partial V}{\partial z'} \Big|_{z_0} - V(x', y', z_0) \right. \\ &\quad \left. \times \frac{\partial}{\partial z'} \left( \frac{1}{r} \right) \Big|_{z_0} \right] dx' dy' \end{aligned} \quad (5.85)$$

where  $r = [(x-x')^2 + (y-y')^2 + (z_0 - \Delta z - z')^2]^{1/2}$ . This expression cannot be easily evaluated, because the vertical derivative of  $V$  is not generally available.

To understand how we can eliminate  $\partial V / \partial z'$  in (5.85), let us consider a new harmonic function  $U$  in  $\mathbf{R}$ . If we apply the second Green's identity (4.67) to the pair  $(U, V)$  and insert the result in the representation formula (4.70), we obtain:

$$\begin{aligned} V(P) &= \frac{1}{4\pi} \oint_{S(\mathbf{R})} \left[ \left( U + \frac{1}{r} \right) \frac{\partial V}{\partial n} - V \frac{\partial}{\partial n} \left( U + \frac{1}{r} \right) \right] dS \end{aligned} \quad (5.86)$$

To eliminate the first part of the integrand, we must find a harmonic function  $U$  such that  $U + 1/r = 0$  on  $S(\mathbf{R})$ . It is easy to verify that the function:

$$\begin{aligned} U(x', y', z') &= - \left[ (x-x')^2 + (y-y')^2 + (z_0 + \Delta z - z')^2 \right]^{-1/2} \\ &\equiv - \frac{1}{\rho(x', y', z')} \end{aligned} \quad (5.87)$$

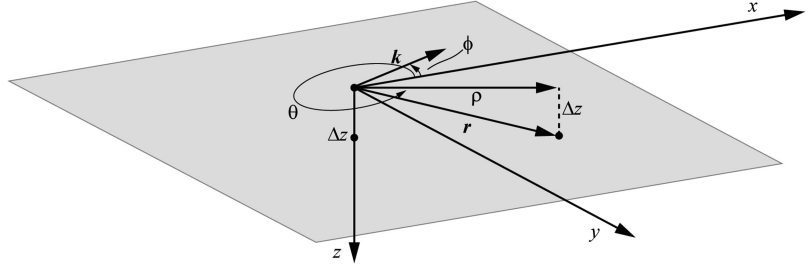
is harmonic in  $\mathbf{R}$  and satisfies the required condition  $U + 1/r = 0$  on the plane  $z' = z_0$ . Furthermore,  $U(x', y', z') \rightarrow 0$  on the hemisphere surface as its radius increases. Therefore, substituting (5.86) and (5.87) in (5.85) gives:

$$\begin{aligned} V(x, y, z_0 - \Delta z) &= - \frac{1}{4\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} V(x', y', z_0) \frac{\partial}{\partial z'} \left( \frac{1}{r} - \frac{1}{\rho} \right) \Big|_{z_0} dx' dy' \\ &= \frac{\Delta z}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{V(x', y', z_0)}{[(x-x')^2 + (y-y')^2 + \Delta z^2]^{3/2}} dx' dy' \end{aligned} \quad (5.88)$$

This is the upward continuation integral. It allows to calculate the potential at any point above

a plane, given the values on that level surface. Clearly, in practice the numerical evaluation of

**Fig. 5.24** Polar coordinates for the computation of integral (5.94)



the upward continuation integral is performed on a finite but sufficiently large rectangle below the point  $(x, y, z_0 - \Delta z)$ . An alternative approach to the upward continuation is based on the Fourier domain representation, which allows to determine  $V$  on any plane, given the values of the function on a reference level surface. We note that the integral (5.88) is a form of two-dimensional convolution of  $V$  by a harmonic function  $\psi$ :

$$V(x, y, z_0 - \Delta z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} V(x', y', z_0) \times \psi(x - x', y - y', \Delta z) dx' dy' \quad (5.89)$$

where

$$\psi(x - x', y - y', \Delta z) = \frac{\Delta z}{2\pi [(x - x')^2 + (y - y')^2 + \Delta z^2]^{3/2}} \quad (5.90)$$

Therefore, we can perform a Fourier transformation of both sides of (5.89) and apply the convolution property. If  $\tilde{V}$  is the upward continuation of  $V$ , then:

$$\mathcal{F}[\tilde{V}] = \mathcal{F}[V] \mathcal{F}[\psi] \quad (5.91)$$

To determine the Fourier transform of  $\psi$ , it is useful to rewrite (5.90) as follows:

$$\psi(x - x', y - y', \Delta z) = -\frac{1}{2\pi} \frac{\partial}{\partial \Delta z} \frac{1}{r} \quad (5.92)$$

where  $r$  is the distance of the point  $(x, y, z_0 - \Delta z)$  from points on the plane  $z' = z_0$  (Fig. 5.23). The Fourier transform of (5.92) gives:

$$\mathcal{F}[\psi] = -\frac{1}{2\pi} \frac{\partial}{\partial \Delta z} \mathcal{F}\left[\frac{1}{r}\right] \quad (5.93)$$

To calculate the Fourier transform of  $1/r$ , let us move the observation point to  $(0, 0, 0)$ . In this instance, the Fourier transform is performed over the plane  $z = \Delta z$ , with  $r = [x^2 + y^2 + \Delta z^2]^{1/2}$ . Then,

$$\mathcal{F}\left[\frac{1}{r}\right] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{x^2 + y^2 + \Delta z^2}} e^{-ik \cdot \mathbf{p}} dx dy \quad (5.94)$$

This integral can be evaluated more easily passing to polar coordinates  $(\rho, \theta, \phi)$ , as shown in Fig. 5.24.

We have:

$$\mathcal{F}\left[\frac{1}{r}\right] = \int_0^{+\infty} \int_0^{2\pi} \frac{\rho}{\sqrt{\rho^2 + \Delta z^2}} e^{-ik\rho \cos(\theta - \phi)} d\rho d\theta \quad (5.95)$$

Making the substitution  $\theta' = \theta - \phi$  and taking into account that  $\exp(-ik\rho \cos\theta')$  is periodic with period  $\pi$ , we obtain:

---


$$\mathcal{F}\left[\frac{1}{r}\right] = \int_0^{+\infty} \frac{\rho}{\sqrt{\rho^2 + \Delta z^2}} \left[ \int_{-\phi}^{2\pi-\phi} e^{-ik\rho \cos \theta'} d\theta' \right] d\rho = \int_0^{+\infty} \frac{\rho}{\sqrt{\rho^2 + \Delta z^2}} \left[ \int_0^{2\pi} e^{-ik\rho \cos \theta} d\theta \right] d\rho \quad (5.96)$$


---

The integral over  $\theta$  is a Bessel function of the first kind:

$$J_0(x) = J_0(-x) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{-ix \sin \alpha} d\alpha$$

$$= \frac{1}{2\pi} \int_0^{2\pi} e^{-ix \cos \theta} d\theta$$

Therefore,

$$\mathcal{F}\left[\frac{1}{r}\right] = 2\pi \int_0^{+\infty} \frac{\rho J_0(k\rho)}{\sqrt{\rho^2 + \Delta z^2}} d\rho$$

$$\equiv 2\pi \int_0^{+\infty} f(\rho) J_0(k\rho) \rho d\rho = 2\pi F_0(k) \quad (5.97)$$

where  $F_0(k)$  is a Hankel transform of order zero. For  $k$  and  $a$  positive real numbers, this transformation gives (e.g., Poularikas 2010):

$$\mathcal{H}_0 \left[ \frac{1}{(\rho^2 + a^2)^{1/2}} \right] = \frac{e^{-ak}}{k}$$

Therefore, using this result in (5.97) we obtain:

$$\mathcal{F}\left[\frac{1}{r}\right] = 2\pi \frac{e^{-k\Delta z}}{k} \quad (5.98)$$

Finally, substitution into (5.93) gives:

$$\mathcal{F}[\psi] = -\frac{1}{2\pi} \frac{\partial}{\partial \Delta z} \left( 2\pi \frac{e^{-k\Delta z}}{k} \right) = e^{-k\Delta z} \quad (5.99)$$

where  $k > 0$  and  $\Delta z > 0$ . Therefore, the Fourier transform of the upward continuation

of  $V$  is simply obtained from the Fourier transform of  $V$  by multiplying this function by  $\exp(-k\Delta z)$ :

$$\mathcal{F}[\tilde{V}] = \mathcal{F}[V] e^{-k\Delta z} \quad (5.100)$$

This solution shows that all wavenumbers are attenuated in the upward continuation, and that the degree of attenuation increases with the step  $\Delta z$  and with the wavenumber  $k$ . Clearly, this approach can be used when the potential is transformed from one plane to another, because the step  $\Delta z$  must be constant in this approach.

---

## Problems

1. Use Magan to analyse the data in exercise\_5.1.zip. Determine the stages, the average linear velocity for each stage, and the angular velocities assuming an angular distance  $\theta = 50^\circ$  from the Euler poles;
2. Use Magan to analyse the data in exercise\_5.2.zip. This exercise requires to deal with spreading asymmetry. Determine the stages and the average linear velocity for each stage;
3. Use Magan to analyse the data in exercise\_5.3.zip. This exercise includes a possible ridge jump. Determine the stages and the average linear velocity for each stage;
4. Use Magan to analyse the data in exercise\_5.4.zip. This exercise includes several magnetic profiles from the same area. Determine the stages and use a GIS to visually build isochrons. Then, use the procedure described in the book to build a kinematic model;

## References

- Backus G, Gilbert F (1968) The resolving power of gross earth data. *Geophys J R Astron Soc* 16:169–205
- Blakely RJ (1996) Potential theory in gravity and magnetic applications. Cambridge University Press, Cambridge, 441 pp
- Bott MHP (1967) Solution of the linear inverse problem in magnetic interpretation with application to oceanic magnetic anomalies. *Geophys J R Astron Soc* 13: 313–323
- Bott MHP, Hutton MA (1970) A matrix method for interpreting oceanic magnetic anomalies. *Geophys J R Astron Soc* 20:149–157
- Chang T (1993) Spherical regression and the statistics of tectonic plate reconstructions. *Int Stat Rev* 61(2): 299–316
- Hellinger SJ (1981) The uncertainties of finite rotations in plate tectonics. *J Geophys Res* 86:9312–9318
- Kirkwood BH, Royer J-Y, Chang TC, Gordon RG (1999) Statistical tools for estimating and combining finite rotations and their uncertainties. *Geophys J Int* 137: 408–428
- Klitgord KD, Schouten H (1986) Plate kinematics of the central Atlantic. In: Vogt PR, Tucholke BE (eds) *The geology of North America*, vol. M, the western north Atlantic region. Geological Society of America, Boulder, pp 351–378
- Korhonen JV et al (2007) Magnetic anomaly map of the world (and associated DVD), scale: 1:50,000,000, 1st edn. Commission for the Geological Map of the World, Paris
- Luyendyk APJ (1997) Processing of airborne magnetic data. AGSO. *J Aust Geol Geophys* 17(2):31–38
- Matias LM, Olivet J-L, Aslanian D, Fidalgo L (2005) PLACA: a white box for plate reconstruction and best-fit pole determination. *Comput Geosci* 31: 437–452
- Mauring E, Kihle O (2006) Leveling aerogeophysical data using a moving differential median filter. *Geophysics* 71(1):L5–L11. doi:10.1190/1.2163912
- Mauring E, Beard LP, Kihle O, Smethurst MA (2002) A comparison of aeromagnetic levelling techniques with an introduction to median levelling. *Geophys Prospect* 50(1):43–54
- Minty BRS (1991) Simple micro-levelling for aeromagnetic data. *Explor Geophys* 22(4):591–592
- Müller RD, Roest WR, Royer J-Y, Gahagan LM, Sclater JG (1997) Digital isochrons of the world's ocean floor. *J Geophys Res* 102(B2):3211–3214
- Nabighian MN, Grauch VJS, Hansen RO, LaFehr TR, Li Y, Peirce JW, Phillips JD, Ruder ME (2005) The historical development of the magnetic method in exploration. *Geophysics* 70(6):33ND–61ND. doi:10.1190/1.2133784
- Parker RL (1972) The rapid calculation of potential anomalies. *Geophys J R Astron Soc* 31:447–455
- Parker RL (1974) A new method for modeling marine gravity and magnetic anomalies. *J Geophys Res* 79(14):2014–2016
- Parker RL, Huestis SP (1974) The inversion of magnetic anomalies in the presence of topography. *J Geophys Res* 79(11):1587–1593
- Poularikas AD, (2010) *The handbook of formulas and tables for signal processing*, vol. 13, CRC Press 1103 pp
- Rowley DB, Lottes AL (1988) Plate-kinematic reconstructions of the North Atlantic and Arctic: late Jurassic to present. *Tectonophysics* 155:73–120
- Royer JY, Müller RD, Gahagan LM, Lawver LA, Mayes CL, Nürnberg D, Sclater JG (1992) A global isochron chart. Technical Rep. University of Texas Institute for Geophysics, Austin. 117, 38 pp
- Sander EL, Mrazek CP (1982) Regression technique to remove temporal variation from geomagnetic survey data. *Geophysics* 47(10):1437–1443
- Sandwell DT, Smith WHF (1997) Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *J Geophys Res* 102:10039–10054
- Schettino A (2012) Magan: a new approach to the analysis and interpretation of marine magnetic anomalies. *Comput Geosci* 39C:135–144. doi:10.1016/j.cageo.2011.07.007
- Schettino A, Scotese CR (2005) Apparent polar wander paths for the major continents (200 Ma – present day): a paleomagnetic reference frame for global plate tectonic reconstructions. *Geophys J Int* 163(2): 727–759
- Schettino A, Turco E (2006) Plate kinematics of the Western Mediterranean region during the Oligocene and early Miocene. *Geophys J Int* 166(3): 1398–1423
- Stock JM, Molnar P (1983) Some geometrical aspects of uncertainties in combined plate reconstructions. *Geology* 11:697–701
- Talwani M, Ewing M (1960) Rapid computation of gravitational attraction of three-dimensional bodies of arbitrary shape. *Geophysics* 25(1):203–225
- Talwani M, Worzel J, Landisman M (1959) Rapid gravity computations for two-dimensional bodies with application to the Mendocino submarine fracture zone. *J Geophys Res* 64(1):49–59
- Tivey MA (1996) Vertical magnetic structure of ocean crust determined from near-bottom magnetic field measurements. *J Geophys Res* 101(B9):20275–20296. doi:10.1029/96JB01307
- Vine FJ, Matthews DH (1963) Magnetic anomalies over oceanic ridges. *Nature* 199:947–949
- Won IJ, Bevis M (1987) Computing the gravitational and magnetic anomalies due to a polygon: algorithms and Fortran subroutines. *Geophysics* 52(2): 232–238
- Yarger HL, Robertson RR, Wentland RL (1978) Diurnal drift removal from aeromagnetic data using least squares. *Geophysics* 43(6):1148–1156

## Abstract

Even in times preceding the worldwide acceptance of the plate tectonics paradigm, paleomagnetic data have furnished the most striking evidence of Wegener's continental drift theory. In this chapter, I present selected topics from paleomagnetism, which are essential for using paleomagnetic data in plate kinematics modelling. First, I introduce Neel's theory of thermo-remanent magnetization, which represents the form of primary remanent magnetism acquired by most igneous rocks. Then, the basics of paleomagnetic sampling and statistics are discussed. Finally, the chapter presents some advanced applications, such as the construction of apparent polar wander paths and true polar wander, as well as a global plate motions model since the Triassic.

## 6.1 Néel's Theory of Single-Domain TRM

In Chap. 3, we have introduced the physics of ferromagnetism in its generality, without considering the different kinds of NRM that are carried by Earth's rocks. Here we are going to describe TRM, which is the form of primary remanent magnetism acquired by most igneous rocks and some high-grade metamorphic rocks when they are cooled in a weak field  $H$  below Curie's temperature. This kind of NRM is considered the most reliable record of the past directions and intensities of the geomagnetic field, especially when it is carried by SD grains. In general, the TRM of SD grains is very stable over geological times, despite geomagnetic polarity reversals, re-

heating, and other secondary processes. However, such a long-term stability does not imply an exact conservation of the total magnetization vector, because for  $H = 0$  the final state of equilibrium of a rock containing a large ensemble of SD grains requires in any case that  $M_r = 0$ . This demagnetized limit equilibrium state is not a consequence of the fact that the magnetization vectors of the individual grains have decreased to zero, because an SD grain is *always* magnetized to saturation (see Sect. 3.5). The total magnetization of an ensemble of SD grains decays to zero as a consequence of the tendency for the magnetization vectors of the individual grains to acquire a random alignment. At any temperature, the thermal energy is responsible for this relaxation towards a demagnetized equilibrium state. Now we are going to introduce the classic theory

of Néel (1949), which represents still today the fundamental starting point for a description of single-domain TRM.

Let us consider an assemblage of spheroidal SD grains aligned with the external field  $\mathbf{H}$ . This configuration could be attained by an ensemble of grains characterized by *uniaxial anisotropy*, meaning that there exists a unique preferred (“easy”) axis of magnetization (in either direction). In this instance, each grain has two minimum-energy states, with magnetization  $M_1 = +M_S$  (grain moment aligned with  $\mathbf{H}$ ) or  $M_2 = -M_S$ . In general, from (3.69) we see that the potential energy associated with a grain having magnetization  $\mathbf{M}$  and volume  $V$  is given by:

$$U = -V\mathbf{M} \cdot \mathbf{B} = -\mu_0 V\mathbf{M} \cdot \mathbf{H} \quad (6.1)$$

Therefore, the minimum-energy states of a grain in the assemblage will be:  $U_1 = -\mu_0 VM_S H$  and  $U_2 = +\mu_0 VM_S H$ . If the number of grains is sufficiently large, the equilibrium configuration for this assemblage is a Maxwell-Boltzmann distribution of the magnetization directions, in which the average magnetization is determined by the Boltzmann partition function:

$$M_{eq} = \frac{\sum_{i=1}^2 M_i e^{-U_i/kT}}{\sum_{i=1}^2 e^{-U_i/kT}} = M_S \tanh\left(\frac{\mu_0 VM_S H}{kT}\right) \quad (6.2)$$

This formula shows that  $M_{eq} = 0$  for  $H = 0$ . If  $N$  is the total number of grains in the ensemble, at any time  $t$  we have that  $n$  grains are in a state with parallel alignment with the external field (state 1), and  $N - n$  have anti-parallel alignment. To evaluate the probability of a transition between the two states, we must first determine the energy barrier separating them. To this purpose, it is necessary to take into account that in addition to the strong exchange interaction described in Chap. 3 between the atoms of a ferromagnetic material, there exist dipole-dipole interactions and interactions between the magnetic moments

and the electric field of the crystal lattice (spin-orbit interactions) (e.g., Kachkachi et al. 2000). Although weak, these interactions operate over a long range, so that their contribution is not generally negligible. Indeed, they introduce anisotropy in the system, that is, a dependence of the total energy from the direction of magnetization. Here we shall consider only the dipole-dipole interactions, which play an important role in small systems such as SD grains. The potential energy associated with these forces is given by (3.73). Considering all dipole pairs in a SD grain, the total magnetostatic energy of the dipole-dipole interaction can be written as follows:

$$U_D = -\frac{\mu_0 \mu_B^2}{4\pi} \sum_{i \neq j} \left[ 3 \frac{(\mathbf{S}_i \cdot \mathbf{r}_{ij})(\mathbf{S}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5} - \frac{\mathbf{S}_i \cdot \mathbf{S}_j}{r_{ij}^3} \right] \quad (6.3)$$

where the sum is extended to all dipole pairs at sites  $i$  and  $j$  and  $r_{ij}$  is their distance. In the case of ellipsoidal grains with volume  $V$ , it is possible to show that (6.3) determines a macroscopic *shape anisotropy*, which is a form of magnetic anisotropy (Kachkachi et al. 2000), and the total magnetostatic energy for the dipole-dipole interactions assumes the following simple expression:

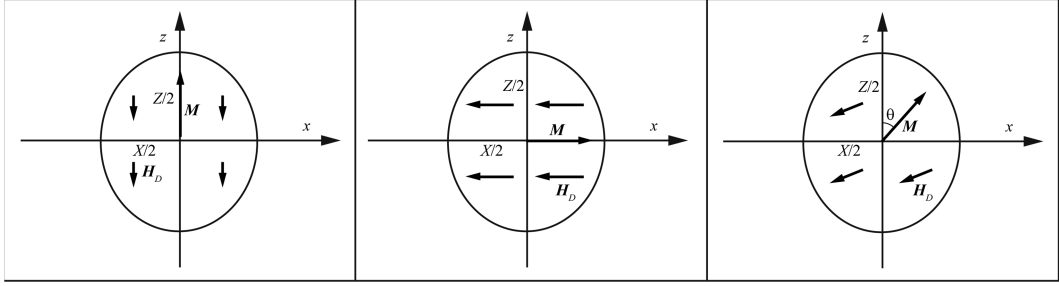
$$U_D = \frac{1}{2} \mu_0 V \left( D_x M_x^2 + D_y M_y^2 + D_z M_z^2 \right) \quad (6.4)$$

where  $D_x$ ,  $D_y$ , and  $D_z$  are positive quantities called *demagnetizing factors* for reasons that will be clear shortly. These quantities depend from the grain geometry and size. For a prolate spheroid with axes  $X$ ,  $Y = X$ , and  $Z > X$ , they are given by:

$$D_z = \frac{1 - e^2}{2e^3} \left[ \lg \left( \frac{1 + e}{1 - e} \right) - 2e \right] \\ D_x = D_y = \frac{1}{2} (1 - D_z) \quad (6.5)$$

where  $0 < e < 1$  is the ellipsoid eccentricity:

$$e = \sqrt{1 - (X/Z)^2}$$



**Fig. 6.1** Internal demagnetizing field in a prolate spheroid. The minimum intensity of  $H_D$  occurs when  $\mathbf{M}$  is parallel to the major axis  $Z$ , because  $D_z \leq D_x = D_y$  (left). Conversely, the maximum intensity is obtained when  $\mathbf{M}$

is parallel to the minor axis  $X$  (center). For any other direction of  $\mathbf{M}$ , the demagnetizing field is not perfectly antiparallel to  $\mathbf{M}$

The dipole-dipole interaction can be represented by an additional magnetic field,  $\mathbf{H}_D$ , which is essentially confined within each grain in the assemblage, so that the total magnetostatic energy is given by:

$$U_T = -\mu_0 V \mathbf{M} \cdot (\mathbf{H} + \mathbf{H}_D) \\ = \mu_0 V M_i H_i - \mu_0 V M_i H_{D,i} \quad (6.6)$$

where we have used (6.1) and the summation convention. Now let us rewrite (6.4) in tensor form:

$$U_D = \frac{1}{2} \mu_0 V M_i D_{ij} M_j \quad (6.7)$$

where the tensor  $\mathbf{D}$  is defined as:  $D_{ij} \equiv D_i \delta_{ij}$ . A comparison of (6.7) with (6.6) gives:

$$H_{D,i} = -\frac{1}{2} D_{ij} M_j \quad (6.8)$$

Therefore, the internal field  $\mathbf{H}_D$  opposes magnetization. This is the reason why this field, which represents the magnetostatic dipole-dipole interaction within a grain, is referred to as the *demagnetizing field*. For the same reason, the tensor  $D_{ij}$  is called the *demagnetizing tensor*. In some cases, for example when the magnetization is aligned with one of the major axes of an ellipsoidal grain,  $\mathbf{D}$  reduces to a scalar, so that  $D_{ij} = D \delta_{ij}$ . Now we want to determine the total magnetostatic energy for a spheroidal SD grain as a function of the magnetization direction. Figure 6.1 illustrates three possibilities. In the

general case that  $\mathbf{M}$  forms an angle  $\theta \neq n\pi/2$  ( $n \geq 0$ ) with the major axis  $Z$ , the demagnetizing field  $\mathbf{H}_D$  is not antiparallel to  $\mathbf{M}$ , as it results:

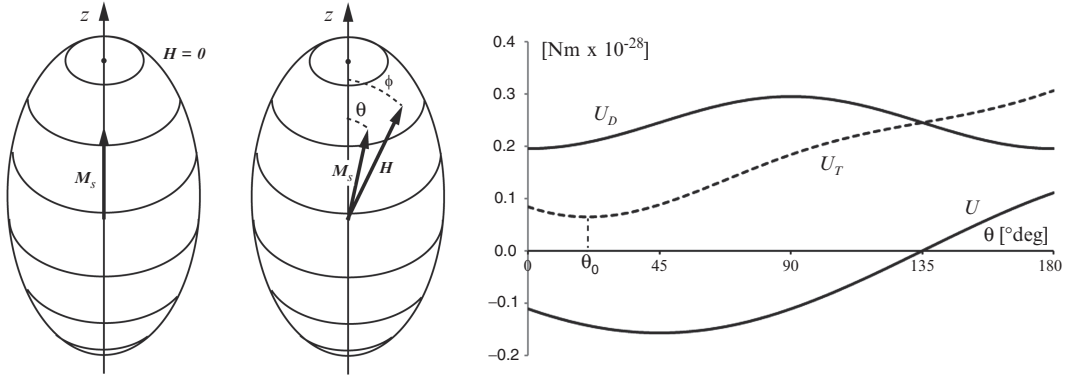
$$\mathbf{H}_D = -\frac{1}{2} \mathbf{M} (D_x \sin \theta \mathbf{i} + D_y \cos \theta \mathbf{k}) \quad (6.9)$$

Therefore, writing the magnetization vector as:  $\mathbf{M} = M \sin \theta \mathbf{i} + M \cos \theta \mathbf{k}$ , we see that the magnetostatic dipole-dipole energy is given by:

$$U_D = -\mu_0 V \mathbf{M} \cdot \mathbf{H}_D \\ = \frac{1}{2} \mu_0 V M^2 [D_z + (D_x - D_z) \sin^2 \theta] \quad (6.10)$$

As pointed out by Dunlop and Özdemir (1997), the difference  $D_x - D_z$  is often more important than  $D_x$  and  $D_z$  separately. Expression (6.10) shows that  $U_D$  is anisotropic with respect to the shape of the grains. This shape anisotropy is uniaxial in the case of prolate spheroids, which have a unique easy axis of magnetization coinciding with the major axis ( $\theta = 0^\circ$  or  $\theta = 180^\circ$ ). Now let us assume that an SD spheroidal grain is magnetized at saturation along its major (easy) axis. If we apply an external field  $\mathbf{H}$  at an angle  $\phi$  with respect to the major axis, the magnetization vector rotates by an angle  $\theta$  with respect to the major axis, as illustrated in Fig. 6.2.

During this rotation the total energy  $U_T$  changes as illustrated in Fig. 6.2, until it reaches



**Fig. 6.2** Rotation of the SD grain magnetization after application of an external field  $\mathbf{H}$ , whose axis forms an angle  $\phi$  with the easy grain axis. The plot on the right shows the total energy (dashed line) as a function

of the rotation angle  $\theta$ . The curves have been traced assuming  $\phi = 45^\circ$ ,  $V = 1.25 \times 10^{-22} \text{ m}^3$ ,  $M_S = 1 \text{ A/m}$ ,  $H = 0.1 \text{ A/m}$ , and  $X/Z = 0.71 (e = 0.7)$ . In this example, the total energy minimum is attained for  $\theta = 20^\circ$

a minimum at an equilibrium angle  $\theta_0$ . For any rotation angle  $\theta$ , the total energy is given by the sum of the potential energy associated with the external field  $\mathbf{H}$  (Eq. 6.1) and the internal demagnetizing energy  $U_D$ :

$$\begin{aligned} U_T(\theta, \phi) &= -\mu_0 V \mathbf{M}_S \cdot \mathbf{H} - \frac{1}{2} \mu_0 V \mathbf{M}_S \cdot \mathbf{H}_D = \\ &= -\mu_0 V M_S H \cos(\phi - \theta) \\ &\quad + \frac{1}{2} \mu_0 V M_S^2 [D_z + (D_x - D_z) \sin^2 \theta] \end{aligned} \quad (6.11)$$

The plot of  $U_D$  shows that in absence of external field the demagnetizing field provides an energy barrier for the complete reversal of the grain magnetization. To determine the equilibrium angle in (6.11), we simply set to zero the first derivative of  $U_T$ :

$$\begin{aligned} 0 = \frac{\partial U_T}{\partial \theta} \bigg|_{\theta=\theta_0} &= -\mu_0 V M_S H \sin(\phi - \theta_0) \\ &\quad + \frac{1}{2} \mu_0 V M_S^2 (D_x - D_z) \sin 2\theta_0 \end{aligned} \quad (6.12)$$

Hence,

$$H \sin(\phi - \theta_0) = \frac{1}{2} M_S (D_x - D_z) \sin 2\theta_0 \quad (6.13)$$

For  $\phi = 0$ , Eq. (6.13) assumes the form:

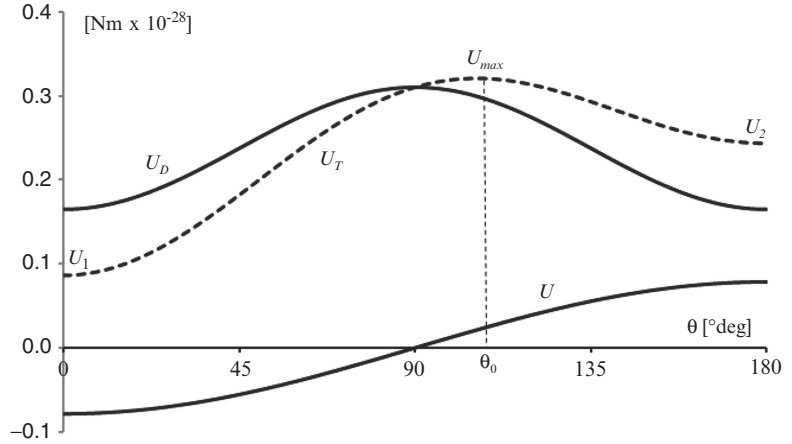
$$\cos \theta_0 = -\frac{H}{M_S (D_x - D_z)} \equiv -\frac{H}{H_c} \quad (6.14)$$

This equation has solutions only for  $|H| \leq H_c$ . The quantity  $H_c$  is called *microscopic coercivity* or simply *micro-coercivity* of the SD grain. For  $|H| > H_c$  there is only one minimum for  $U_T$ , thereby, the grain magnetization undergoes an irreversible rotation to the unique stable configuration. Now let us come back to the Maxwell-Boltzmann distribution of the magnetization directions for an assemblage of spheroidal SD grains (Eq. 6.2). In this instance  $\phi = 0$ , thereby at any time  $t$ , for  $n$  grains, we have that  $\theta = 0$ , while  $N - n$  grains have anti-parallel alignment, so that  $\theta = 180^\circ$ . Taking into account of the demagnetizing energy, the corresponding energy levels for these states are:

$$\begin{aligned} U_1 &= -\mu_0 V M_S H + \frac{1}{2} \mu_0 V M_S^2 D_z \\ U_2 &= +\mu_0 V M_S H + \frac{1}{2} \mu_0 V M_S^2 D_z \end{aligned} \quad (6.15)$$

Switching from state 1 to state 2 implies a rotation  $\theta = 180^\circ$ . As shown in Fig. 6.3, this transition requires passing the energy peak at  $\theta = \theta_0$ . To rotate from state 1 to state 2, the thermal energy must be in excess of  $\Delta U_{12} = U_{\max} - U_1$ .

**Fig. 6.3** Total energy of a prolate spheroidal SD grain, magnetized in the direction of the easy axis ( $\phi = 0$ , dashed line) as a function of the rotation angle  $\theta$ . The curves have been traced assuming  $V = 1.25 \times 10^{-22} \text{ m}^3$ ,  $M_S = 1 \text{ A/m}$ ,  $H = 0.05 \text{ A/m}$ , and  $X/Z = 0.60$  ( $e = 0.8$ ). In this example, the total energy maximum is attained for  $\theta = 106^\circ$



Similarly, to have a rotation from state 2 to state 1, the thermal energy must be in excess of  $\Delta U_{21} = U_{\max} - U_2$ . Insertion of (6.14) in (6.11) gives:

$$U_{\max} = \mu_0 V M_S \frac{H^2}{H_c} + \frac{1}{2} \mu_0 V M_S^2 \times \left[ D_z + (D_x - D_z) \left( 1 - \frac{H^2}{H_c^2} \right) \right] \quad (6.16)$$

Therefore, using (6.15) we see that the two energy barriers for state transitions are given by:

$$\begin{aligned} \Delta U_{12} &= \frac{1}{2} \mu_0 V M_S H_c \left( 1 + \frac{H}{H_c} \right)^2 \\ \Delta U_{21} &= \frac{1}{2} \mu_0 V M_S H_c \left( 1 - \frac{H}{H_c} \right)^2 \end{aligned} \quad (6.17)$$

Clearly, for  $H > 0$ , the transition from state 1 to state 2 is more difficult than the reverse one. Therefore, the number of grains,  $n$ , in state 1 increases progressively until the thermal equilibrium is attained. A kinetic equation that describes this process can be obtained assuming that the number of transitions per unit time from one state to another is proportional to the corresponding population of grains in the initial state. Therefore,

$$\dot{n}(t) = K_{21} [N - n(t)] - K_{12} n(t) \quad (6.18)$$

where  $K_{12}$  and  $K_{21}$  are, respectively, the probability of a transition from state 1 ( $\theta = 0^\circ$ ) to state 2 ( $\theta = 180^\circ$ ) and from state 2 to state 1. At any time, the average magnetization is given by:

$$\begin{aligned} M(t) &= \frac{-[N - n(t)] M_S + n(t) M_S}{N} \\ &= \frac{2n(t) - N}{N} M_S \end{aligned} \quad (6.19)$$

In terms of  $M$ , Eq. (6.18) assumes the form:

$$\dot{M}(t) + K M(t) = (K_{21} - K_{12}) M_S \quad (6.20)$$

where:

$$K = \frac{1}{\tau} \equiv K_{12} + K_{21} \quad (6.21)$$

The solution to Eq. (6.20) is immediate:

$$M(t) = M_0 e^{-t/\tau} + M_{eq} (1 - e^{-t/\tau}) \quad (6.22)$$

where  $M_0 = M(0)$  and  $M_{eq} = (K_{21} - K_{12}) M_S / K$  is the limit equilibrium state for  $t \rightarrow \infty$ . A direct formula for  $M_{eq}$  is expression (6.1). The quantity  $\tau$  is called *relaxation time*. It can be calculated from the probabilities of state transitions  $K_{12}$  and  $K_{21}$  using (6.21). These quantities obey to an Arrhenius equation for the temperature dependence of transition rates, with thermal activation

energies  $\Delta U_{12}$  and  $\Delta U_{21}$  and pre-exponential factor  $1/\tau_0$  (Dunlop and Özdemir 1997):

$$\begin{aligned}
 K_{12} &= \frac{1}{\tau_0} \exp\left(-\frac{\Delta U_{12}}{kT}\right) \\
 &= \frac{1}{\tau_0} \exp\left[-\frac{\mu_0 V M_S H_c}{2kT} \left(1 + \frac{H}{H_c}\right)^2\right] \\
 K_{21} &= \frac{1}{\tau_0} \exp\left(-\frac{\Delta U_{21}}{kT}\right) \\
 &= \frac{1}{\tau_0} \exp\left[-\frac{\mu_0 V M_S H_c}{2kT} \left(1 - \frac{H}{H_c}\right)^2\right]
 \end{aligned} \tag{6.23}$$

where we have used (6.17) and  $\tau_0 \sim 10^{-9}$ – $10^{-10}$  s is the atomic time interval of reorganization between successive state transitions. This quantity determines the frequency at which the grains try to switch their direction of magnetization. When  $H > 0$ ,  $\Delta U_{12} > \Delta U_{21}$  and rotations from  $\theta = 180^\circ$  to  $\theta = 0^\circ$  are favoured, so that  $K_{12} < K_{21}$ . If  $H$  is sufficiently large, state 1 is strongly favoured and  $M_{eq} \rightarrow M_S$ . Similarly, if  $H < 0$  and sufficiently large, then state 2 is strongly favoured and  $M_{eq} \rightarrow -M_S$ . In this instance, we can write:

$$\frac{1}{\tau} \cong \frac{1}{\tau_0} \exp\left[-\frac{\mu_0 V M_S H_c}{2kT} \left(1 - \frac{|H|}{H_c}\right)^2\right] \tag{6.24}$$

for  $|H|$  sufficiently large. Conversely, for  $H = 0$  we have that  $\Delta U_{12} = \Delta U_{21}$  and Eqs. 6.21 and 6.23 give:

$$\frac{1}{\tau} = \frac{2}{\tau_0} \exp\left[-\frac{\mu_0 V M_S H_c}{2kT}\right] \tag{6.25}$$

According to (6.22),  $M(t)$  will be close to the equilibrium value  $M_{eq}$  after a time interval coinciding with the relaxation time  $\tau$ . The factor  $U_B = \frac{1}{2}\mu_0 V M_S H_c$  is usually referred to as the *blocking energy*. Note that both  $M_S$  and  $H_c$  in this expression depend upon  $T$ . From (6.24) to (6.25), we see that the relaxation time strongly depends from the temperature  $T$ . For temperatures close to the Curie temperature  $T_c$ , the

blocking energy is small and the equilibrium is attained in a short time interval. This situation is referred to as an *unblocked* or *superparamagnetic* condition. At room temperature, where  $U_B$  is large, the equilibrium is attained very slowly, so that the non-equilibrium state can be considered as frozen or *blocked* in the system. For a given grain volume  $V$ , there is a *critical blocking temperature*,  $T_B$ , at which  $\tau$  becomes small (e.g., 60–100 s) during a heating experiment. Conversely, during a rapid cooling from above  $T_B$ , the equilibrium value of magnetization  $M_{eq}$  is frozen in the sample as TRM. This is the main conclusion of Néel's theory of thermal relaxation. On the basis of this model, TRM can be removed only reheating a sample to its original blocking temperature. In the next section, we shall see that this feature furnishes a method for removing secondary TRM from rock samples.

## 6.2 Paleomagnetic Sampling and Cleaning Procedures

Well before the discovery of marine magnetic anomalies in the early 1960s, paleomagnetism provided the strongest evidence for continental drift, although not yet for plate tectonics in the modern acceptance. This important branch of Earth Sciences supplied a large number of data from all continents since the 1950s, which proved that both igneous and sedimentary rocks formed at latitudes that did not generally coincide with their present day latitudes. This conclusion was supported by the anomalous *inclination* of NRM vectors observed in rock samples of various ages. Furthermore, these data showed that going back to ages older than Pliocene, the *declination* of the magnetization vectors was distributed differently from what we expected on the basis of secular variation. The strike of the horizontal component of magnetization seemed to migrate away from the Earth's spin axis in so far as the rock ages increased. As pointed out by Irving (2005), until the late 1960s the dominant belief among geophysicists and geologists, especially in the United States, was a form of *fixism*, which assumed that

the continents were not, and had never been, subject to horizontal motion. Curiously, even some famous seismologists, like Sir Harold Jeffreys, were proud opponents of *mobilism* and the theory of Wegener (1912) about continental drift (Frankel 2012). However, by the end of 1950s, a large amount of evidence for continental drift had accumulated, and most of this information came from paleomagnetic data collected on continental crust.

Paleomagnetic sampling on continents is usually structured on a hierarchical basis. At the top level, we have *rock units*, which often coincide with geologic formations or with one of their members. Sampling is performed at several *sites* across a rock unit to give a unique estimation of the NRM direction associated with the rock formation. In general, at least ten sites are required to achieve acceptable confidence limits in the estimated paleomagnetic direction. A site can be a lava flow, a dike, a pluton, or any other igneous body unit. In sedimentary rocks, sites usually coincide with specific layers in the stratigraphic succession. For each site, paleomagnetists collect six to eight *samples* using a portable drilling device. The original orientation of the individual specimens is accurately annotated for the subsequent reconstruction of the direction of magnetization after laboratory treatment.

Once the specimens have been brought to a paleomagnetic lab, they are processed to isolate their primary remnant magnetization, a procedure that is referred to as *cleaning*. Usually, NRM of the order of  $10^{-5}$  A/m are sufficient to obtain meaningful results. The modern measuring device for the determination of the NRM of a sample is the *cryogenic magnetometer*. This device uses a complex sensor based on superconductivity, which is called SQUID (Superconducting QUantum Interference Device). It can measure NRM of rock specimens with total magnetic moment  $MV \leq 10^{-10}$  Am<sup>2</sup>. The NRM vectors measured on a set of  $N$  samples from  $B$  sites generally include two components: a primary NRM acquired during rock formation (TRM, CRM, or DRM), and a secondary component that is gradually acquired during the geological time

(the *viscous* remnant magnetization, or VRM) or even after sampling (Eq. 3.79). Clearly, the secondary NRM must be considered as a form of noise in the applications of paleomagnetism to plate kinematics, thereby, a series of laboratory procedures have been designed to remove this component of magnetization. To eliminate a secondary NRM, a rock sample is subject to *partial demagnetization*, which operates on the component of magnetization with lower stability. This low-stability component usually coincides with the secondary NRM that we want to remove, whereas the high-stability component that is isolated through this procedure often coincides with the primary NRM. However, exceptions exist, so that the high-stability component is usually referred to as the *characteristic component* of remnant magnetization (ChRM), which does not necessarily coincide with the primary NRM.

There are two basic techniques for accomplishing partial demagnetization of a sample. In the *alternate field (AF) demagnetization*, a specimen is exposed for about 1 min to an alternating magnetic field with linearly decaying intensity from an initial amplitude  $H_{AF} \sim 100$  mT. The typical frequency of this signal is 400 Hz, and the device allows to rotate automatically the sample in order to align in turn all the specimen axes with the applied field. At the time  $t$  when a positive peak  $|H^+(t)|$  is reached, only the domains with micro-coercivity  $H_c \leq |H^+(t)|$  align their spins with the external field. Because the amplitude of each half-cycle is smaller than the previous one, after a half-period  $T/2$  a slightly smaller set of domains with micro-coercivity  $H_c \leq |H^-(t + T/2)| < |H^+(t)|$  will align their spins in the opposite (“down”) direction. Therefore, if the rate of decrease of the AF is not large, only a small fraction of spins will be frozen in the “up” direction, so that the total magnetic moment of the grains in these two intervals of micro-coercivity  $H_c$  will be approximately zero. Similarly, at the next step, only a small fraction of spins will be frozen in the “down” direction, thereby, after several cycles the net magnetization along each axis will be removed from the sample for all

grains having relaxation times corresponding to micro-coercivity  $H_c \leq |H(0)|$ . Then,  $H(0)$  is increased and the procedure is repeated to remove the next more stable component, which should have a different direction of magnetization. If the new component has the same direction of magnetization of the one detected at the previous step, then it is interpreted as the ChRM. AF demagnetization is effective when the dominant ferromagnetic mineral is a titanomagnetite. In these rocks, secondary NRM is mostly associated with multi-domain grains (see Sect. 3.5), whereas the ChRM is retained by single-domain grains with higher micro-coercivity.

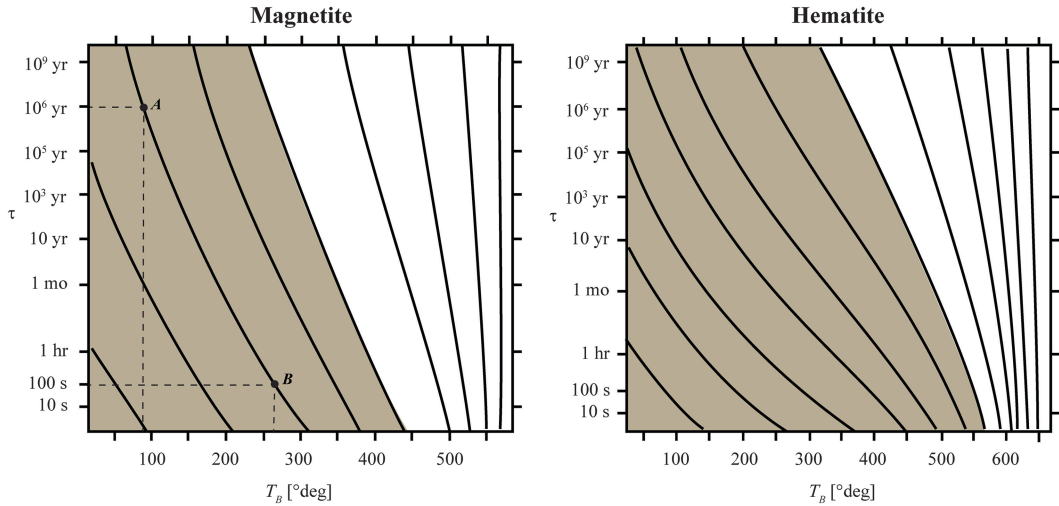
The second fundamental technique for the removal of secondary NRM is *thermal demagnetization*. This method requires heating of a specimen to high temperature below the Curie point of the constituent ferromagnetic minerals, then cooling to room temperature in zero magnetic field. It is based on the thermal relaxation of an assemblage of SD grains described in the previous section. A real rock sample can be considered as formed by multiple ensembles of grains with different physical and geometrical characteristics. Grains with different volumes  $V$  have distinct blocking temperatures  $T_B$ , so that when the rock cools and passes through the various  $T_B$ , the relaxation times of the corresponding grain assemblages increase quickly. Therefore, different equilibrium magnetizations are frozen during this process, so that any subsequent variation in the direction of the external field at lower temperatures does not affect these components. This means that the TRM is not acquired at one time just below the Curie temperature, but during a long time interval over a set of blocking temperatures. Therefore, reheating a sample to a temperature  $T < T_c$  implies unblocking of the frozen magnetizations for all grain populations with  $T_B < T$  and the consequent removal of these components of TRM. Expression (6.24) shows that the relaxation time of can vary over a wide range. SD grains with small values of  $\tau$  are

called *superparamagnetic grains*: their remnant magnetization decays quickly to zero after removal of the magnetizing field. These grains are those that more likely acquire VRM through the geological time. For fixed temperature, relation (6.24) shows that the relaxation time varies with the grain volume  $V$  and the micro-coercivity  $H_c$ . Grains with low values of the product  $VH_c$  will have shorter relaxation time with respect to grains with higher values of this parameter. Good paleomagnetic recorders must have values of  $\tau$  of the order of several hundred Myrs. However, even superparamagnetic grains are converted into stable grains at low temperature. We have seen that the temperature at which this transition occurs is the blocking temperature,  $T_B$ . Any SD grain is superparamagnetic between the Curie temperature and the blocking temperature.

In the previous section, we have shown that the stability of the TRM acquired by an assemblage of SD grains is expressed in terms of relaxation time  $\tau$ . According to Eqs. 6.23 and 6.25 this quantity increases rapidly when the temperature  $T$  decreases. Let us consider an SD grain assemblage with a fixed value of  $VH_c$ . If the TRM of these grains has a relaxation time  $\tau$  of geological length for some temperature  $T$ , it is possible to determine the blocking temperature at laboratory time scale, namely, the temperature at which a sample must be heated to reset its TRM through a zero external field in a short time interval, say  $\tau_B = 60\text{--}100$  s. In fact, using (6.25) we have that:

$$\frac{T_B \lg(2\tau_B/\tau_0)}{M_S(T_B)H_c(T_B)} = \frac{T \lg(2\tau/\tau_0)}{M_S(T)H_c(T)} \quad (6.26)$$

Plots of the blocking temperature as a function of the relaxation time are shown in Fig. 6.4. These plots can be used to determine the blocking temperature as a function of the initial relaxation time. For example, an assemblage of SD magnetite grains with relaxation time  $\tau = 1$  Myr at  $93^\circ\text{C}$  is expected to have acquired substantial VRM if it has been kept at this temperature for



**Fig. 6.4** Relation between blocking temperature,  $T_B$ , and relaxation time,  $\tau$ , for magnetite and hematite. A line on these diagrams represents a set of pairs  $(\tau, T)$  that resets the TRM of an SD assemblage (Redrawn from Pullaiah et al. (1975))

1 Myr (point A in Fig. 6.4). However, if we heat the sample to  $270^\circ\text{C}$  for 100 s in a zero magnetic field, then this secondary magnetization is unblocked and reset to zero (point B in Fig. 6.4). We also note that grain assemblages associated with curves that are placed on the right side of the diagrams (high unblocking temperatures) are reset by a small increase of temperature. This is a consequence of the rapid increase of  $\tau$  for decreasing  $T$  when the assemblage has  $T_B$  close to the Curie temperature. Therefore, grain assemblages in the grey regions of Fig. 6.4, which acquire VRM at relatively low temperatures ( $250\text{--}350^\circ\text{C}$ ) over geological time intervals ( $\sim 10$  Myrs), are unstable carriers of primary TRM. Conversely, grains in the white regions of Fig. 6.4 have blocking temperatures within  $\sim 100^\circ\text{C}$  of  $T_c$  for any relaxation time, so that  $T_B$  is insensitive to  $\tau$  and resetting their TRM is more difficult. These grain assemblages are the main carriers of primary TRM. The plots in Fig. 6.4 predict that primary TRM can be retained after a heating episode within the greenschist metamorphic range ( $300\text{--}500^\circ\text{C}$ ) but not within the amphibolite range (above  $550^\circ\text{C}$ ).

### 6.3 Paleomagnetic Directions

The result of paleomagnetic sampling and the subsequent laboratory treatment is a set of  $N$  magnetization vectors  $(\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_N)$  for a rock unit of known radiometric or stratigraphic age. In plate tectonics, we are generally interested only in the *paleomagnetic directions*, and not in the magnitude of these vectors. Therefore, the data set is usually expressed in terms of  $N$  pairs  $(I_k, D_k)$  in a local reference frame,  $I_k$  being the inclination of the  $k$ -th magnetization vector and  $D_k$  being its declination. Paleomagnetic directions observed at a single site are generally scattered, for example as a consequence of inaccurate orientation of the specimens. However, such within-site scattering is flanked by an inter-site dispersion, associated with the secular variation of the geomagnetic field (see Sect. 4.3). In fact, the sampling procedures are designed in such a way that the samples of a single site have approximately the same age, so that variations of magnetization between the different sites of a

rock unit are representative of secular variation. A key assumption in paleomagnetism is that the  $N$  paleomagnetic directions sampled across a rock unit form a statistically meaningful data set for the determination of a time-averaged paleomagnetic field that can be adequately approximated by a geocentric dipole field.

To determine a time-averaged paleomagnetic field from a collection of paleomagnetic directions, researchers use Fisher's (1953) statistics. The average paleomagnetic direction is calculated simply as the vector sum of the individual vectors, normalized to unity. Therefore,

using (2.29) we see that the Cartesian coordinates  $(X, Y, Z)$  of the estimated mean direction are given by:

$$\begin{aligned} X &= \frac{1}{R} \sum_{k=1}^N \cos I_k \cos D_k; \\ Y &= \frac{1}{R} \sum_{k=1}^N \cos I_k \sin D_k; \quad Z = \frac{1}{R} \sum_{k=1}^N \sin I_k \end{aligned} \quad (6.27)$$

where  $R \leq N$  is the magnitude of the resultant vector:

$$R = \sqrt{\left( \sum_{k=1}^N \cos I_k \cos D_k \right)^2 + \left( \sum_{k=1}^N \cos I_k \sin D_k \right)^2 + \left( \sum_{k=1}^N \sin I_k \right)^2} \quad (6.28)$$

Finally, using the inverse transformation (2.30), we see that the declination and inclination of the mean direction are given by:

$$\begin{cases} D = \arctan(Y/X) \\ I = \arcsin(Z) \end{cases} \quad (6.29)$$

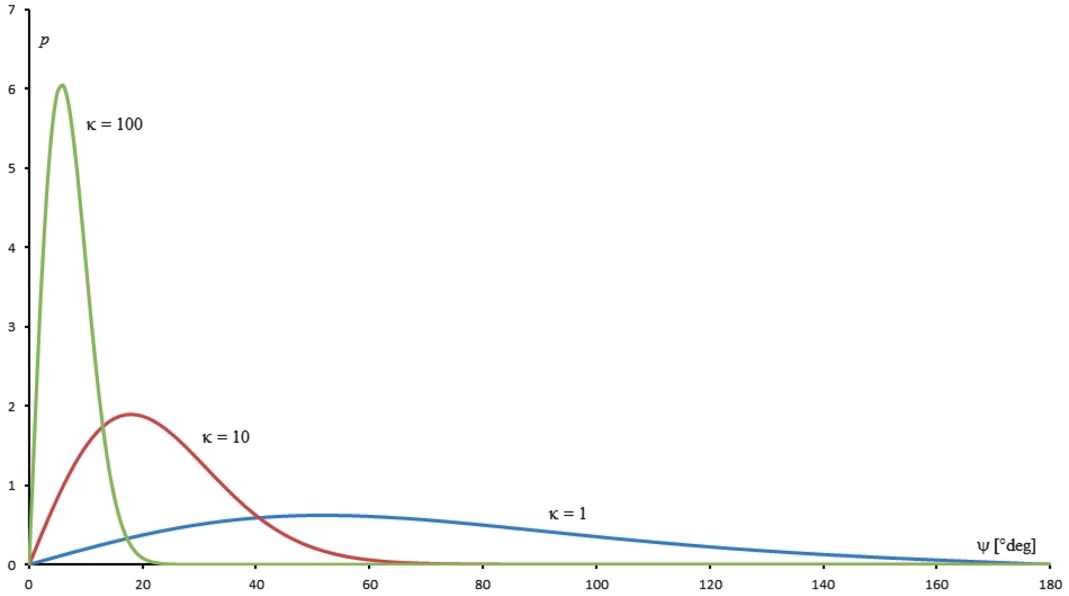
The statistical parameters of uncertainty associated with the estimated mean direction are calculated assuming that the data set of paleomagnetic directions can be modelled as a *Fisher distribution*, a spherical analogue of the Gaussian distribution, in which the concentration of unit vectors about the mean is proportional to  $\exp(\kappa \cos \psi)$ , where  $\kappa$  is a *precision* parameter and  $\psi$  is the angle between an observation and the true mean direction. The precision parameter describes the dispersion of the points about the mean. When  $\kappa$  is small, the distribution is highly dispersed, whereas for large  $\kappa$  it is concentrated about the mean. In a Fisher distribution, the probability to find a unit vector in a region  $\mathbf{R}$  of the spherical surface of radius 1 is given by:

$$P(\mathbf{R}) = \int_{\mathbf{R}} \frac{\kappa}{4\pi \sinh(\kappa)} \exp(\kappa \cos \psi) dS \quad (6.30)$$

The constant factor  $\kappa/(4\pi \sinh \kappa)$  in (6.30) ensures that the probability normalizes to unity when  $\mathbf{R}$  coincides with the whole sphere. The element area  $dS$  in the integral (6.30) is a spherical surface element, which at angular distance  $\psi$  is given by:  $dS = \sin \psi d\psi d\phi$ ,  $\phi$  being the azimuth of the point about the mean. The distribution is uniformly distributed with respect to the azimuthal angle  $\phi$ , thereby, integrating over  $\phi$  we have that the probability  $P$  to find an observation within an angle  $\psi$  from the true mean is given by:

$$\begin{aligned} P(\psi) &= \int_0^\psi \frac{\kappa}{2 \sinh(\kappa)} \exp(\kappa \cos \psi') \sin \psi' d\psi' \\ &= \frac{e^\kappa - e^{\kappa \cos \psi}}{e^\kappa - e^{-\kappa}} \end{aligned} \quad (6.31)$$

The probability density function  $p$  associated with (6.31) is shown in Fig. 6.5 for three different precision parameters. Note that the presence of  $\sin \psi$  in (6.31) determines a maximum for  $p$  that is offset with respect to  $\psi = 0$ . From (6.31) it is possible to calculate the angle  $\psi_{95}$  within which we can find an observation with a probability of



**Fig. 6.5** Probability density function,  $p = p(\psi)$ , associated with the distribution (6.31). For any angle  $\psi$ , the probability to find an observation within a band of width  $d\psi$  between  $\psi$  and  $\psi + d\psi$  is given by  $p(\psi)d\psi$

95 %:

$$\psi_{95} = \arccos \left[ \frac{1}{\kappa} \lg (0.05e^{\kappa} + 0.95e^{-\kappa}) \right] \quad (6.32)$$

$$\alpha_{95} = \arccos \left\{ 1 - \frac{N-R}{R} \left[ \left( \frac{1}{0.05} \right)^{\frac{1}{N-1}} - 1 \right] \right\} \approx \frac{140^\circ}{\sqrt{kR}} \quad (6.34)$$

Fisher (1953) showed that for  $\kappa \geq 3$  the best estimate of the precision parameter is given by:

$$k = \frac{N-1}{N-R} \quad (6.33)$$

This is a best estimate in the sense that  $1/k$  is both a minimum variance and unbiased estimator of  $\kappa^{-1}$  (McFadden 1980). Fisher (1953) also proved that if we take groups of  $N$  observations from the distribution (6.30), then the directions of the resultant vectors, which are the estimated means, belong themselves to a Fisher distribution about the true mean, with precision parameter  $\kappa R$ . Therefore, he deduced that for  $\kappa \geq 3$  the true mean direction of the distribution has 95 % probability to lie within a spherical circle of radius  $\alpha_{95}$  about the resultant vector  $(X, Y, Z)$ , where:

This is the 95 % confidence circle used in paleomagnetism. In general, the averaging procedure is applied at different levels. At the lowest level, for each sample that includes several specimens, the ChRM directions of the single specimens are averaged. Then, site-means are calculated from the sample means of each site. Finally, site-means are averaged to give the final paleomagnetic direction of the rock unit. However, in the next section we shall see that averaging of paleomagnetic data continues even at higher levels, when they are used to determine the kinematics of tectonic plates. Although there are no strong rules, generally paleomagnetists consider  $k > 30$  and  $\alpha_{95} < 15^\circ$  as minimum acceptability parameters for site means (e.g., Butler 1992). Apart from the determination of the confidence circle, a series

of statistical tests is performed before drawing geological conclusion from paleomagnetic data. These tests, which are described in detail in the book of McElhinny and McFadden (2000), include:

- Testing whether an apparent outlier is truly discordant with other observations, so that it should be rejected;
- Testing whether two or more sets of paleomagnetic observations could have been drawn from a common Fisher distribution, so that they would have a common true mean direction and a common precision parameter  $\kappa$ ;
- Testing whether a scattered set of paleomagnetic directions have been sampled from a uniform random population;
- Testing whether a set of observed paleomagnetic directions conforms with a Fisher distribution, so that the statistical methods described above are applicable.

Here we are going to describe the last of these tests, which is of fundamental importance in the plate kinematics applications. The test requires more than one step. First, the azimuthal angle  $\phi$  must be distributed uniformly between 0 and  $2\pi$ . Secondly, the distribution of the polar angles  $\psi$  must be conform to the probability density (6.31). From (6.31), we see that the probability to find an observation between angles  $\psi_1$  and  $\psi_2$  about the true mean is given by:

$$P(\psi_1, \psi_2) = \frac{e^{\kappa \cos \psi_1} - e^{\kappa \cos \psi_2}}{2 \sinh \kappa} \quad (6.35)$$

Therefore, for  $N$  observations, the expected frequency of polar angles  $\psi_1 \leq \psi \leq \psi_2$  is given by:

$$\hat{f}(\psi_1, \psi_2) = \frac{N}{2 \sinh k} (e^{k \cos \psi_1} - e^{k \cos \psi_2}) \quad (6.36)$$

where  $k$  is given by (6.33). Similarly, the expected frequency of azimuthal angles  $\phi_1 \leq \phi \leq \phi_2$  is given by:

$$\hat{f}(\phi_1, \phi_2) = \frac{N}{2\pi} (\phi_2 - \phi_1) \quad (6.37)$$

To test either the azimuthal or the radial distribution, we separate the observations into  $m$  polar or azimuthal classes. For each class, the statistic  $X^2$  is calculated by the following expression:

$$X^2 = \sum_{i=1}^m \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} = \sum_{i=1}^m \frac{f_i^2}{\hat{f}_i} - N \quad (6.38)$$

where the  $f_i$  are observed frequencies, while the  $\hat{f}_i$  are calculated by (6.36) for polar classes and from (6.37) in the case of azimuthal classes. If the observations were drawn from a Fisher distribution, then the statistic  $X^2$  has a  $\chi^2$  distribution with  $\nu$  degrees of freedom, where:

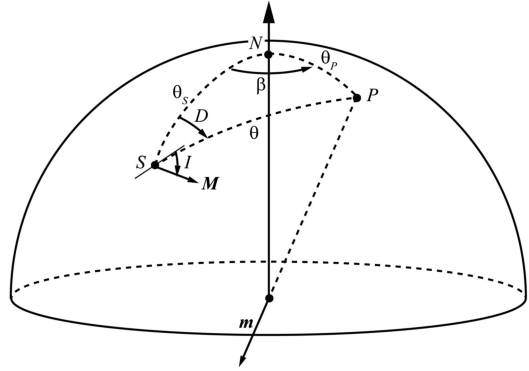
$$\nu = m - 1 - \Pi \quad (6.39)$$

and  $\Pi$  is the number of parameters that have been replaced by their maximum likelihood estimates to calculate the observed frequencies. For example, testing a data set of geomagnetic poles against the GAD hypothesis would require just one estimated parameter in (6.36), because the true mean is known and coincides with the North Pole. Therefore, we would have that  $\Pi = 1$  in the radial distribution test and  $\Pi = 0$  in the azimuthal test. However, typically, we do not know the true mean direction, so that it will be estimated through expressions (6.27)–(6.28). In this case we have that  $\Pi = 3$  in the radial distribution test and  $\Pi = 2$  in the azimuthal test.

## 6.4 Paleopoles and Apparent Polar Wander Paths

The declination  $D$  and inclination  $I$  of a rock unit, which result from averaging site means, are quantities that depend from the geographic position of the rock formation and from the tectonic history of the continent to which these rocks belong. Assuming that the corresponding time-averaged magnetization has been acquired by exposition to a geocentric dipole field, we can easily determine the coordinates of the *paleomagnetic pole*, or

**Fig. 6.6** Geometrical relations between a rock unit at location  $S$ , whose time-averaged magnetization vector  $\mathbf{M}$  has declination  $D$  and inclination  $I$ , and the corresponding paleopole, at location  $P$ , of a geocentric paleomagnetic dipole field with magnetic moment  $\mathbf{m}$ . Angle  $\theta$  is the site paleo-colatitude



simply the *paleopole*, associated with this field. In fact, for a magnetic dipole field, the inclination is related to the distance from the field pole by the dipole equation (4.49), while the declination tells us the direction where to find this pole. Figure 6.6 illustrates the geometric relationships between the various parameters.

Let  $S \equiv (\theta_S, \phi_S)$  be the geographic location (colatitude and longitude) of a rock unit, for which a time-averaged remnant magnetization vector  $\mathbf{M}$  with declination  $D$  and inclination  $I$  has been determined. If this is considered as equivalent to an NRM that was acquired by exposition to a geocentric dipole field, then from the dipole equation (4.49) we have that the paleocolatitude  $\theta$  of the site can be calculated readily by the following formula:

$$\theta = \cot^{-1} \left( \frac{1}{2} \tan I \right) \quad (6.40)$$

To determine the colatitude of the paleopole  $P$ , we can use the spherical version of the law of cosines:

$$\cos \theta_P = \cos \theta_S \cos \theta + \sin \theta_S \sin \theta \cos D \quad (6.41)$$

A little bit more complicate is to determine the paleopole longitude  $\phi_P$ . Let  $\beta$  be the longitude difference between paleopole and site. By the spherical version of the law of sines we have that:

$$\frac{\sin \theta}{\sin \beta} = \frac{\sin \theta_P}{\sin D} \quad (6.42)$$

Therefore,

$$\beta = \arcsin \left( \frac{\sin D \sin \theta}{\sin \theta_P} \right) \quad (6.43)$$

This formula constrains  $\beta$  to be in the range  $[-\pi/2, +\pi/2]$ . In fact, there is a source of ambiguity arising from the fact that a longitude difference  $\beta$  and a difference  $\pi - \beta$  give the same sine. Therefore, using (6.43) we cannot distinguish a situation where a paleopole  $P$  lies in the same hemisphere of  $S$  from a situation in which  $P$  is in the opposite hemisphere. Applying again the law of cosines to the spherical triangle  $(S, P, N)$  of Fig. 6.6, we see that the two situations give:

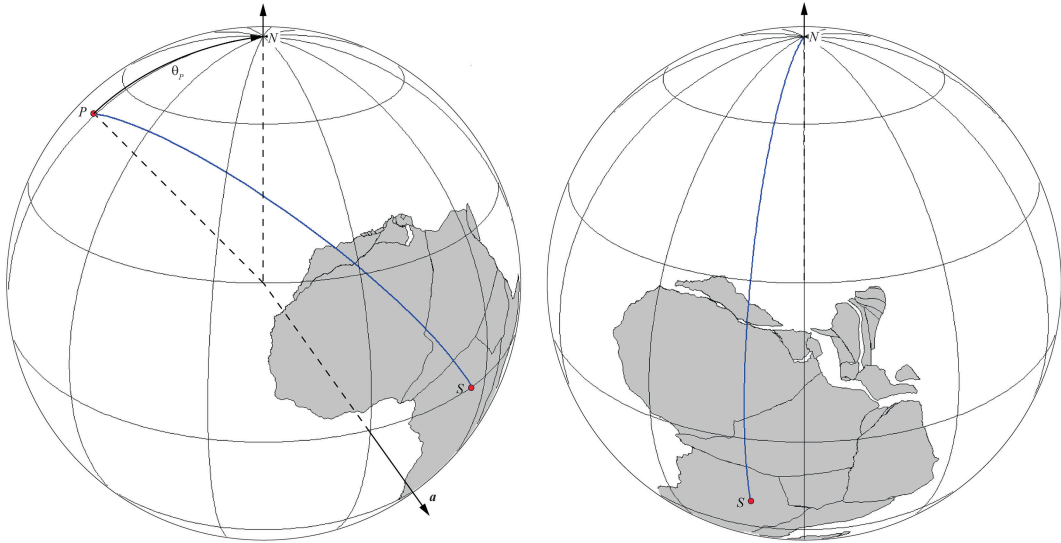
$$\begin{aligned} \cos \theta &= \cos \theta_P \cos \theta_S + \sin \theta_P \sin \theta_S \\ \begin{cases} \cos \beta & S \text{ and } P \text{ in the same hemisphere} \\ \cos(\pi - \beta) & S \text{ and } P \text{ in opposite hemispheres} \end{cases} \end{aligned} \quad (6.44)$$

Therefore,

$$\cos \theta = \cos \theta_P \cos \theta_S \pm \sin \theta_P \sin \theta_S \cos \beta \quad (6.45)$$

Now we note that  $\sin \theta_P \sin \theta_S \cos \beta \geq 0$  in any case, thereby, the two situations can be distinguished comparing  $\cos \theta$  with  $\cos \theta_P \cos \theta_S$ :

$$\beta = \begin{cases} \phi_P - \phi_S & \text{for } \cos \theta \geq \cos \theta_P \cos \theta_S \\ \phi_S - \phi_P + \pi & \text{for } \cos \theta < \cos \theta_P \cos \theta_S \end{cases} \quad (6.46)$$



**Fig. 6.7** Transformation of a paleomagnetic field at  $t = 140$  Ma (early Cretaceous), with pole at  $P$  in the reference frame of Africa (present day coordinates), into a GAD field. The transformation is performed applying

an Euler rotation that moves  $P$  to  $N$ . In this example, the Euler axis of rotation,  $a$ , has pole at  $(0^\circ, \phi_P + 90^\circ)$  and the rotation angle is  $-\theta_P$

Therefore, solving for  $\phi_P$  gives:

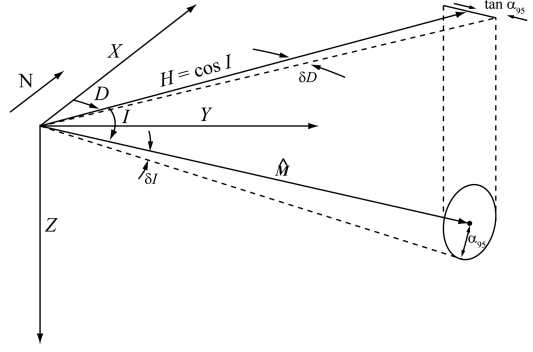
$$\phi_P = \begin{cases} \phi_S + \beta & \text{for } \cos \theta \geq \cos \theta_P \cos \theta_S \\ \phi_S + \pi - \beta & \text{for } \cos \theta < \cos \theta_P \cos \theta_S \end{cases} \quad (6.47)$$

Differently from a mean paleomagnetic direction ( $D, I$ ), which is a site-dependent quantity, a paleomagnetic pole of age  $t$  (the mean age attributed to the rock unit) is a *global* quantity that represents the apparent location of the geographic North Pole at time  $t$  in a *reference frame fixed to the continent to which the site  $S$  belongs* (Fig. 6.7). This is clearly a consequence of the GAD hypothesis (see Sect. 4.3), which states that a time-averaged geomagnetic field approximately coincides with a GAD field not only in the present epoch but also in the geologic past. Therefore, any Euler rotation that brings a paleopole to the geographic North Pole will also transform the time-averaged paleomagnetic field associated with the paleopole into a GAD field. This rotation coincides with a transformation from a reference frame where the continent is at rest in the present day position to a geocentric reference frame in which the paleopole  $P$  coincides with the geo-

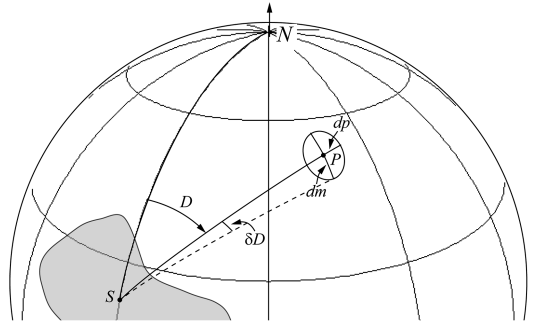
graphic North Pole  $N$ , as shown in Fig. 6.7, and the continent has been restored to its original orientation with respect to the spin axis. However, as a GAD field is symmetric with respect to the spin axis, there is no way to establish a paleolongitude of the site  $S$  relative to the present day location. In other words, there are infinitely many rotations that moves  $P$  to  $N$ , which determine infinitely many longitudinal displacements of the site  $S$ . All these reconstructions are admissible, and there is no way to determine the correct longitude variation of  $S$  from the present day location to its reconstructed position. A common method to perform paleomagnetic reconstructions is illustrated in Fig. 6.7. In this approach, the paleopole is moved to the North Pole along its own meridian of longitude through an Euler rotation about an equatorial pole at  $(0, \phi_P + 90^\circ)$  by an angle  $-\theta_P$  (e.g., Ziegler et al. 1983; Schettino and Scotese 2005). We shall come back to this point in the next section.

Once that we have determined the paleopole corresponding to a paleomagnetic direction ( $D, I$ ), the next step is assigning uncertainty parameters to the location  $(\theta_P, \phi_P)$ . To this purpose, let us first consider the uncertainty of a paleomagnetic

**Fig. 6.8** Determination of the inclination and declination uncertainty from the confidence cone  $\alpha_{95}$



**Fig. 6.9** Colatitude error,  $dp$ , and relationship between declination uncertainty,  $\delta D$ , and trasversal uncertainty in the paleopole direction,  $dm$



direction in terms of confidence limits for the inclination and the declination, as illustrated in Fig. 6.8. From simple trigonometry, we have that:

$$\delta I = \alpha_{95}; \quad \tan \delta D = \frac{\tan \alpha_{95}}{\cos I} \quad (6.48)$$

where the second of these equations is usually expressed in the approximated form:

$$\delta D \cong \frac{\alpha_{95}}{\cos I} \quad (6.49)$$

These uncertainties can be easily converted into a confidence oval about the paleopole, with mutually orthogonal semi-axes  $dp$  and  $dm$ , as illustrated in Fig. 6.9. Applying the spherical law of sines we obtain:

$$\frac{\sin dm}{\sin \delta D} = \sin p \quad (6.50)$$

Therefore,

$$dm \cong \delta D \sin p = \frac{\alpha_{95} \sin p}{\cos I} \quad (6.51)$$

The colatitude error  $dp$  can be calculated using the dipole equation (6.40). It results:

$$\begin{aligned} dp &= -\frac{d\theta}{dI} \delta I = -\alpha_{95} \frac{d\theta}{dI} = \frac{2\alpha_{95}}{1 + 3\cos^2 I} \\ &= \frac{1}{2} \alpha_{95} (1 + 3\cos^2 \theta) \end{aligned} \quad (6.52)$$

The pair  $(dp, dm)$  is called *oval of 95 % confidence about the paleopole*. In an alternative approach, each site direction is converted first into a *virtual geomagnetic pole (VGP)*. Then, the resulting set of VGPs is considered as a Fisher distribution and analyzed using the technique described above. In this instance, conventions require that the precision parameter and the 95 % confidence cone be indicated by the capital symbols  $K$  and  $A_{95}$ , respectively (e.g., Van der Voo 1993). It can be shown that the approximate relation between the confidence cone  $A_{95}$  and the uncertainties  $dp$  and  $dm$  is:

$$A_{95} \approx \sqrt{dp dm} \quad (6.53)$$

Paleopoles are fundamental quantities in the application of paleomagnetism to

plate kinematics. Their importance arises from the capability to predict both the paleolatitude and the declination at *any* reference site on a tectonic plate. Therefore, combining paleopoles of different age, determined from different rock units, it is possible to build curves that describe the variations of paleolatitude and declination through the geological time at a common reference site (Van der Voo 1993; Schettino and Scotese 2005). Let  $\mathbf{p}$  and  $\mathbf{s}$  be the unit vectors associated respectively with a paleopole  $P$  and a reference site  $S$ . The Cartesian coordinates of these vectors can be calculated from the corresponding geographic coordinates using the transformation equations (2.27). The predicted paleolatitude at  $S$  can be determined promptly by the following formula:

$$\lambda = \frac{\pi}{2} - \arccos(\mathbf{p} \cdot \mathbf{s}) \quad (6.54)$$

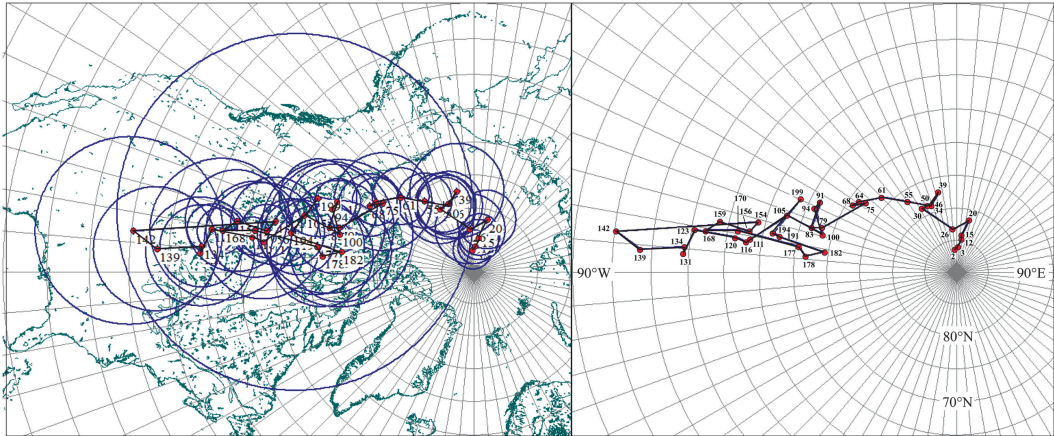
Calculating the predicted declination is likewise simple. We first apply (6.46) to determine  $\beta$ , then the declination can be obtained by Eq. (6.42). Although this technique can be applied using individual paleopoles of different ages from the same continent, the large amount of paleomagnetic data collected during the last decades and the discrepancy between results of similar age has led to the development of statistical techniques for the determination of the “best” mean paleomagnetic pole from a data set formed by several paleopoles of similar age. Similarly, several methods have been proposed for the determination of best-fit time series of paleolatitude and declination, and for the construction of curves on the globe that represent the migration of the mean paleomagnetic pole of a continent through the geological time. The latter curves, which are called *apparent polar wander paths* (APW paths), represent the most important result of paleomagnetism for the study of continental drift. The determination of mean paleopoles is not much different from the calculation of individual paleopoles from data sets of VGPs. However, in this instance the data are always weighted through their confidence cones  $A_{95}$  and possibly through their age uncertainties.

For example, if we wish to determine the mean paleopole for North America during the Paleocene, we should select all the paleopoles obtained from rocks on the cratonic part of this continent, with an age between 65.5 and 56 Ma, and satisfying some quality criteria (e.g., Van der Voo 1990, 1993). To this purpose, we could use the Global Paleomagnetic Database (GPMDB) (McElhinny and Lock 1990), which is a structured publicly accessible data set of paleopoles (e.g., Schettino and Scotese 2001). Most authors perform a pre-selection of the data according to more or less subjective data reliability criteria. As an example, Van der Voo (1990, 1993) proposed seven reliability criteria that could be satisfied by a paleopole. In this approach, a “quality factor”  $0 \leq Q \leq 7$  is assigned to each datum, which indicates the number of criteria that are satisfied. Van der Voo (1990, 1993) suggests that  $Q \geq 3$  is the minimum requirement for a paleopole to be included in the computation of a representative mean. Let  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N$  be  $N$  unit vectors associated with the selected paleopoles, and let  $w_1, w_2, \dots, w_N$  be weights assigned to each paleopole on the basis of their quality parameters. For example, we could simply set  $w_i = 1/A_{95,i}$  for  $i = 1, 2, \dots, N$ , but other more complex weighting schemes are possible. In some cases (e.g., Torsvik et al. 2001, 2008), even Van der Voo’s quality factor  $Q$  has been used as a weighting factor, although  $Q$  is all except a physical quantity. Once a weighting scheme has been selected, the mean paleopole  $\mathbf{p}$  is calculated by the following formula:

$$\mathbf{p} = \frac{1}{w} \sum_{i=1}^N w_i \mathbf{p}_i \quad (6.55)$$

where  $w$  is a normalization factor, which ensures that  $|\mathbf{p}| = 1$ . To determine the uncertainty parameters of the mean paleopole, we can use again Fisher statistics, granted that the spatial distribution of the data is conformal to a Fisher distribution. This test can be performed using the technique described in Sect. 6.3.

Let us consider now the procedure for constructing APW paths. As mentioned above,



**Fig. 6.10** A sliding-window APW path for the African craton since 200 Ma (Besse and Courtillot 2002), obtained using a 10 Myrs window and 5 Myrs steps. Blue circles are

$A_{95}$  confidence cones (left). Red dots are mean paleopole locations. The right pane shows more clearly the complexity of this path, which includes several loops and hairpins

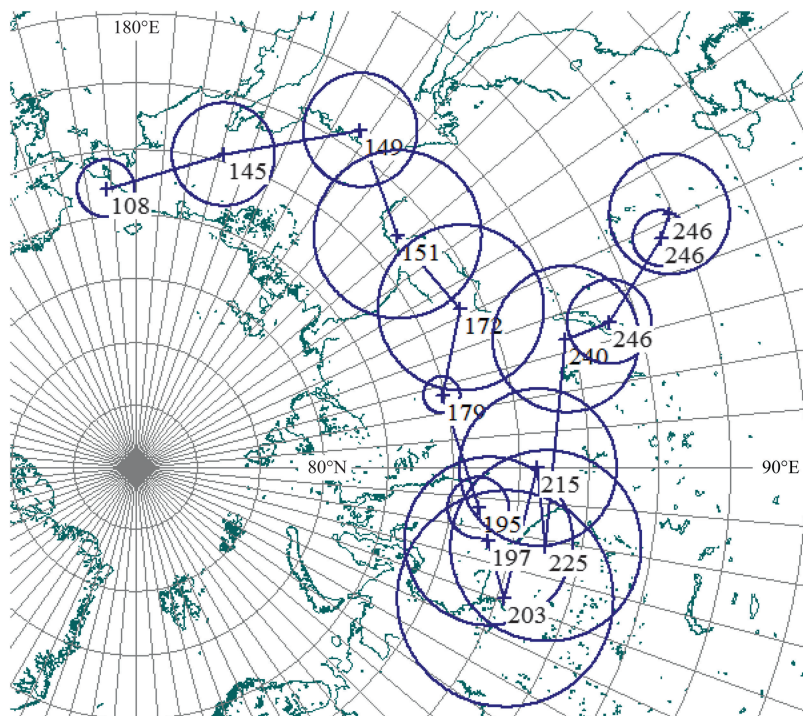
an APW path is a time series of *mean* paleopoles  $p(t_1), p(t_2), \dots, p(t_N)$ , where  $t_1 < t_2 < \dots < t_N$ . Therefore, a vector  $p(t_i)$  is always calculated by a weighted averaging formula like (6.55). Geometrically, the time series is represented by a curve on the unit sphere, formed by a sequence of great circle arcs linking the paleopoles. A potential source of ambiguity in the construction of these curves arises from the geomagnetic field reversals. However, in paleomagnetic databases like the GPMDB such ambiguity is overcome by assigning to each listed paleopole a normal polarity. Eventually, in the case of southern hemisphere continents, it is possible to reverse all the paleomagnetic poles and visualize the APW paths as inverted polarity curves. The first APW paths were built selecting groups of paleopoles according to their *geologic* age. For example, one could start determining the average of all paleopoles having a stratigraphic or radiometric age in the lower Cretaceous time interval, obtaining the representative mean paleopole for the lower Cretaceous. Then, the procedure was repeated for the upper Cretaceous paleopoles, and so on. This simple method was the only possible approach when the number of available paleopoles satisfying minimum reliability criteria was small. Starting from the 1990s, the publication of many new results allowed more refined analyses of the paleomagnetic fields.

Today, there are three general approaches to the construction of APW paths. In the *sliding window* method (e.g., Harrison and Lindh 1982), the paleomagnetic poles with ages falling within a time window of fixed width (for example, 30 Myrs) are averaged to determine a mean paleopole, which it is attributed an age equal to the central age of the interval, or equal to the mean age of the averaged paleopoles. Then, the window is moved by an assigned step (for example, 10 Myrs) and the procedure is repeated. An example of sliding-window APW path is shown in Fig. 6.10.

A major problem with APW paths like that illustrated in Fig. 6.10 is that they show a level of detail higher than what is statistically justifiable. This is quite evident by comparing the confidence limits shown in Fig. 6.10 (left panel) with the average distance between consecutive paleopoles and with the hairpin turns of the path. Therefore, it is likely that several consecutive mean paleopoles of this APW path would pass a test that establishes if two poles have been drawn from a common distribution (e.g., McElhinny and McFadden 2000).

A second, more rigorous, approach to the construction of APW paths consists into the selection of a few reliable paleomagnetic poles of different age, without applying time averaging (e.g., Gordon et al. 1984;

**Fig. 6.11** Triassic – early Cretaceous APW path for N. America, based on a compilation of 15 high-quality paleopoles (May and Butler 1986). Circles are 95 % confidence cones

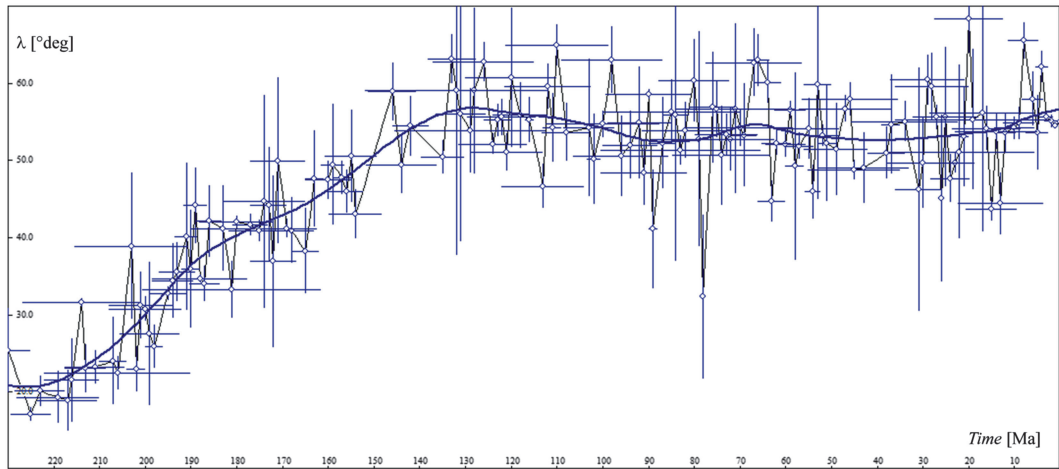


Kent and Van der Voo 1990; May and Butler 1986; van Fossen and Kent 1992; Kent and Witte 1993). According to May and Butler (1986), the most important factor controlling the accuracy of APW paths is the reliability of the selected data base, thereby, these authors propose to select only high-quality paleopoles and to evaluate their time sequence. Clearly, this technique relies on an accurate selection of the data, which will be paleopoles determined using the most severe demagnetization analyses and from a large number of samples and sites. A good example of this class of APW paths is shown in Fig. 6.11.

Although the approach of constructing APW paths from few reliable paleopoles does not suffer the problems of the sliding-window method, it relies too strongly on the more or less subjective process of paleopole selection, which does not guarantee the correctness of each selected datum. For example, only two of the three early Triassic (246 Ma) paleopoles in Fig. 6.11 may have been drawn from the same Fisher distribution. Therefore, the third paleopole could be either correct (in fact, it is located close to the 240 Ma result) or the other two paleopoles are correct and this

result should be discarded. Actually, in the selection of reliable paleopoles for the construction of APW paths, it is necessary taking into account that even the “best” paleomagnetic direction can produce a *wrong* paleopole if it is attributed an incorrect age.

The third approach to the construction of APW paths tries to overcome the limitations of the methods described above through statistical regression techniques. In this instance, APW paths are built by fitting smoothed regression curves on the sphere through swaths of paleomagnetic data. Parker and Denham (1979) were the first to propose an interpolation method based on cubic splines. Similarly, Thompson and Clark (1981) used weighted, least-squares cubic splines to fit smoothed curves to the colatitudes and longitudes of paleopoles from North America and Europe. Musgrave (1989) applied a modified version of the weighted least-squares regression method to a study of Cretaceous and Cenozoic Australian paleomagnetic data, while Jupp and Kent (1987) developed a sophisticated fitting algorithm based on spherical smoothing splines. Examples of APW paths based on the method of



**Fig. 6.12** Predicted paleolatitudes for a reference point in N. America at (55°N, 90°W) since the late Triassic, based on the global compilation of Schettino and Scotese

(2005). The regression curve is a natural cubic spline, built using a smoothing parameter  $\beta = 5$

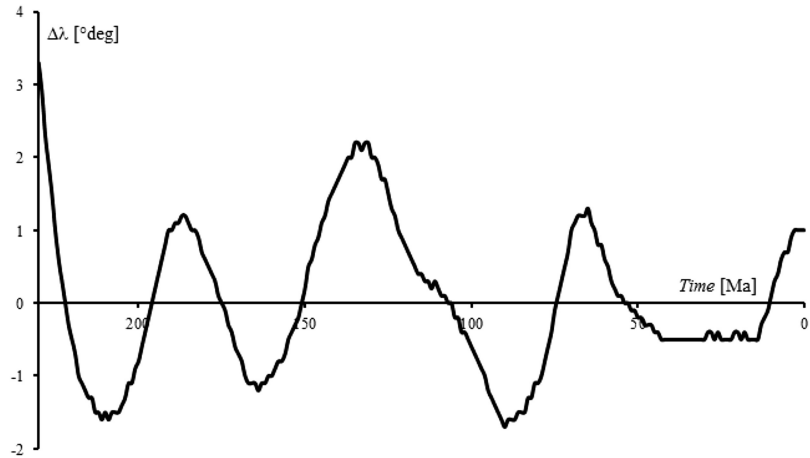
Jupp and Kent (1987) can be found in Torsvik et al. (1996, 2012). An alternative technique of construction of smoothed APW paths has been proposed more recently by Schettino and Scotese (2005). This approach tries to overcome the limits of the spherical splines smoothing algorithms, in which the amount of smoothing is chosen arbitrarily by the researcher, and the regression may determine a best fitting curve that is not so best with regard to the local geology at individual sites. To understand the problem, let us consider the plot of Fig. 6.12, which shows the predicted paleolatitudes of a point on the North American craton since the late Triassic.

The plot in Fig. 6.12 has been built using (6.54) and combining paleopoles from N. America with paleopoles from other continents, which were rotated into N. American coordinates using the rotation model of Schettino and Scotese (2005). We note that the overall trend of paleolatitude for the selected reference point is an approximately linear increase by  $\sim 36^\circ$  from the late Ladinian (230 Ma) to the Barremian (130 Ma), then a more or less constant paleolatitude until recent times. However, the smoothing spline curve of regression, which has  $4.8^\circ$  rms error of residuals, shows a sequence of second-order low-amplitude oscillations about the general trend. These oscillations could be interpreted as real

cycles having geological significance. However, if we used a greater smoothing parameter, say  $\beta = 300$ , in order to generate a regression curve that is more representative of the general trend, the maximum displacement of the spline curve in Fig. 6.12 from the new representative trend would be only  $\Delta\lambda = 3.3^\circ$  (Fig. 6.13). Therefore, the predicted paleolatitude oscillations would have amplitude that is less than the standard deviation of the residuals about the regression line!

This example can be extended to the spherical regression curves that are used in the modelling of APW paths. It shows that the smoothing parameter of a spline regression curve cannot be chosen arbitrarily, but it should be compatible with the dispersion of the data about the regression curve. Another more critical problem of the “crude” statistical approach will be discussed now. To this end, it will be useful to examine in detail some key features of the approach of Schettino and Scotese (2005). These authors compiled a list of paleopoles for each continent by filtering data in the GPMDB according to some minimum-reliability criteria ( $B \geq 4$ ,  $N/B \geq 4$ ,  $A_{95} \leq 15^\circ$ , cleaning procedure code  $\geq 2$ , and half-interval of age uncertainty  $\leq 20$  Myrs). In the analysis of a continent, the paleopoles belonging to other plates were rotated into the local coordinate system of the continent

**Fig. 6.13** Variations of a paleolatitude regression spline with smoothing  $\beta = 5$  (Fig. 6.12) about a spline regression plot having smoothing parameter  $\beta = 300$



being analyzed. The possibility to perform such operation is quite intuitive. If we know the Euler pole that restores the position of a continent with respect to a reference plate at time  $t$ , for example the Brazilian craton with respect to central Africa at 130 Ma, then applying this rotation to its paleopoles of age  $t$  gives a new set of paleomagnetic poles that are fully compatible with those obtained by sampling directly on the reference continent. We say that the paleopoles have been rotated into the coordinate system of the reference continent. Such *paleopole transfer* technique has been applied by several authors to fill gaps in the paleomagnetic record of some continents (especially Africa) and to build global APW paths (e.g., Besse and Courtillot 1991).

The next step in the technique of Schettino and Scotese (2005) was to select an arbitrary reference point on each continent and to build paleolatitude and declination plots relative to these points. Then, the two time series were analyzed independently through a non-parametric spline regression technique. Parametric regression models assume that the form of the regression function,  $\mu$ , is known except for a finite number of parameters. Non-parametric regression models, on the other hand, only require some qualitative properties for  $\mu$ , for example that  $\mu$  belongs to some functional space  $\Gamma$ . An overall measure of performance of an estimator can be expressed by a combination of goodness of the fit and smoothness of the curve  $\mu = \mu(t)$ . For example,

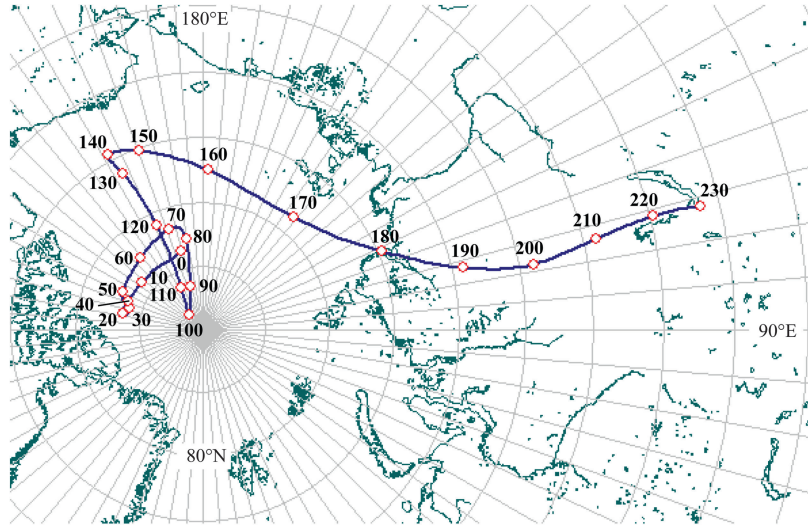
$$\chi^2 = (1 - q) \frac{1}{n} \sum_{i=1}^n w_i [x_i - \mu(t_i)]^2 + q \int_a^b \left( \frac{d^m \mu}{dt^m} \right)^2 dt \quad (6.56)$$

where  $x_i = x(t_i)$  are the  $n$  observed values with positive weights  $w_i$ ,  $m$  is a positive integer, and the parameter  $0 < q < 1$  is used as a balancing factor between goodness-of-fit and smoothing of the estimator. If we set:  $\beta \equiv q/(1 - q)$ , then the functional to be minimized,  $\chi^2$ , can be rewritten as:

$$\chi^2 = \frac{1}{n} \sum_{i=1}^n w_i [x_i - \mu(t_i)]^2 + \beta \int_a^b \left( \frac{d^m \mu}{dt^m} \right)^2 dt \quad (6.57)$$

The minimizing function,  $\mu$ , is a *natural smoothing spline* estimator (Eubank 1999) and the parameter  $\beta$ , which controls balancing between goodness-of-fit and smoothness, is called the *smoothing parameter*. When  $\beta$  is large ( $q \sim 1$ ) smoothness is favoured, whereas estimators having large  $m$ -th derivative are penalized. Conversely, small values of  $\beta$  ( $q \sim 0$ ) tend to select classic least squares estimators and privilege goodness-of-fit. It can be shown that smoothing spline estimators are *natural splines*, that is, piecewise polynomials subject to a

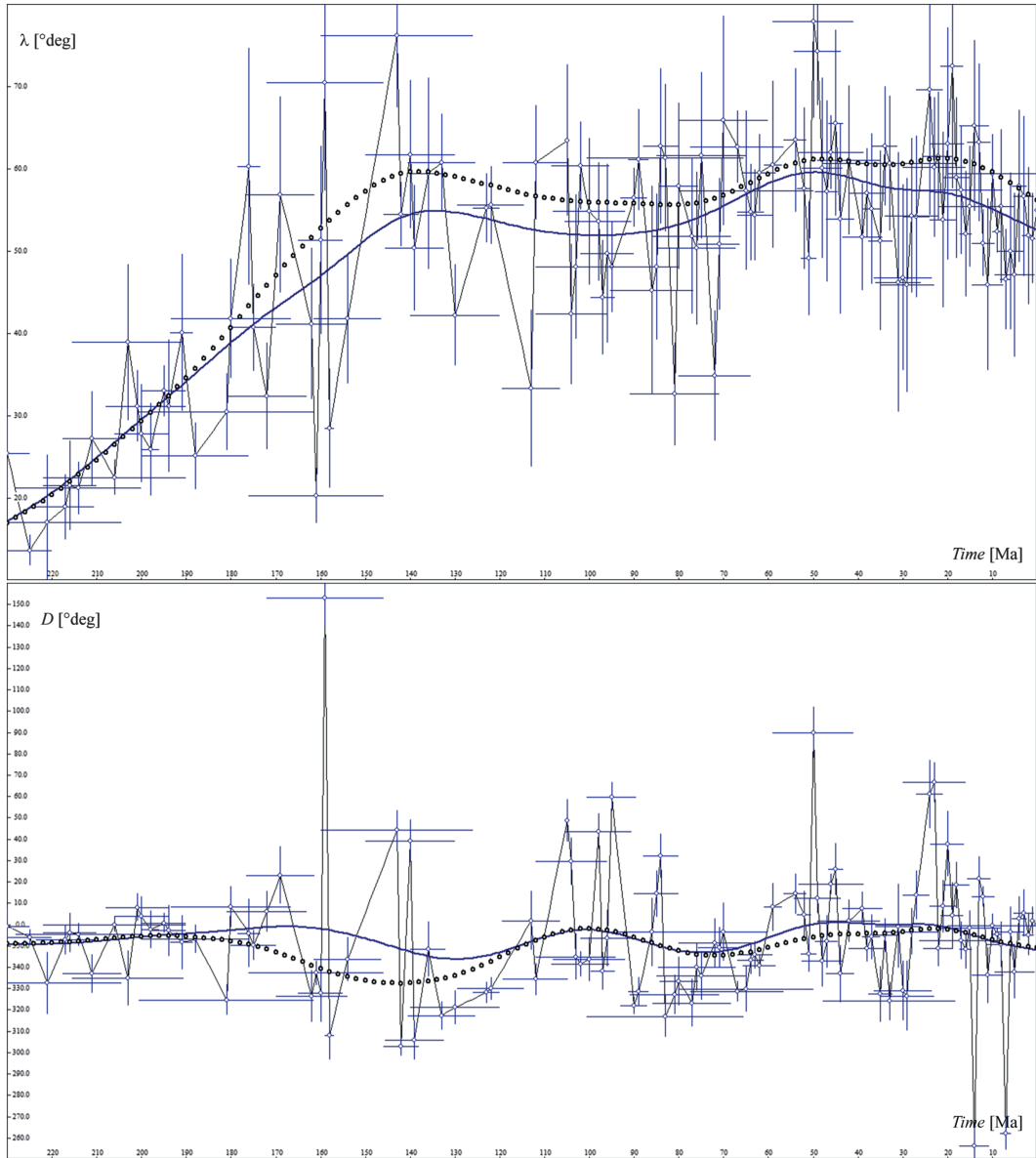
**Fig. 6.14** Spherical spline APW path for N. America. This best-fitting curve has been generated using the algorithm of Jupp and Kent (1987) and a smoothing parameter  $\beta = 50$ . North American paleopoles have been selected from the GPMDB using the minimum-reliability criteria of Schettino and Scotese (2005), but no further filtering has been applied. Confidence limits are not displayed for clarity



maximum number of continuity constraints. The segmented nature of these functions gives them more flexibility than polynomials and allows a better adaption to the local characteristics of the data. The parameters  $\beta$  and  $m$  can be chosen manually, according to some known properties of the estimator, or computed by algorithms that take into account of the actual data and some optimality criteria. When the only assumption about  $\mu$  is that this function is continuously differentiable and has a square integrable second derivative, then  $m = 2$ , and the resulting estimator is a cubic smoothing spline.

The independent statistical analysis of paleolatitude and declination time series represented a major point of controversy, to such an extent that the a famous geophysical journal rejected to publish the paper submitted by the proponents of this method. The criticsers of the approach followed by Schettino and Scotese (2005) questioned the fact that *independent* regressions of declination and paleolatitude data were statistically incorrect (e.g., Ruiz-Martínez et al. 2012), because declination and inclination (hence paleolatitude) are *not* independent variables. To understand this point, let us assume that a time series of unit vectors  $\mathbf{u}_i = \mathbf{u}(t_i)$  is converted into a sequence of site-dependent paleolatitude and declination pairs  $(\lambda_i = \lambda(t_i), D_i = D(t_i))$ . If we perform independent statistical regressions of these two time

series and then convert back the smoothed pairs  $(\lambda'(t), D'(t))$  into unit vectors  $\mathbf{u}'(t)$ , then these vectors will *not* form, in general, a best fitting spherical regression curve of the original time series. In this instance, we say that  $\lambda$  and  $D$  are not independent each other from the statistical point of view. In fact, while  $\lambda'(t)$  and  $D'(t)$  will satisfy at best the maximum likelihood principle for the corresponding curves of paleolatitude and declination, their combination in the spherical variable  $\mathbf{u}'(t)$  will not necessarily fit at best the unit vectors  $\mathbf{u}_i$  in the least-squares sense. Conversely, unit vectors sampled along a real best fitting curve on the sphere (e.g., a spherical spline) will not generally produce paleolatitude and declination pairs that fit at best the corresponding time series. For example, Fig. 6.14 shows a spherical smoothing spline APW path for the North American craton, generated using the algorithm of Jupp and Kent (1987) and without paleopole transfer from other continents. We may use this APW path to get smoothed curves of paleolatitude and declination for a reference point in North America. However, Fig. 6.15 shows that these functions would *not* coincide with best fitting smoothed curves of paleolatitude and declination, which can be built using the same data set and a similar degree of smoothing. In general, the curves of predicted paleolatitude and declination generated from smoothed APW paths will not represent best



**Fig. 6.15** Predicted paleolatitudes (*top*) and declinations (*bottom*) for a point in N. America at ( $55^{\circ}\text{N}$ ,  $90^{\circ}\text{W}$ ) since the late Triassic, based on the same data used for the construction of the APW path of Fig. 6.14. No attempt

has been done to remove outliers. The *solid lines* represent cubic spline regression curves, built using a smoothing parameter  $\beta = 50$ . *Dots* are paleolatitudes and declinations predicted by the spherical spline APW path of Fig. 6.14

fitting estimators of the corresponding time series, although the APW paths from which they are derived are by themselves best fitting spherical curves of paleomagnetic pole time series on the unit sphere.

In summary, we can choose to have alternatively best fitting regression curves of pale-

olatitude and declination, sacrificing the possibility to have a best fitting APW path through the paleopoles, or the vice versa. What is the “correct” or the “best” approach? It depends from what we consider to be our data. For the critics of the method proposed by Schettino and Scotese (2005), the primary input data

in the construction of an APW path are mean paleopoles. In this instance, a smoothed APW path must have the property to be a best fitting regression curve on the sphere in the least-squares sense, and there is no necessity to have best fitting paleolatitude or declination plots. Conversely, in the approach of Schettino and Scotese (2005) declination and inclination represent the primary physical observables, from which it is possible to *calculate* a paleomagnetic pole using some assumptions (e.g., the GAD hypothesis). In this view, the data are paleomagnetic directions determined at sampling sites, and the utility of paleopoles resides exclusively in their capability to predict inclination and declination at any other site, as well as in the possibility to determine total reconstruction poles for the continent.

Therefore, in the view of these authors, the important time series to be analyzed, and for which best fitting regression curves were searched, were the sequences of paleolatitude and declination at representative reference sites and *not* the paleopole time series.

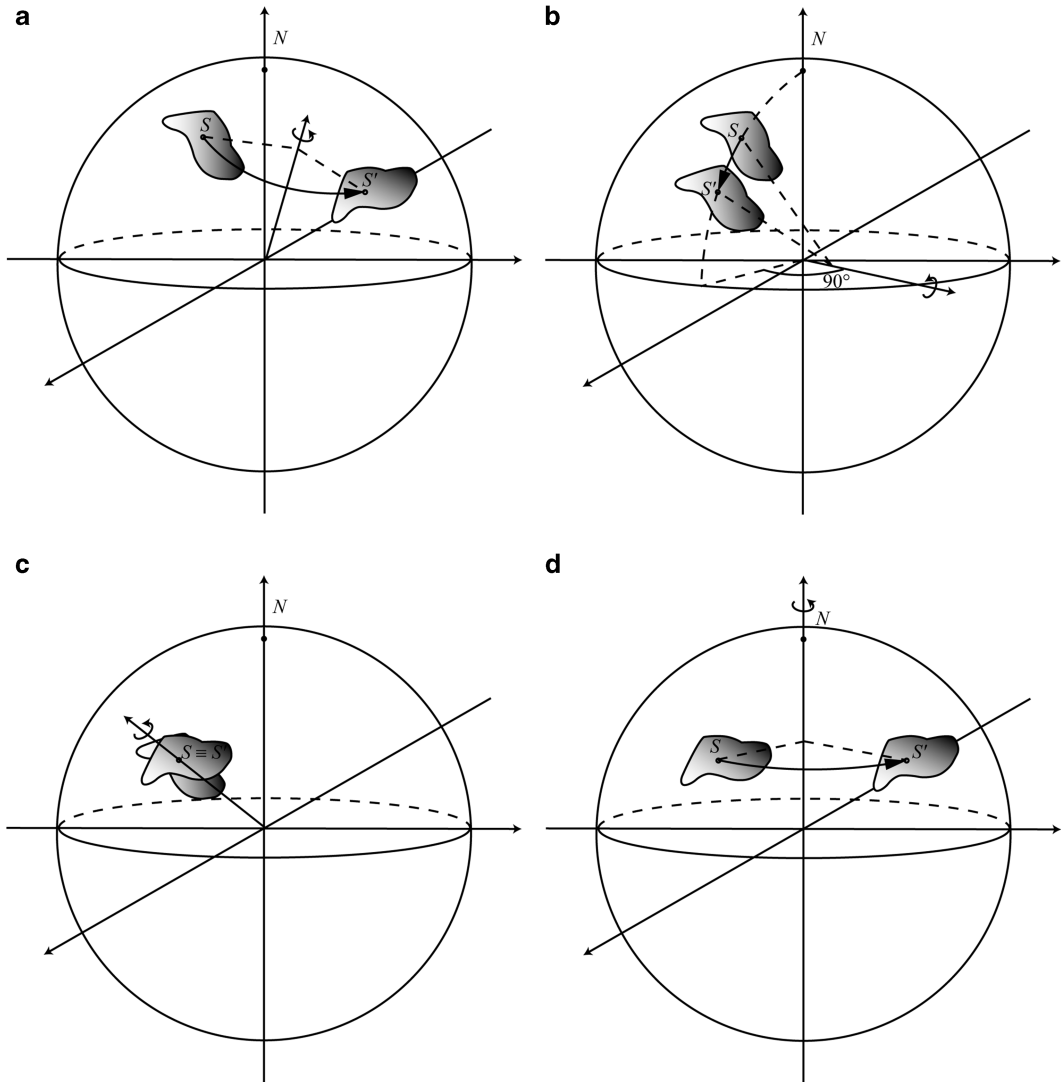
The importance of these variables was also stressed observing that the declination  $D$  and the paleolatitude  $\lambda$  at a reference site are kinematic quantities that completely describe the motion of a tectonic plate with respect to a paleomagnetic reference frame. In fact, it is easy to show that any motion on the sphere can be resolved into three elementary Euler rotations, namely: (a) A pure  $N$ – $S$  rotation, associated with an Equatorial Euler pole and variations of latitude of the reference site  $S$ ; (b) A pure rotation about the vertical axis at the reference site, which is responsible for changes in declination, and (c) A rotation about the spin axis ( $z$  axis), which only changes the site longitude. Figure 6.16 illustrates these independent rotations, which can be performed in any order without affecting the final result. Therefore, assuming a paleomagnetic reference frame where the longitude of the site  $S$  is fixed (see Sect. 2.3), we have that the pair  $(\lambda, D)$  completely describes the kinematics of the tectonic plate to which  $S$  belongs. Another reason to prefer declination and paleolatitude regressions sacrificing paleopole fitting is that declination and inclination (hence paleolatitude) are influ-

enced differently by tectonic processes after the acquisition of NRM, which also contributes to their *physical* independence. For example, the site of a paleopole could be located on a second-order tectonic element that experienced a small amount of vertical axis rotation with respect to the main continent. This process could affect significantly the site declination, although its paleolatitude would remain unchanged. On the contrary, a process known as sedimentary *inclination shallowing* (e.g., Arason and Levi 1990) could modify the magnetic inclination of deep-sea sediments during the process of compaction without affecting their declination. As a consequence, paleopoles having either a wrong declination or a wrong paleolatitude have a negative effect on a spherical regression curve, whereas they would be easily detected as outliers on either paleolatitude or declination plots. Therefore, an essential aspect of the approach of Schettino and Scotese (2005) was the detection of anomalous values of the predicted declination or paleolatitude at the reference sites. These outliers were filtered away during the process of construction of a smoothed paleolatitude or declination plot, although the corresponding paleopoles could still contribute to the other curve (respectively declination or paleolatitude). Finally, the resulting spline regression curves were combined to generate spherical smoothed (but not necessarily best fitting in the spherical sense) APW paths.

---

## 6.5 Paleomagnetic Reference Frames

In Chap. 2, when we have described the general structure and the process of construction of plate circuits, we omitted to specify how the finite rotations of the root plate are determined. We have learnt that rotation models are listings of finite reconstruction poles and rotation angles for each identified pair of conjugate plates. However, the finite rotations of the plate associated with the circuit *root* node cannot be referred to any other plate, so that in a rotation table like that illustrated in Fig. 2.29 the reference plate field of the corresponding entries will be set to



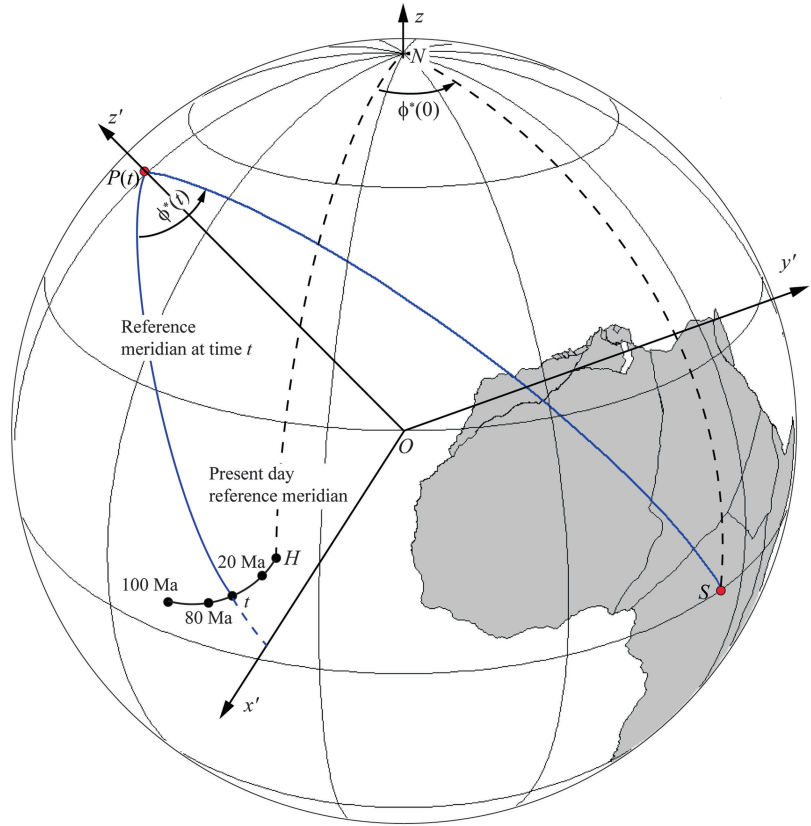
**Fig. 6.16** Decomposition of Euler rotations. Given a rotation on the sphere (a), an arbitrary reference site  $S$  will be subject to a variation of paleolatitude, declination, and longitude. These variations can be described by three

elementary transformations: a rotation about an Equatorial pole, located  $90^\circ$ E from the initial longitude (b), a vertical axis rotation (c), and a rotation about the spin axis (d)

zero. This means that the reconstruction matrices obtained from these entries are immediately *total* reconstruction matrices, which transform a plate from the present day position to the *absolute* position occupied at an assigned time  $T$  in a *paleotectonic* reference frame. Here, we are going to discuss some methods to set up a paleomagnetic frame of reference, which is a common kind of paleotectonic frame, using an APW path. In Chap. 2, we have seen that paleomagnetic frames are geocentric coordinate

systems such that the  $z$  – axis coincides with the *apparent* position of the Earth's spin axis as determined by a paleomagnetic pole. Therefore, these frames differ each other on the basis of how the longitude of a reference point on the root plate changes with respect to the present day value. Clearly, this definition requires a characterization of longitudes that is independent from plate motions, which leads us to abandon the familiar concept of angular distance from the Greenwich meridian.

**Fig. 6.17** The paleolongitude  $\phi^*(t)$  at time  $t$  of a site  $S$  on the reference continent (central Africa in this example) in a paleomagnetic reference frame  $(x', y', z')$  is built using the mean paleopole  $P(t)$  for time  $t$  and a hot spot track. At any time  $t$ , the reference meridian (i.e., the meridian for which  $x' = 0$ ) is defined as the actual meridian of a hot spot location,  $H(t)$ , with respect to  $P(t)$ . It is assumed that  $H$  is the surficial expression of deep processes and that its location does not depend from the motion of the plate where it is currently placed. The black line with age progression of hot spot locations represents the track left by  $H$  during the motion of this plate



A way to define the *paleolongitude* of a point independently from plate motions is based on the individuation of *hot spots* (Morgan 1971) punctual magmatic features at the Earth's surface, which are expressions of deep processes occurring somewhere below the lithosphere and the uppermost asthenosphere (Fig. 6.17). A key feature of hot spots, which is generally interpreted as evidence of their stationary position with respect to the mantle, is that they are placed at the youngest ends of *alignments* of intraplate volcanism with *age progression*. In this instance, the combined sequence of magmatic events in space and time could witness the transit of the tectonic plates where they are placed over quasi-stationary mantle upwellings not related to plate boundary processes. Classic examples of *hot spot tracks* are the Hawaiian and Emperor seamount chains on the Pacific plate. Contrarily to a widespread belief, the construction of reference frames based on hot spot tracks does not require assumption

that hot spots be fixed to the deep mantle, because *any* marker whose geographic location is unconstrained with respect to the surrounding plate and with sufficient age range can be used to define paleolongitudes. Figure 6.17 illustrates how to accomplish this task. If  $H$  is the present day location of a hot spot, we choose an arbitrary reference site  $S$  on the root plate and determine its relative longitude,  $\phi^*(0) = \phi_S - \phi_H$ , with respect to  $H$ . To calculate the paleolongitude at time  $t$ , we choose as reference meridian in the paleomagnetic reference frame  $(x', y', z')$  the meridian passing through the location,  $H(t)$ , where the hot spot magmatism has age  $t$ . The paleolongitude  $\phi^*(t)$  will be given by the angle between the meridian through  $S$  and the reference meridian *in the paleomagnetic frame*. Therefore, the *longitudinal shift* of the reconstructed site  $S$  is given by  $\delta\phi(t) = \phi^*(t) - \phi^*(0)$ .

Hot spot tracks are useful to set up paleomagnetic or pure hot-spot reference frames for

Cenozoic and late Mesozoic plate reconstructions (e.g., Müller et al. 1993). However, the scarcity of hot spot data for times older than the Cretaceous leads to adopt alternative strategies when the reconstructions encompass larger time intervals. The simplest way to define a paleomagnetic coordinate system independently from hot spot tracks is to assign constant paleolongitude  $\phi^*(t) = \phi_0$  to a selected reference site,  $\phi_0$  being the present day longitude of the site. For example, if we choose the central African craton as the root continent of our rotation model, then we could assume that the point with present day coordinates (0°N, 25°E) in central Africa will have constant paleolongitude  $\phi^*(t) = 25^\circ\text{E}$  for any time  $t$ . This approach can be found in Besse and Courtillot (1988). In this instance, the finite rotations corresponding to each paleopole position are obtained applying the transformation:

$$\mathbf{R}(t) = \mathbf{R}(0, \phi_0 + 90, \lambda_0 - \lambda(t)) \mathbf{R}(\lambda_0, \phi_0, D(t)) \quad (6.58)$$

Here,  $D(t)$  and  $\lambda(t)$  are respectively the declination of the reference point  $(\lambda_0, \phi_0)$  and its paleolatitude at time  $t$ , and  $\mathbf{R}(\lambda, \phi, \omega)$  is a rotation matrix about the generic Euler pole  $(\lambda, \phi)$  by an angle  $\omega$ . In this formula,  $D(t)$  and  $\lambda(t)$  are determined by the paleopole position using the method described in Sect. 6.4. We note that the first rotation in (6.58) resets to zero the declination at  $(\lambda_0, \phi_0)$ , while the subsequent rotation moves this point northward or southward to restore its paleolatitude. Both these rotations leave invariant the longitude  $\phi_0$  of the reference point. An alternative, but substantially equivalent, way to set up a paleomagnetic reference frame is to define  $\mathbf{R}(t)$  as the rotation that brings a paleopole  $P$  to the North Pole through its own meridian of longitude, as illustrated in Fig. 6.7. This rotation is given by:

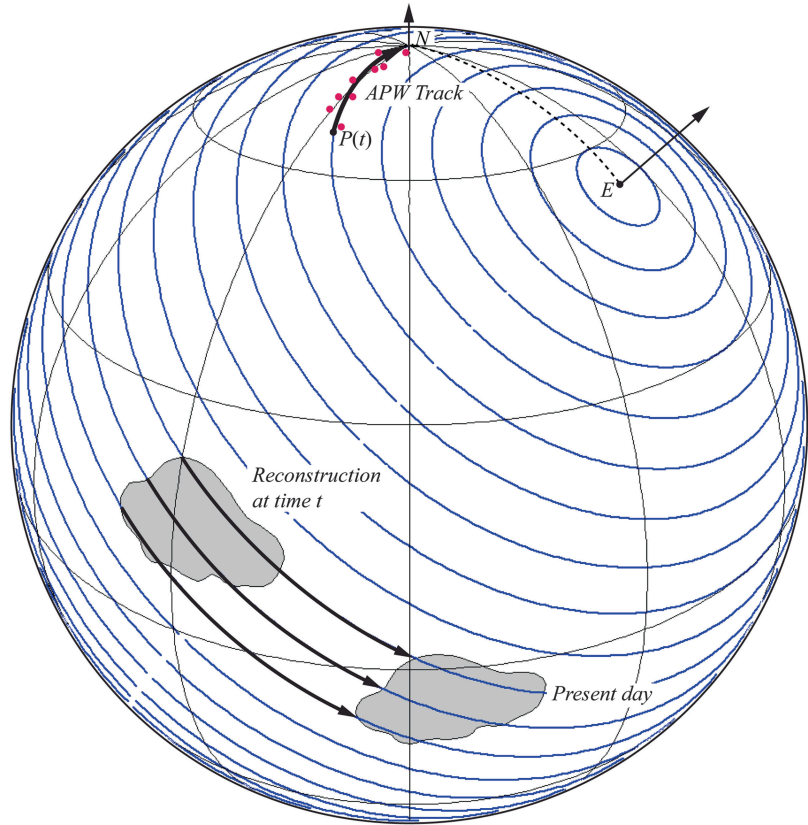
$$\mathbf{R}(t) = \mathbf{R}(0, \phi_P + 90^\circ, -\theta_P) \quad (6.59)$$

where  $(\phi_P, \theta_P)$  are the paleopole longitude and colatitude. Two examples of the use of such a reference frame can be found in Ziegler et al. (1983) and Schettino and Scotese (2005).

Note that the setup of this frame does not require the selection of a reference site. Furthermore, in this case the paleolongitude of any reference point will change with respect to its present day value in a complicated way, although the variation will be very small. In general, any rotation that brings a paleopole  $P$  to the North Pole generates a corresponding paleomagnetic reference frame.

Maybe, the most intriguing way to perform this operation is that proposed by Gordon et al. (1984). These authors observed that for long time intervals plate motions appear to be rotations about fixed Euler poles, so that any point on a tectonic plate moves along a small circle arc. Even the APW paths seemed to be composed by sequences of *tracks* having the geometry of small circle arcs, which were interpreted as the kinematic expression of stages of constant motion about fixed Euler poles. These poles were called *paleomagnetic Euler poles* (PEPs). In this view, the *cusps* of an APW path, which link consecutive small circles, correspond to times of rapid change in the plate motion. An example of PEP analysis is illustrated in Fig. 6.18, in the case of a track ending at the North Pole. In general, the transformation from the present day geographic coordinate system to a paleomagnetic reference frame is accomplished through  $k$  rotations about PEPs if a paleopole is placed along the  $k$ -th track. Apparently, the technique proposed by Gordon et al. (1984) allows to determine the true *paleolongitude variation* of a reference site  $S$  with respect to its present day value. To this purpose, these authors drew an analogy with the motion of a tectonic plate with respect to hot spots. They observed that alignments of volcanic activity with regular age progression, such as the Hawaiian and Emperor seamount chains, have the geometry of small circle arcs. Therefore, in this case we can find Euler poles by fitting small circles to hot spot tracks. If we assume that the hot spots are fixed with respect to the lower mantle, then a rotation about a hot spot Euler pole can be considered as a motion expressed in a lower mantle reference frame. By analogy, a rotation about a PEP would be expressed in a *unique* “absolute” reference frame. However, Gordon et al. (1984) did not

**Fig. 6.18** Reconstruction of a continent through PEP analysis. A small circle arc about a PEP,  $E$ , is used to fit a sequence of paleopoles (red dots). For a track that ends at the North Pole, a paleomagnetic reference frame for time  $t$  is found rotating  $P(t)$  to the North pole about the PEP  $E$



explain how to set up such an “absolute” frame. Therefore, although PEP analysis furnishes an intriguing way to represent the motion of a root continent, it has not received much favor from the community of paleomagnetists (e.g., Butler 1992; Van der Voo 1993), possibly as a consequence of the difficulties to define precisely the reference frame.

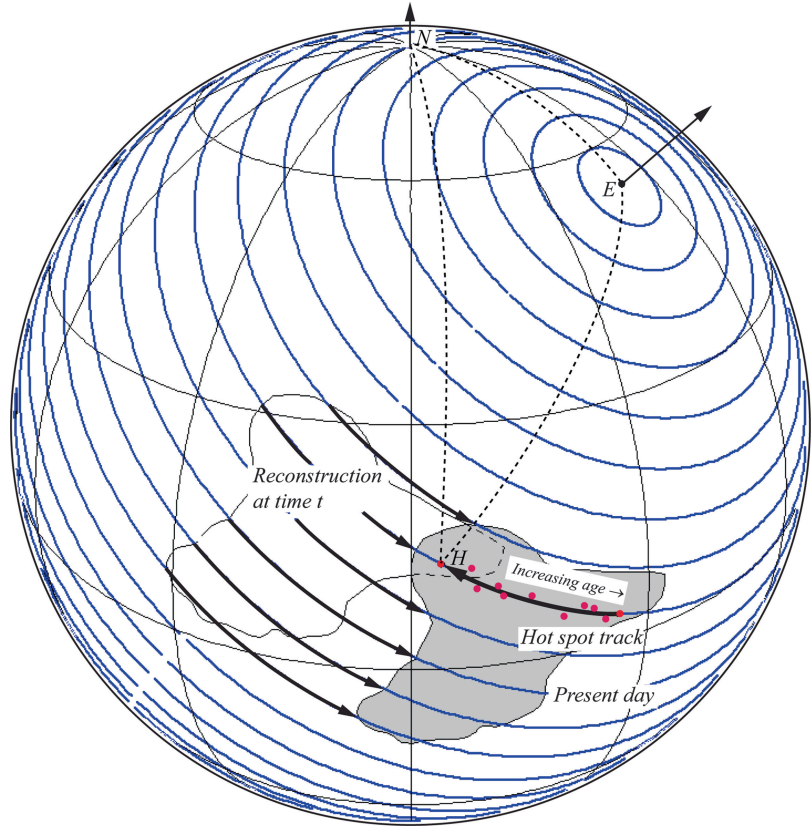
## 6.6 True Polar Wander

*True polar wander* (TPW) can be defined as a rotation of the whole outer shells of the solid Earth (crust plus mantle) with respect to the spin axis, although this rotation would be indistinguishable from a motion of the lone lithosphere. Such definition, which can be found in several papers about this subject (e.g., Jurdy and Van der Voo 1974; Gordon 1987; Evans 2003), requires a precise definition of what is intended by “rotation of the whole Earth”, and a specification

of the reference frame where the rotation can be observed. Unfortunately, the definitions that can be found in the published literature are often vague or inadequate. For example, Steinberger and Torsvik (2008) built a reference frame assuming that the  $z$ -axis was aligned with the spin axis, while the  $y$ -axis was chosen in the Equatorial plane, passing through the meridian where the “continents’ centre of mass” is placed. Apart from the difficulty to find the “continents’ centre of mass”, this definition is somewhat confusing, because the centre of mass of a system of plates (or continents) is not independent from global plate motions, thereby, it should not be used to set up a reference frame for the detection of TPW. As shown in the previous section, a suitable way to define a coordinate system that is independent from plate motions requires using hot spot tracks, which are expressions of processes occurring somewhere below the lithosphere.

For example, if a hot spot track exists on the root plate, we can eventually fit a small circle

**Fig. 6.19** Determination of a hot-spot Euler pole (HEP). In this instance, a small circle arc about a HEP,  $E$ , is used to fit a chain of volcanoes (red dots).  $H$  is the location of present day hot spot volcanism



arc about a *hot spot Euler pole* (HEP) to the hot spot track observed on this plate (Gordon et al. 1984). This method is illustrated in Fig. 6.19. In this instance, for a set of  $N$  plates related by a circuit  $C$ , we can determine the Euler vectors  $\omega_i$  describing the motion of each plate in the global circuit. In fact, if  $\omega_r$  is the Euler vector of the reference plate (with Euler pole in  $E$ , see Fig. 6.19) and  $\omega_{ir}$  is the angular velocity of the  $i$ -th plate relative to the root plate, then its absolute angular velocity will be given by:  $\omega_i = \omega_{ir} + \omega_r$ . Then, using (2.69) we see that an Euler vector describing the rotation of the whole lithosphere can be written as follows:

$$\Omega = \frac{3}{8\pi} \sum_i Q^i \omega_i \quad (6.60)$$

where the components of the tensors  $Q^i$  are given by (2.63). The vector  $\Omega$  can be used to establish the existence of TPW. To this purpose,  $\Omega$  must be decomposed into an Equatorial component and a

component aligned with the spin axis. As noted by Jurdy and Van der Voo (1974), only the Equatorial component will contribute to TPW, in so far as it shifts the whole lithosphere from the current location of the spin axis. Therefore, we can define TPW as the result of a rotation of the whole lithosphere about an *equatorial* Euler pole. Given that it is unlikely for  $\Omega$  to be aligned with the spin axis at any time, we see that TPW is a *common* feature of global plate motions, which also affects at the same time *all* APW paths. A series of questions immediately follows (apart from those regarding the dynamics associated with TPW). For example, what is the amount of TPW during the geologic past? Does it average to zero? Second, how are APW paths affected by TPW? Third, what is the kinematic relation between TPW and hot spots? Although these questions have been addressed in a number of articles since the 1970s, no satisfactory answer has been found so far and TPW is still at the frontier of paleomagnetic research.

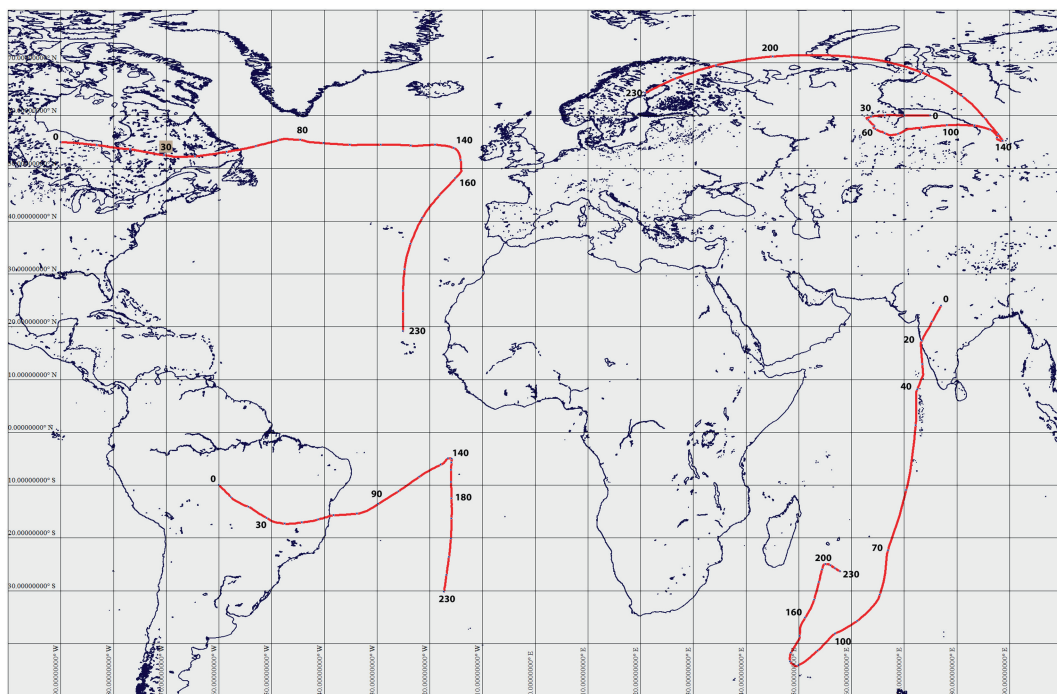
## 6.7 Velocity Fields and Acceleration Fields During the Cenozoic and the Mesozoic

Now we are going to describe some insightful features of global plate motions since the late Triassic through an unconventional approach, based on an analysis of velocity and acceleration fields. To this end, we will use the techniques described in Sect. 2.8 to determine the absolute linear velocity of a point (Eq. 2.51) and its acceleration. For simplicity, plate motions will be represented in the paleomagnetic reference frame of Schettino and Scotese (2005), but the validity of the results extends to more sophisticated coordinate systems. Our starting point is the observation that the absolute motion of a reference point on any given plate, which can be determined using the methods described in Sect. 2.7, is far from being a random walk. Figure 6.20 illustrates the absolute

trajectories of reference sites on four continents since 230 Ma (late Ladinian). An important feature that these paths have in common is that *they are chains of small circle arcs*. Furthermore, it is possible to show that the velocity along each small circle arc is approximately constant, which suggests us that the motion of any point can be described by a sequence of rotations about fixed Euler axes.

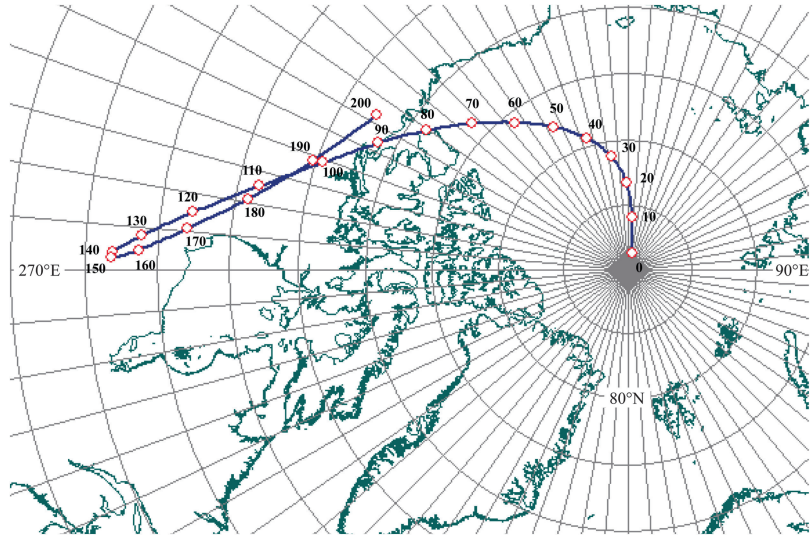
It is not difficult to prove that this is in part a consequence of the fact that the relative motion between two neighbor plates in a circuit is always a sequence of stage rotations (see Sect. 2.7). The technique to calculate the path of relative motion travelled by a reference point has been described in Sect. 2.8. Extending the method to absolute motions is an elementary task: we simply substitute the relative linear velocity appearing in (2.56), which could not be associated with a stage rotation, by an absolute velocity  $\mathbf{v}$ :

$$\mathbf{r}(T + \delta T) = \mathbf{r}(T) + \mathbf{v}(\mathbf{r}) \delta T \quad (6.61)$$



**Fig. 6.20** Trajectories of reference points in N. America, S. America, Eurasia, and India during the last 230 Myrs in the paleomagnetic reference frame of Schettino and Scotese (2005). Numbers represent ages along the paths

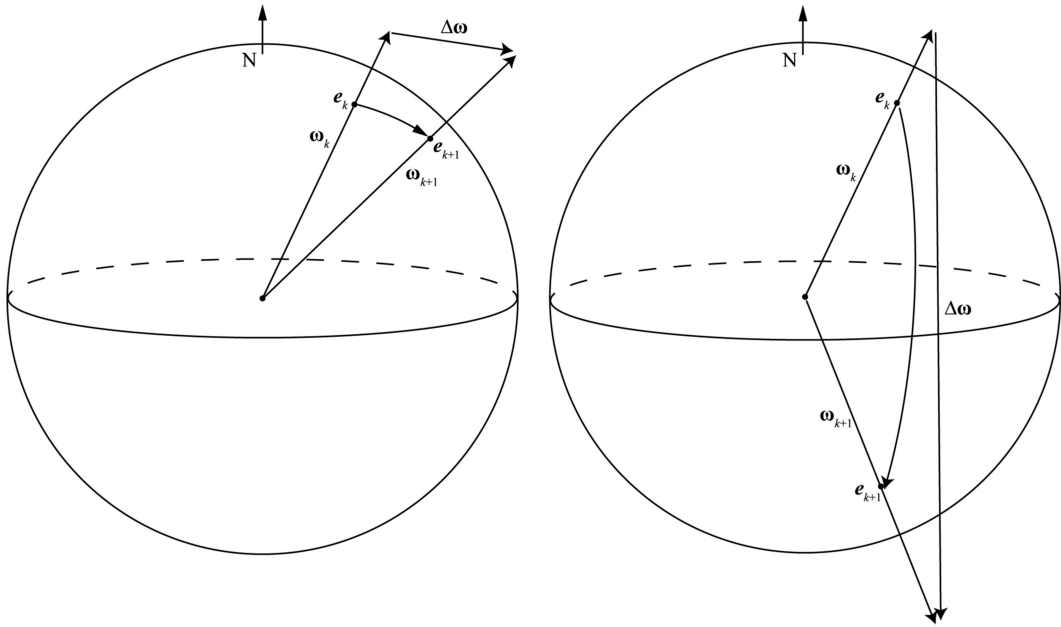
**Fig. 6.21** Smoothed APW path of central Africa since the early Jurassic (modified from Schettino and Scotese 2005). Numbers represent ages along the path



where  $T$  is the time and  $\mathbf{r}$  is the position vector of the reference point. To determine the absolute velocity  $\mathbf{v}$ , we perform a traversal of a rotation tree from the selected plate to the root node, calculating and summing at each step the relative velocity of the current node with respect to its parent (Eq. 2.51). However, we know that any node in a rotation tree represents a *conjugate* plate for the parent node, thereby, the velocities are always calculated from stage poles. Consequently, the relative motion of any plate with respect to the root node can be represented by a sequence of stage rotations. This is not sufficient to explain the geometry of motion illustrated in Fig. 6.20, because the absolute motion of the root plate could be a random sequence of instantaneous rotations, at least in principle. However, as we have seen in section Sect. 6.5, paleomagnetic data suggest that even the motion of the root plate can be described by a sequence of rotations about PEPs. This observation results not only from the study of Gordon et al. (1984) but also from an independent analysis of the geometry of the smoothed APW paths proposed by Schettino and Scotese (2005). For example, it can be proved rigorously that the smoothed African APW path can be divided into four small circle tracks, as suggested by a visual inspection of Fig. 6.21 (Schettino and Scotese 2000).

Therefore, we infer that the absolute motion of any plate can be described by a sequence of rotations about fixed Euler poles, as illustrated in Fig. 6.20 by the paths travelled by reference sites. Such conclusion has important consequences when we try to link plate kinematics with mantle geodynamics, because it implies that plate motions proceed for long time intervals with constant angular momentum, so that the total torque exerted on a tectonic plate is zero. This in turn implies that in normal conditions the lithosphere moves in equilibrium conditions, such that the resistive forces opposing the motion over the fluid asthenosphere are balanced by active plate boundary forces. More specifically, the existence of stationary plate motions that are maintained for long time intervals implies that the dominant driving force represented by the gravitational pull of subducting slabs is always balanced by the resistive viscous drag exerted by the asthenosphere (Chase 1979). However, in Chap. 13 we shall see that episodes of accelerated plate motion are possible and result from currents in the asthenosphere, anomalous ridge push, or continental collisions.

Now we are going to give a precise characterization of the global tectonic stages since the late Triassic, as well as of possible events of accelerated motion during this time interval. In general,



**Fig. 6.22** Accelerations at stage transitions.  $e_k$  and  $e_{k+1}$  are two successive stage poles. Euler vectors  $\omega_k$  and  $\omega_{k+1}$  are built using the versors through  $e_k$  and  $e_{k+1}$  and the angular velocities of the corresponding stages. Although

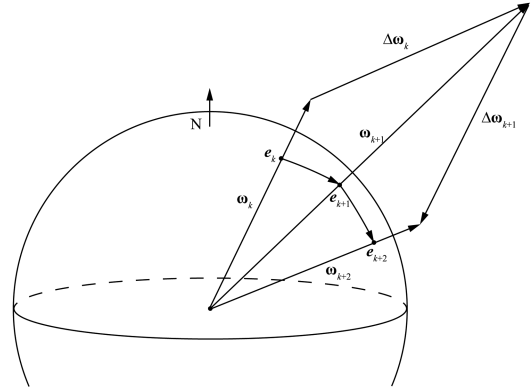
the variation  $\Delta\omega$  can be quite large, the magnitude of the angular velocity variation,  $\Delta|\omega| = |\omega_{k+1}| - |\omega_k|$  is generally small

stage transitions are associated with discontinuous variations of the stage pole about which a tectonic plate moves, so that we can define an *angular acceleration* vector  $\Delta\omega \equiv \omega_{k+1} - \omega_k$  at the end of the  $k$ -stage (Fig. 6.22). Just as in the example of Fig. 6.22, these transitions do not generally involve large variations of plate velocity, because in most cases we have only a change in the location of the stage pole, such that  $\Delta|\omega| \ll |\Delta\omega|$ . In principle, phases of accelerated plate motion should be associated with a series of very short stages with increasing (or decreasing) velocity. In practice, the analysis of marine magnetic anomalies often allows to determine only an *average* stage velocity, whereas the details of the spreading process are obscured. The reason is that stage boundaries are usually associated with sharp variations of curvature in the trend of oceanic fracture zones (corresponding to changes of stage pole location), only exceptionally with variations of plate velocity. Therefore, we will be able to identify phases of accelerated motion, which must not be confused with stages

of elevated but constant velocity, only by the individuation of *short asynchronous pulses* of fast (or slow) velocity of individual plates, not necessarily associated with plate boundary processes (e.g., collisions). In this case, we will observe a significant positive (or negative) peak in the magnitude of the angular acceleration of a *single* plate at a stage boundary, followed by a comparable peak of opposite sign at the successive stage boundary (Fig. 6.23).

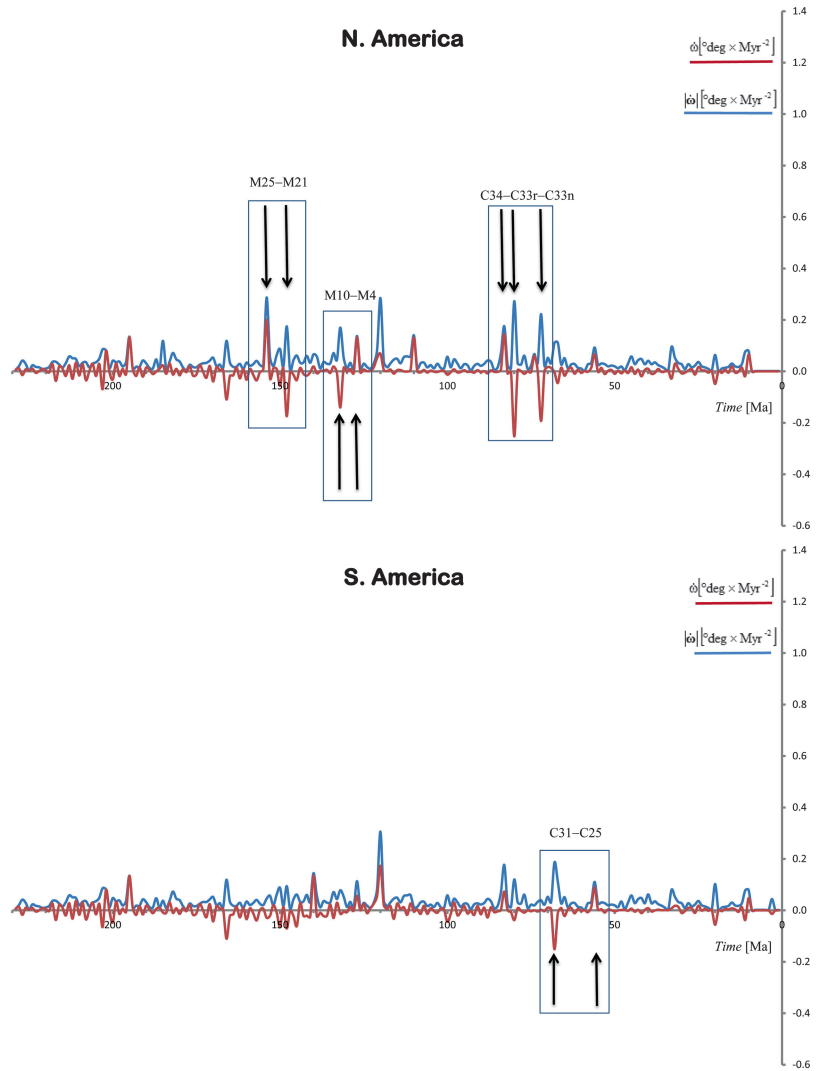
Plots of magnitude of the angular acceleration vector,  $|\dot{\omega}|$ , and of variations of angular velocity,  $\dot{\omega}$ , for nine large plates since the late Triassic are illustrated in Figs. 6.24, 6.25, 6.26, 6.27, 6.28, 6.29, 6.30, 6.31, and 6.32. Peaks in the magnitude of the angular acceleration vector are associated with large displacements of the stage pole during stage transitions, but not necessarily with non-equilibrium states. Conversely, alternate pairs of large peaks in the plots of scalar acceleration are indicative of non-equilibrium states of *traction* for the lithosphere, if they represent asynchronous events. For example, a striking peak of

**Fig. 6.23** Detection of accelerated non-equilibrium states. The transitions from stage pole  $e_k$  to  $e_{k+1}$  and from  $e_{k+1}$  to  $e_{k+2}$  are associated with large variations, of opposite sign, of the magnitude of the angular velocity



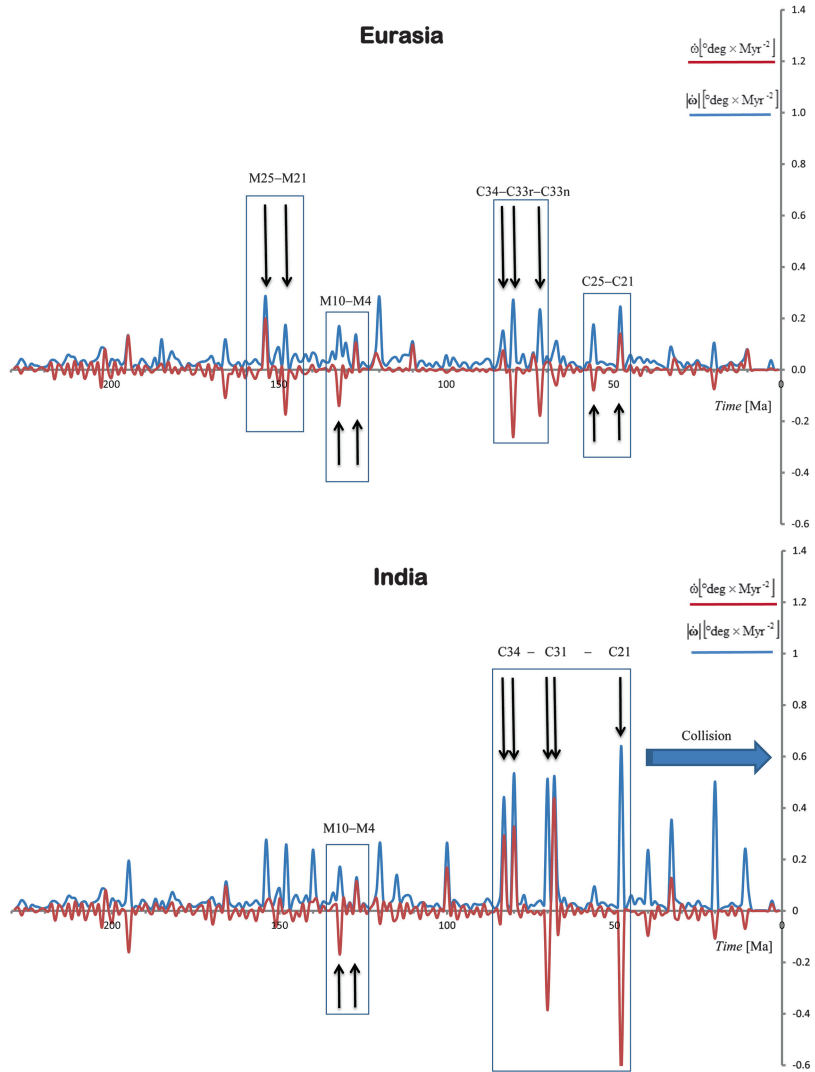
**Figs. 6.24 and 6.25**

Magnitude of the angular acceleration vector (*blue line*) and variations of angular velocity (*red line*) for N. America and S. America since the late Triassic (230 Ma). Boxes indicate possible events of traction



**Figs. 6.26 and 6.27**

Magnitude of the angular acceleration vector (*blue line*) and variations of angular velocity (*red line*) for Eurasia and India since the late Triassic (230 Ma). Boxes indicate possible events of traction



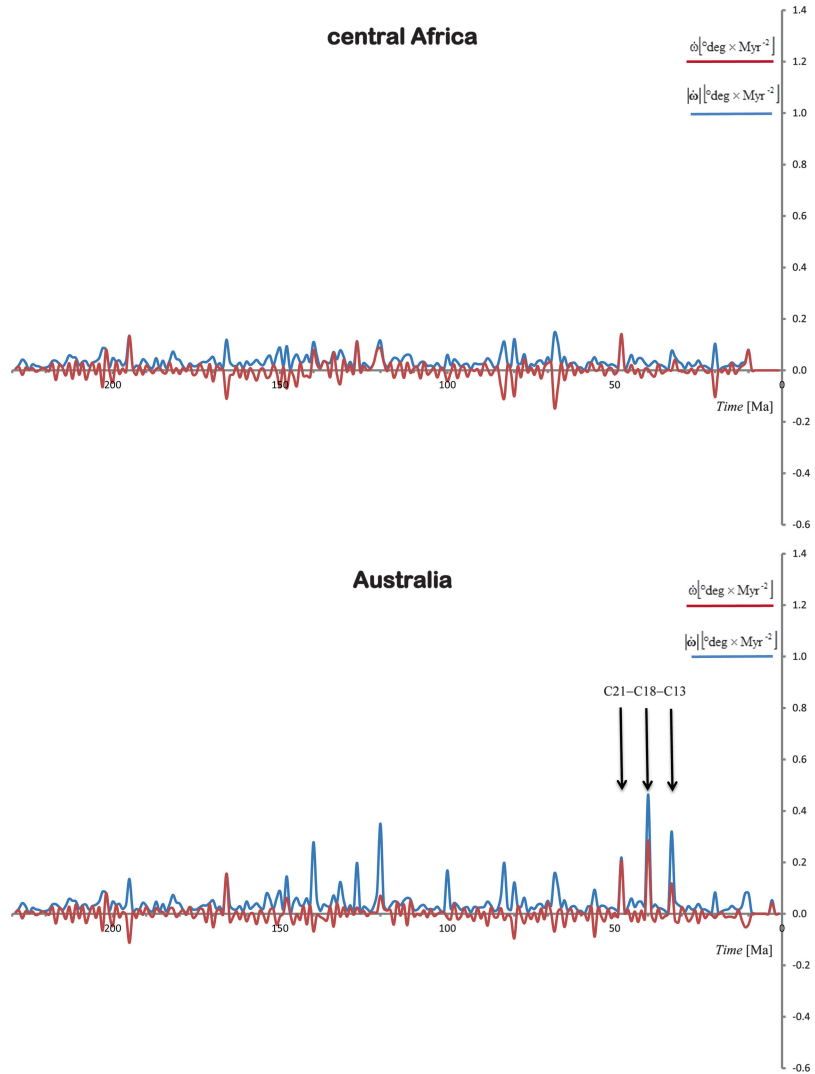
$|d\omega/dt|$  can be observed at 120 Ma (anomaly M0) in the plots of N. America, S. America, Eurasia, India, Australia, and E. Antarctica.

This peak is associated with a phase of global reorganization of the plate boundaries during the early Aptian, but no events of traction can be observed at that age. However, large peaks of scalar acceleration are always accompanied by peaks in the magnitude of the angular acceleration vector. The plots in Figs. 6.24, 6.25, 6.26, 6.27, 6.28, 6.29, 6.30, 6.31, and 6.32 show that central Africa, east Antarctica and possibly Australia never experienced a remarkable episode of traction, although the latter was subject to three

strong increments of angular velocity during the Tertiary that could be indicative of ongoing accelerated motion (Figs. 6.28, 6.29). The oldest event of lithospheric traction during the time interval considered here possibly occurred between 154 Ma (M25 – Kimmeridgian) and 148 Ma (M21 – Tithonian). It consisted into a clockwise acceleration of Laurasia about a pole located in the present day Xinjiang region, central Asia ( $44.88^{\circ}\text{N}$ ,  $88.45^{\circ}\text{E}$ ). The event can be observed on the plots of both N. America and Eurasia (Figs. 6.24, 6.25, 6.26, 6.27), which at that time formed a unique Laurasian plate, and coincided with a phase of accelerated spreading

**Figs. 6.28 and 6.29**

Magnitude of the angular acceleration vector (*blue line*) and variations of angular velocity (*red line*) for central Africa and Australia since the late Triassic (230 Ma)



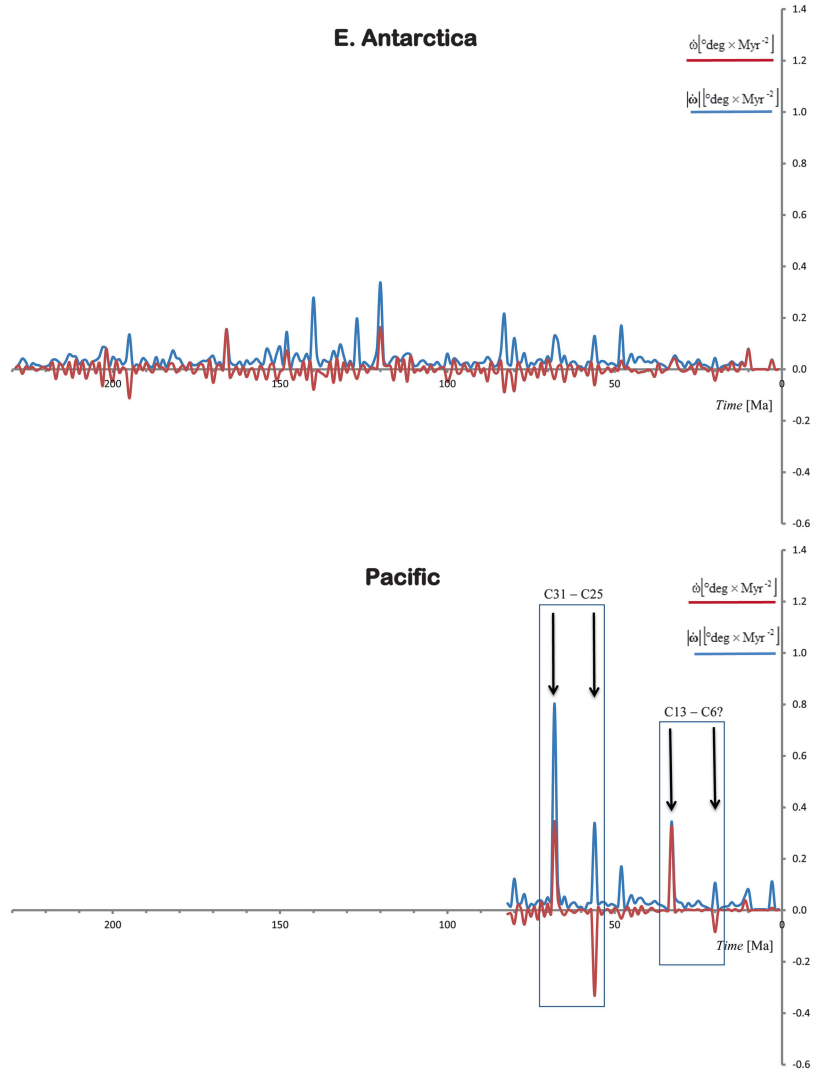
in the central Atlantic. It is interesting to note that the time interval of this event does not overlap with the first episode of magmatism in the New England seamounts chain, which is  $\sim 103$  Ma (Duncan 1984). Therefore, it is not possible in this case to associate the lithospheric traction with mantle plume magmatism. A paleotectonic reconstruction at 154 Ma (M25 – Kimmeridgian) is shown in Fig. 6.33, which illustrates both the instantaneous velocity field of the major continents at that time and the  $\sim 0.20^{\circ}\text{Myr}^{-2}$  clockwise acceleration field of Laurasia. Undoubtedly, this event could have triggered a phase of increased subduction in

the Tethys, associated with the obduction of the Vardar ophiolites onto the Dinaric, Pannonian, and Menderes-Tauride domains of east Europe and Turkey (Schettino and Turco 2011).

Apart from the traction event that affected Laurasia, the global velocity field shown in Fig. 6.33 suggests a phase of relevant TPW, associated with a clockwise rotation of the lithosphere about an Equatorial pole. However, this is only a possibility, which cannot be demonstrated merely through the observation of continental lithosphere velocity fields. The next probable episode of traction was a slowdown of Laurasia and India between

**Figs. 6.30 and 6.31**

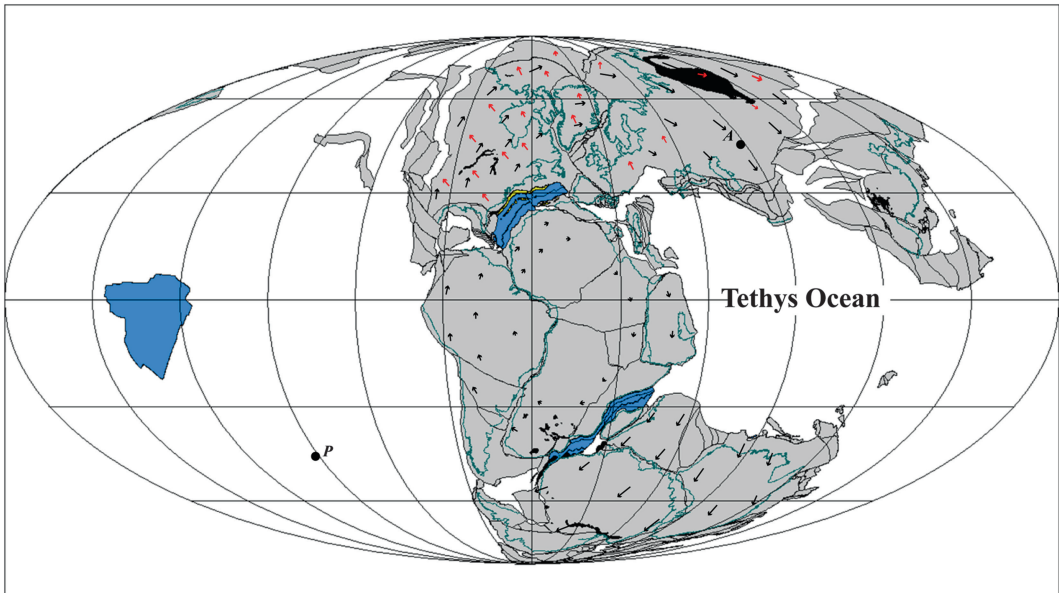
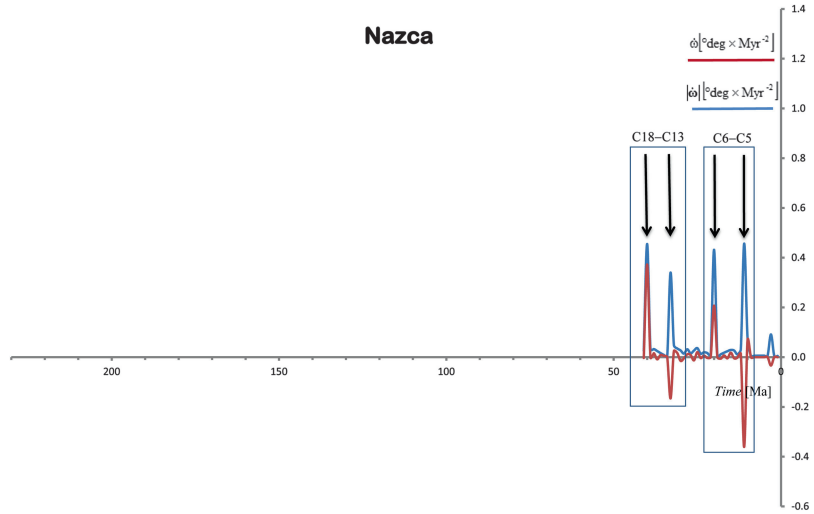
Magnitude of the angular acceleration vector (*blue line*) and variations of angular velocity (*red line*) for east Antarctica and Pacific since the late Triassic (230 Ma). *Boxes* indicate possible events of traction



132 Ma (M10 – Hauterivian) and 127 Ma (M4 – Barremian) (Figs. 6.24, 6.25, 6.26, 6.27). However, in this instance lower magnitudes of angular acceleration (respectively  $-0.14$  and  $-0.17^{\circ}\text{Myr}^{-2}$ ) were accompanied by locations of either poles or antipoles of acceleration not far from these continents. Consequently, the effective accelerations were low and there is more than one doubt that this time interval was effectively a phase of accelerated motion for these plates. Figure 6.34 shows a reconstruction at 132 Ma (M10 – Hauterivian). It is interesting to note that in this reconstruction both S. America and Africa have negligible angular angular velocities.

The successive episode of non-equilibrium kinematics occurred at  $\sim 84$  Ma (C34–C33 boundary, Santonian) and influenced once again the motion of N. America, Eurasia, and India. However, while the episode of accelerated motion ended during C33n ( $\sim 78$  Ma – Campanian) in the case of N. America and Eurasia (Figs. 6.24, 6.25, 6.26, 6.27), India continued to increase its velocity until C21 ( $\sim 48$  Ma – Lutetian) (Figs. 6.26, 6.27). Furthermore, while the former plates experienced modest increases of angular velocity, respectively  $0.14$  and  $0.07^{\circ}\text{Myr}^{-2}$ , India accelerated northward by  $0.29$ ,  $0.33$ , and  $0.44^{\circ}\text{Myr}^{-2}$  during three distinct pulses between

**Fig. 6.32** Magnitude of the angular acceleration vector (*blue line*) and variations of angular velocity (*red line*) for Nazca since the late Triassic (230 Ma). *Boxes* indicate possible events of traction

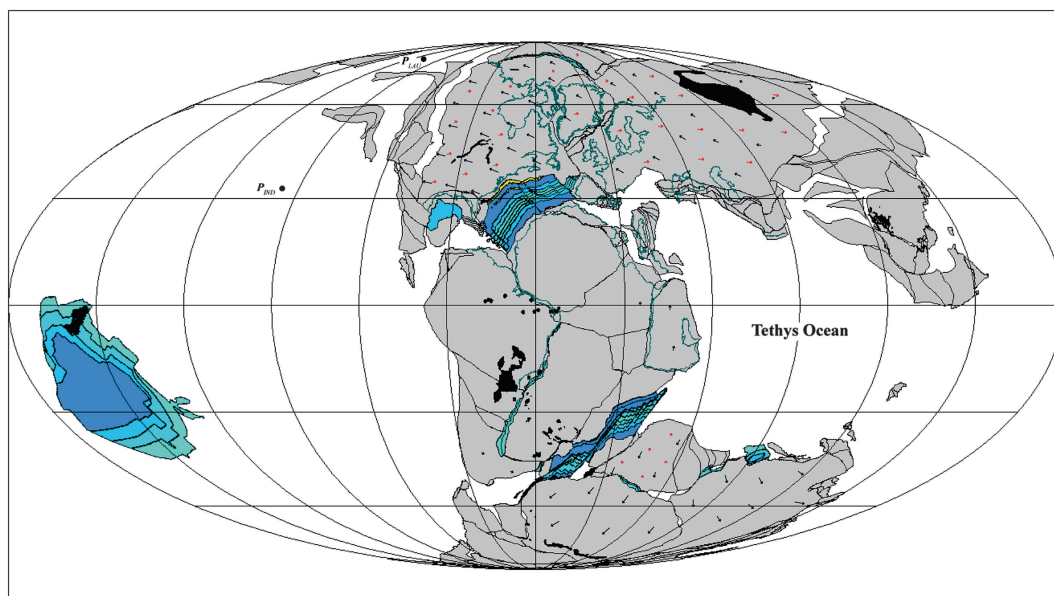


**Fig. 6.33** Reconstruction at 154 Ma (M25 – Kimmeridgian), illustrating the acceleration field of Laurasia (*red arrows*) and the global plate velocities (*black arrows*).

*A* and *P* are respectively the antipole and the pole of acceleration. *Black* areas are LIPS. Coastlines are overlaid for reference

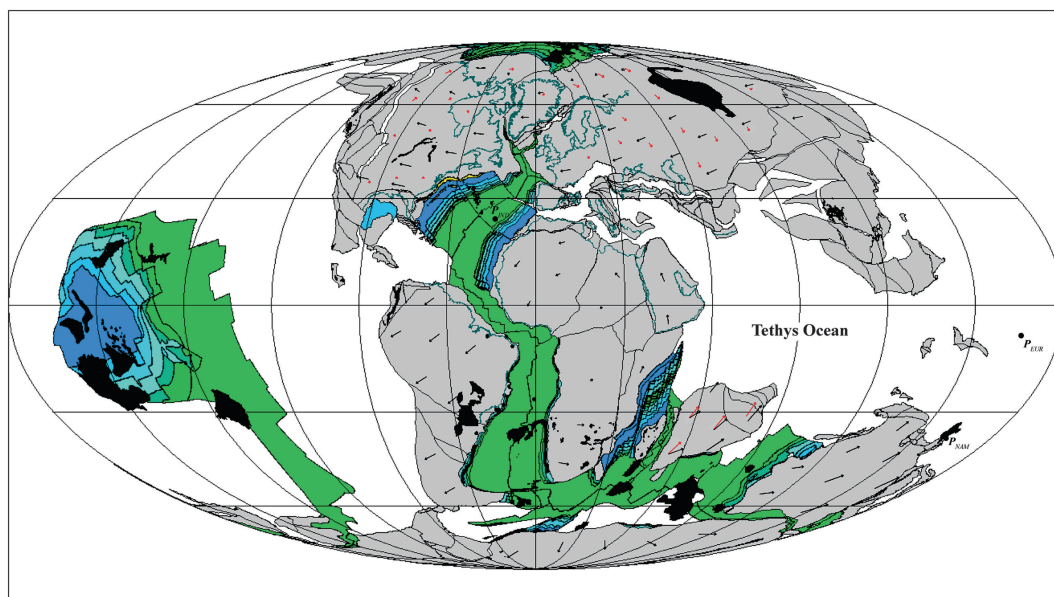
C34 and C21. Cande and Stegman (2011) have suggested that the northward acceleration of India during the Cretaceous to Eocene time interval was determined by the push exerted by the Reunion plume head. Conversely, no clear connection exists between the traction of N. America – Eurasia and plume magmatism. Figure 6.35 shows a reconstruction at 84 Ma (M10 – Santonian), where we can observe a

possible episode of TPW, associated with a counterclockwise rotation of the lithosphere about an Equatorial pole. A remarkable feature of this reconstruction is the considerable *linear* acceleration of N. America and Eurasia, despite their modest increase of *angular* velocity. This depends from the location of the acceleration poles or antipoles, which are both far from these plates. This is an important point: *the location*



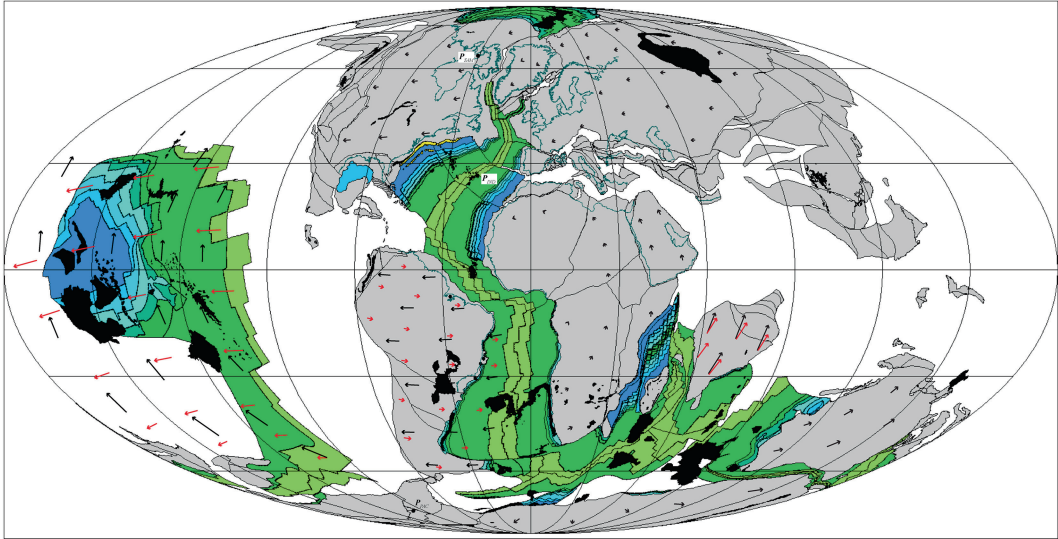
**Fig. 6.34** Reconstruction at 132 Ma (M10 – Hauterivian), illustrating the small acceleration field of Laurasia and India (red arrows) and the global plate velocities

(black arrows).  $P_{IND}$  and  $P_{LAU}$  are respectively the poles of acceleration of India and Laurasia. Black areas are LIPS. Coastlines are overlaid for reference



**Fig. 6.35** Reconstruction at 84 Ma (C34 – Santonian), illustrating the acceleration field of N. America, Eurasia, and India (red arrows) and the global plate velocities

(black arrows).  $P_{IND}$ ,  $P_{EUR}$ , and  $P_{NAM}$  are respectively the poles of acceleration of India, Eurasia, and N. America. Black areas are LIPS. Coastlines are overlaid for reference



**Fig. 6.36** Reconstruction at 68 Ma (C31 – Maastrichtian), illustrating the acceleration field of the Pacific, S. American, and Indian plates (*red arrows*) and the global plate velocities (*black arrows*).  $P_{IND}$ ,  $P_{PAC}$ , and  $P_{SAM}$  are

respectively the poles of acceleration of India, Pacific, and S. America. *Black* areas are LIPS. Coastlines are overlaid for reference

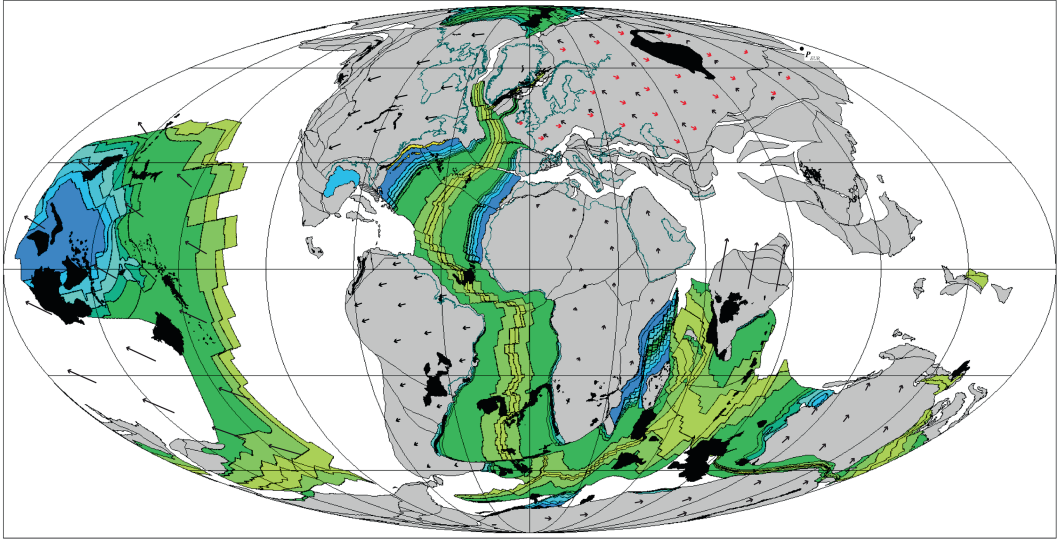
of a pole of acceleration is often more important than its magnitude in the recognition of traction episodes.

Figure 6.35 also shows that S. America was moving rapidly southwestward during the Santonian, while the southern part of central Africa was moving eastward at moderate velocity. These motions followed an episode of almost instantaneous increase of angular velocity during the very short chron M0 (~125 Ma), which was accompanied by a relevant change of the stage pole locations. It is not possible to prove that this was an event of traction, because we do not know the details of the spreading process during the long CQZ, although it is interesting to note that the spike of  $\dot{\omega}$  coincided with the onset of the Tristan de Cunha plume magmatism in the South Atlantic.

A new set of traction episodes occurred between chrons C31 (~68 Ma – Maastrichtian) and C25 (~57 Ma – Thanetian). Apart from the continued northward acceleration of India, during this stage both S. America and the Pacific plate experienced traction, but with opposite polarities. In fact, while S. America was *decelerating* in this time interval (Figs. 6.24,

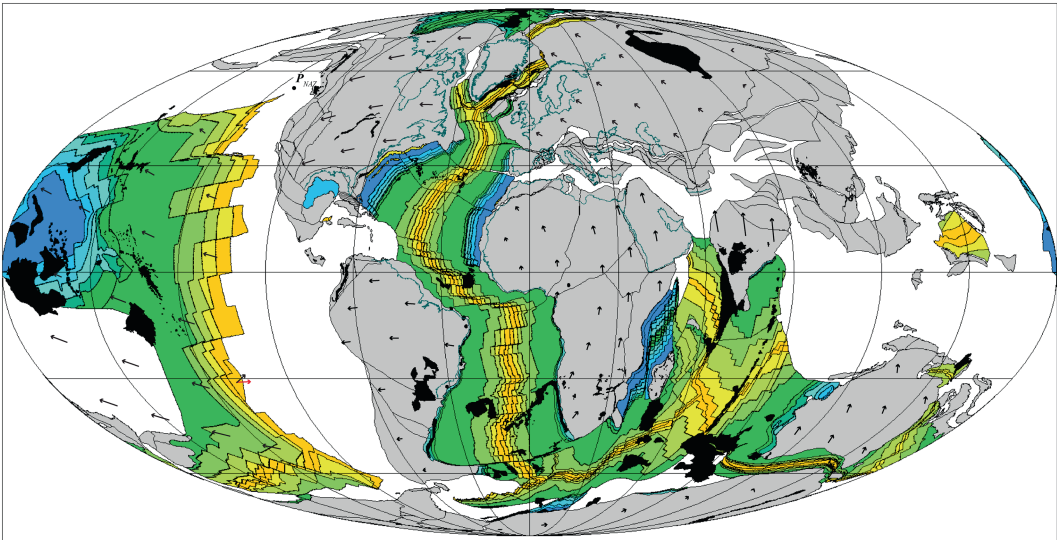
6.25), the Pacific plate *increased* its angular velocity and changed dramatically the stage pole location (Figs. 6.30, 6.31). A reconstruction at ~68 Ma is shown in Fig. 6.36, which illustrates the strong eastward acceleration of the Pacific plate between chrons C31 and C25. Such acceleration was necessarily accompanied by increased spreading rate along the East Pacific Rise (EPR), which separated the Pacific domain from the Farallon plate (e.g., Cande et al. 1988). However, there is no direct evidence that this episode was associated with plume activity in the Pacific area.

An interesting episode of *deceleration* was experienced by Eurasia between C25 (~57 Ma – Thanetian) and C21 (~47 Ma – Lutetian). Figures 6.26, 6.27 shows that this event was associated with a remarkable variation of the stage pole, while Figs. 6.26, 6.27 indicates that the end of this time interval coincided with an abrupt slowdown in the northward drift of India, possibly caused by the incipient collision with the Eurasian margin (e.g., Dewey et al. 1989; Huchon et al. 1994). A plate reconstruction at chron C25 is illustrated in Fig. 6.37.



**Fig. 6.37** Reconstruction at 57 Ma (C25 – Thanetian), illustrating the acceleration field of Eurasia (*red arrows*) and the global plate velocities (*black arrows*).  $P_{EUR}$  is

the pole of acceleration of Eurasia. *Black areas* are LIPS. Coastlines are overlaid for reference

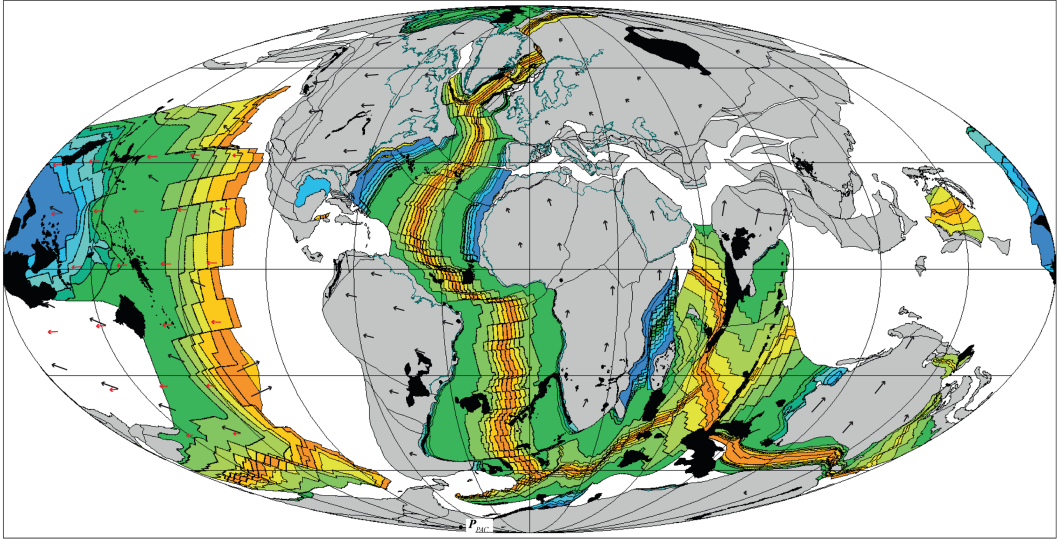


**Fig. 6.38** Reconstruction at 39 Ma (C18 – Bartonian), illustrating the acceleration field of Nazca (*red arrows*) and the global plate velocities (*black arrows*).  $P_{NAZ}$  is

the pole of acceleration of Nazca. *Black areas* are LIPS. Coastlines are overlaid for reference

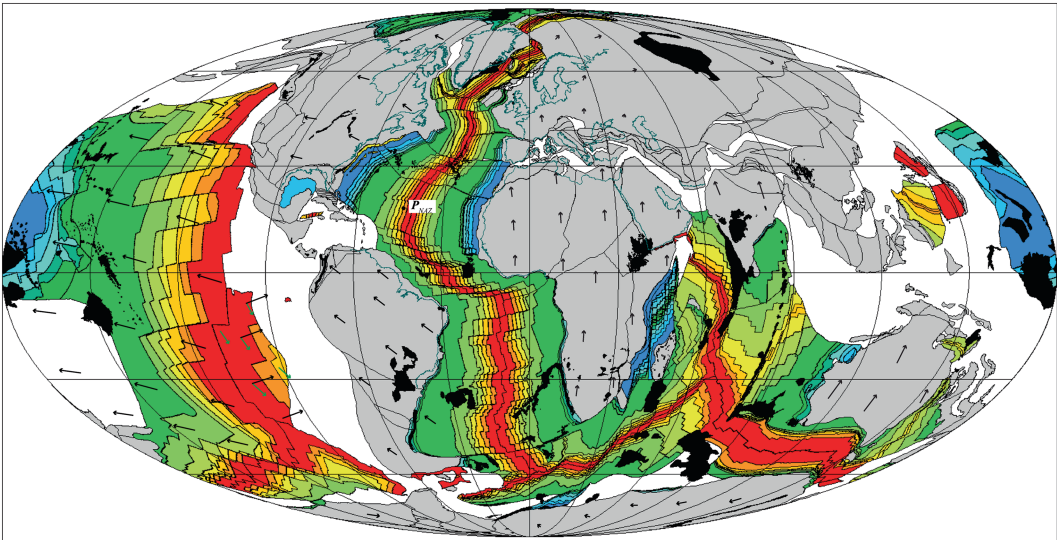
The next episode of traction occurred between C18 (~39 Ma – Bartonian) and C13 (~33 Ma – early Rupelian) and affected the Nazca plate (Fig. 6.32). It consisted into an eastward acceleration (Fig. 6.38), accompanied by a remarkable change of the stage pole. At the end of this stage,

the Pacific plate probably started a phase of westward acceleration and increased spreading rate along the EPR, which lasted at C6 (~20 Ma – Burdigalian) (Figs. 6.35 and 6.39). The last recognized episode of traction occurred once again in the Pacific area and consisted into a southward



**Fig. 6.39** Reconstruction at 33 Ma (C13 – Rupelian), illustrating the acceleration field of Pacific (*red arrows*) and the global plate velocities (*black arrows*).  $P_{PAC}$  is the

pole of acceleration of the Pacific plate. *Black areas* are LIPS. Coastlines are overlaid for reference



**Fig. 6.40** Reconstruction at 19 Ma (C6 – Burdigalian), illustrating the acceleration field of Nazca (*green arrows*) and the global plate velocities (*black arrows*).  $P_{NAZ}$  is

the pole of acceleration of Nazca. *Black areas* are LIPS. Coastlines are overlaid for reference

acceleration of Nazca between C6 and C5 (~11 Ma – Tortonian) (Fig. 6.40). In conclusion, eight major phases of non-equilibrium kinematics could have occurred during the last 230 Myrs. These episodes interested individual plates, which experienced a stage of accelerated motion,

accompanied by fast spreading on one boundary and rapid subduction along an opposite boundary. In some cases, the plates were subject to a sudden slowdown rather than to increasing velocity. In any event, the force exerted by the downgoing slabs was not balanced by the resistive viscous

drag exerted by the fluid asthenosphere, so that additional forces came to play. We shall come back to this point later in this book.

## 6.8 Non-dipole Paleomagnetic Fields

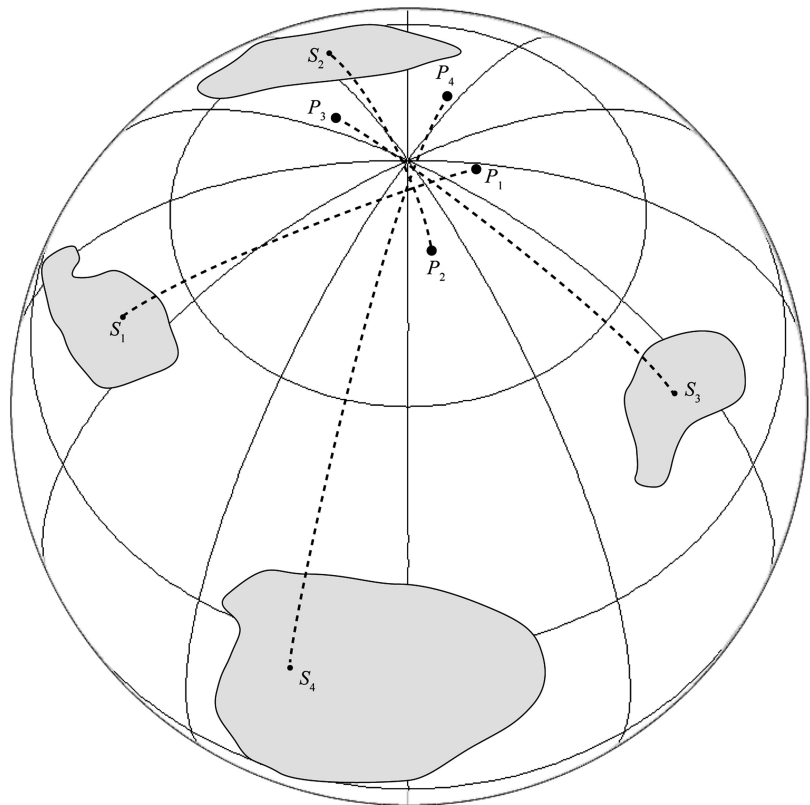
Up to this point, we have assumed that time-averaged paleomagnetic fields are *geocentric dipole* fields. However, several lines of evidence suggest that this is only an approximate picture of the reality, which could not be sufficient to solve some issues. In a study of paleomagnetic results from continental igneous rocks and oceanic sedimentary cores of Quaternary to Recent age, Wilson (1970) showed that individual paleopoles were systematically displaced from the geographic North Pole towards the far side with respect to the sampling site (Fig. 6.41), although they were correctly averaged to a point that statistically coincided with the North Pole. In

particular, the northern and southern hemisphere sea cores had respectively shallower and steeper inclinations, and this could not be a depositional effect. Therefore, Wilson (1970) suggested that the dipole equation (4.49), which implies a GAD field, is only an approximation of the correct formula.

To explain the paleopole distribution, Wilson (1970) proposed that a more refined model of the time-averaged geomagnetic field was that generated by an axial magnetic dipole, displaced northward along the spin axis. In this instance, the formula relating inclination to colatitude for an arbitrary displacement  $\delta z$  is:

$$\tan I = \frac{2a \cos \theta + 3\delta z (3\cos^2 \theta - 1)}{a \sin \theta + 6\delta z \sin \theta \cos \theta} \quad (6.62)$$

where  $a$  is the Earth's radius. This formula is the analogous of the GAD Eq. (4.49) for a magnetic dipole displaced along the  $z$  axis. Wilson's field was already known in geomagnetism, where it



**Fig. 6.41** Far-sided effect. In this example, Quaternary to Recent paleopoles  $P_i$ , sampled at locations  $S_i$  on different continents, are systematically displaced towards the far side from the geographic pole

was used in the non-axial form to approximate the Earth's magnetic field (Bartels 1936). It is usually referred to as the *eccentric dipole field*. Wilson's analysis of inclination data gave an offset  $\delta z = 191 \pm 38$  km. Most interestingly, Wilson (1970) proved that the field produced by the eccentric axial dipole is an axially symmetric field formed by a GAD field plus a quadrupole component. For an axially symmetric potential, we have that the spherical harmonic expansion (4.93) assumes the following simplified form:

$$V(r, \theta) = a \sum_{n=1}^{\infty} \left(\frac{a}{r}\right)^{n+1} g_n^0 P_n(\cos \theta); \quad r \geq a \quad (6.63)$$

which reduces to (4.97) for  $r = a$  (i.e., at the Earth's surface). The vertical and horizontal components of the magnetic field associated with this potential are given by:

$$\begin{cases} Z(r, \theta) = -B_r(r, \theta) = \frac{\partial V}{\partial r} \\ \quad = -\sum_{n=1}^{\infty} (n+1) \left(\frac{a}{r}\right)^{n+2} g_n^0 P_n(\cos \theta) \\ H(r, \theta) = -B_\theta(r, \theta) \\ \quad = \frac{1}{a} \frac{\partial V}{\partial \theta} = \sum_{n=1}^{\infty} \left(\frac{a}{r}\right)^{n+1} g_n^0 \frac{dP_n}{d\theta} \end{cases} \quad (6.64)$$

In the case of Wilson's field, only the dipole and quadrupole components of (6.63) and (6.64) survive, so that the potential  $V$  is given by:

$$V(r, \theta) = a \left[ g_1^0 \frac{a^2}{r^2} P_1(\cos \theta) + g_2^0 \frac{a^3}{r^3} P_2(\cos \theta) \right] \quad (6.65)$$

The Legendre polynomials appearing in (6.65) are given by:  $P_1(\cos \theta) = \cos \theta$  and  $P_2(\cos \theta) = (3 \cos^2 \theta - 1)/2$ . Therefore, we have:

$$V(r, \theta) = a \left[ g_1^0 \frac{a^2}{r^2} \cos \theta + g_2^0 \frac{a^3}{2r^3} (3 \cos^2 \theta - 1) \right] \quad (6.66)$$

At the Earth's surface, the vertical and horizontal components of the corresponding geomagnetic field are given by:

$$\begin{cases} Z(\theta) = -2g_1^0 \cos \theta - \frac{3}{2}g_2^0 (3 \cos^2 \theta - 1) \\ H(\theta) = -g_1^0 \sin \theta - 3g_2^0 \sin \theta \cos \theta \end{cases} \quad (6.67)$$

Therefore, for an eccentric dipole field the dipole equation (4.49) will be substituted by the following more complex expression:

$$\tan I = \frac{Z(\theta)}{H(\theta)} = \frac{2g_1^0 \cos \theta + \frac{3}{2}g_2^0 (3 \cos^2 \theta - 1)}{g_1^0 \sin \theta + 3g_2^0 \sin \theta \cos \theta} \quad (6.68)$$

A comparison of this equation with the Wilson solution (6.62) gives an estimate of the relative importance of the quadrupole term with respect to the dipole strength:

$$\frac{g_2^0}{g_1^0} = \frac{2\delta z}{a} \cong 0.06 \pm 0.01 \quad (6.69)$$

where we have used  $a = 6,371$  km and  $\delta z = 191 \pm 38$  km. In a later study, Wilson and McElhinny (1974) showed that the mean paleomagnetic field for the last 25 Myrs was an eccentric dipole field with  $\delta z = 325 \pm 57$  km. These authors also found a slow rate of change in the long-term structure of the geomagnetic field, hence in the parameter  $\delta z$ , during this time interval. However, the first comprehensive study about the non-dipolar components of the time-averaged geomagnetic field was published by Coupland and Van der Voo in 1980. These authors assumed an axial geometry for the paleomagnetic field and searched the best-fitting low-degree zonal Gauss coefficients,  $g_n^0$ , of a spherical harmonic expansion of the potential. With the available paleomagnetic data, they found that significant departures from the GAD symmetry were associated with the zonal quadrupole and octupole components,

with coefficients  $g_2^0$  and  $g_3^0$ , respectively. This study confirmed the existence of non-dipole components for the time interval between 130 Ma (early Barremian) and the present day, although the octupole component was negligible for times older than 26 Ma. The study of Coupland and Van der Voo (1980) was followed in 1983 and

1984 by two important articles of Livermore and colleagues (Livermore et al. 1983, 1984). These authors confirmed that the best substitute for the dipole equation was the following expression, which can be determined easily from (6.64) retaining only dipole, quadrupole, and octupole components:

---


$$\tan I = \frac{2g_1^0 \cos \theta + \frac{3}{2}g_2^0 (3 \cos^2 \theta - 1) + 2g_3^0 \cos \theta (5 \cos^2 \theta - 3)}{g_1^0 \sin \theta + 3g_2^0 \sin \theta \cos \theta + \frac{3}{2}g_3^0 \sin \theta (5 \cos^2 \theta - 1)} \quad (6.70)$$


---

This formula is usually written in the form:

---

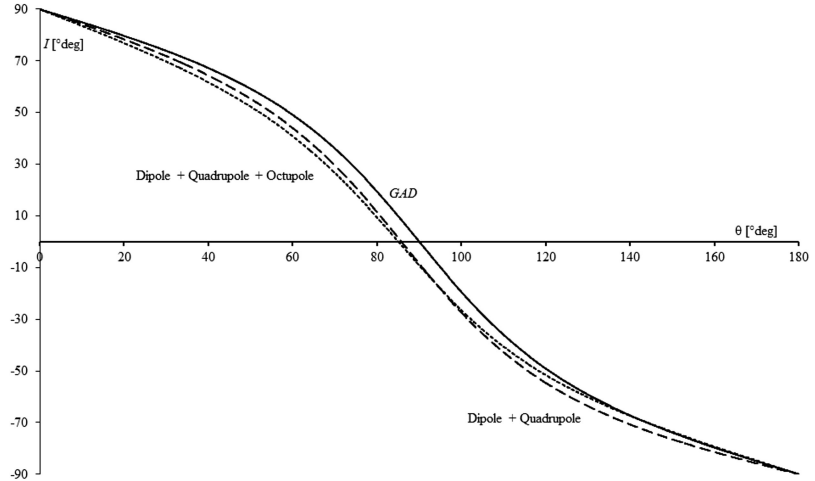

$$\tan I = \frac{2 \cos \theta + \frac{3}{2} (g_2^0/g_1^0) (3 \cos^2 \theta - 1) + 2 (g_3^0/g_1^0) \cos \theta (5 \cos^2 \theta - 3)}{\sin \theta + 3 (g_2^0/g_1^0) \sin \theta \cos \theta + \frac{3}{2} (g_3^0/g_1^0) \sin \theta (5 \cos^2 \theta - 1)} \quad (6.71)$$


---

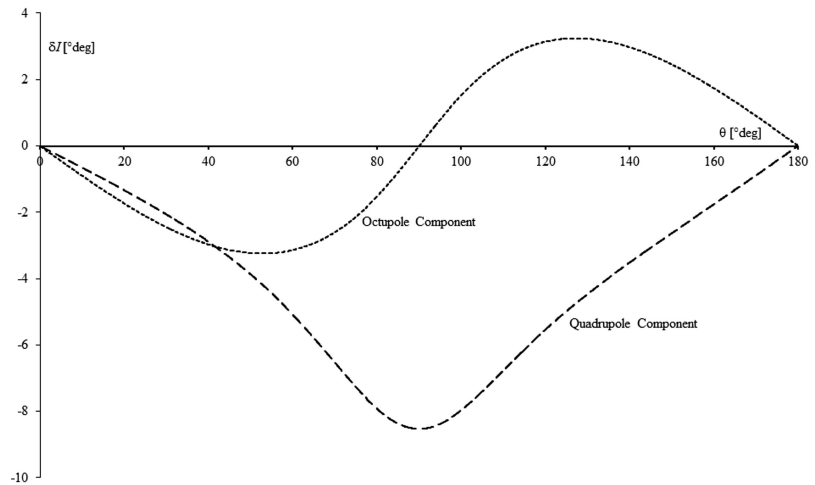
which evidences more clearly the dependence of the inclination from the ratios  $G_2 \equiv g_2^0/g_1^0$  and  $G_3 \equiv g_3^0/g_1^0$ . Livermore and colleagues found that  $G_2 = 0.045 \pm 0.015$  for the last 35 Myrs, while  $G_2 \cong 0.10$  between 40 and 60 Ma (Paleocene – Eocene), and it attained negative values between  $-0.07$  and  $-0.10$  during the Cretaceous and the Jurassic. Regarding the octupole component, these authors suggested the possibility that  $G_3 \cong 0.02$  for the last 5 Myrs, although they questioned that this result may be due to data errors. The effect of low-order non-dipolar components on the observed magnetic inclinations is illustrated in Fig. 6.42. We note that both the quadrupole and octupole components determine *shallower inclinations* in the *northern* hemisphere. However, in the southern hemisphere inclinations are *always* steeper only in the case of dipole-quadrupole fields, because when the octupole components have sufficiently large magnitude the inclination will be shallower in both hemispheres. Inclinations considerably shallower than those predicted from APW paths were effectively observed in central Asia (e.g., Si and Van der Voo 2001 and references therein).

On the basis of these observations, Si and Van der Voo (2001) proposed a paleomagnetic field geometry with negligible quadrupole component and significant octupole component ( $G_3 \geq 0.06$ ) for the time interval between the late Cretaceous and the Tortonian. The difference between inclination anomalies associated with quadrupole and octupole components is illustrated in Fig. 6.43. We note that an octupole inclination anomaly is antisymmetric with respect to the Equator, whereas a quadrupole anomaly is always negative and has its maximum at the Equator. In another paper, Van der Voo and Torsvik (2001) used a data set of N. American and European paleopoles with ages between 300 and 40 Ma, comparing the observed paleolatitudes with theoretical values predicted on the basis of the dipole equation. These authors assumed a paleomagnetic field geometry formed by a GAD field plus a zonal octupole component, obtaining  $G_3 = 0.1$ . In a successive study, Torsvik and Van der Voo (2002) confirmed this value using stable Gondwana paleopoles. One of the major issues addressed by these authors was associated with the classic fits of Pangaea (known as Pangaea A1 configurations, e.g., Van

**Fig. 6.42** Predicted inclination as a function of colatitude for three paleomagnetic field geometries: GAD (*solid line*), a dipole field plus a quadrupole component (*dashed line*), and a field which includes all zonal components up to degree two (*dotted line*). In this example,  $G_2 = 0.10$  and  $G_3 = 0.05$



**Fig. 6.43** Predicted inclination anomaly as a function of colatitude for a quadrupole component (*dashed line*) and an octupole component (*dotted line*). In this example,  $G_2 = 0.10$  and  $G_3 = 0.05$



der Voo 1993), which introduced a discrepancy between Permian-early Triassic paleopoles from Gondwana and those obtained from Laurasian sites.

In this instance, Permian and early Triassic reconstructions did not bring the two data sets sufficiently close to the North Pole, because Gondwana was placed too northwards with respect to Laurasia in the fit of Pangaea. The attempt to keep the GAD hypothesis led to several alternative, more or less creative, fits of this supercontinent, which are referred to as Pangaea A2, B, or C fits. Apparently, the only alternative solution to the paleomagnetic discrepancy, which would have allowed to retain the classic A1 fits (e.g., that illustrated in Fig. 2.30), was to admit the existence

of significant high-degree zonal components in the paleomagnetic fields. However, more recent studies have questioned this interpretation, which does not take into account of a phenomenon known as *sedimentary inclination error*. King (1955) first observed that the inclination of remnant magnetization of artificial sedimentary specimens,  $I$ , was different from that of the magnetizing field,  $I_F$ . He called this difference *inclination error*:

$$\delta I = I_F - I \quad (6.72)$$

The experiments of King (1955) showed that  $\delta I$  is independent from the field strength, but depends from the fraction  $f$  of particles with

sub-spherical geometry, and the corresponding fraction  $1 - f$  of plate-like particles.

He found that the two inclinations were related by the following simple equation:

$$\tan I = f \tan I_F \quad (6.73)$$

The empirical parameter  $f$  is called *flattening factor*. In more recent times, Tauxe and Kent (1984) have confirmed the validity of this equation and showed that  $f \cong 0.55$  for sediments with detrital hematite. This value indicated that the observed shallower inclinations could be associated with inclination errors rather than with non-dipole components of the paleomagnetic field (e.g., Gilder et al. 2003). The present orientation of paleomagnetists regarding the origin of the inclination anomaly is that it is associated with sedimentary inclination errors and/or poor-quality data, while a solution to the Pangaea fit can be found by selecting reliable data, correcting inclination errors, and improving the plate reconstructions (e.g., Domeier et al. 2012).

## Problems

1. Interpreting the APW path of a continent as the trajectory of a moving point at the Earth's surface, explain what kind of information can be obtained about the kinematics at a reference site  $S$ ;
2. Assuming that a tectonic plate is rotating clockwise about an Euler pole located at ( $0^\circ\text{N}, 90^\circ\text{E}$ ) with angular velocity  $\omega = 1^\circ/\text{Myr}$ , determine the northward and eastward components of velocity for a reference point at ( $30^\circ\text{N}, 45^\circ\text{E}$ ), and the rate of variation of its inclination and declination;
3. Given a mean paleopole for India at 100 Ma, describe a method to fit India to Madagascar in three steps;
4. Find an expression/algorithm for calculating the predicted paleolatitude and declination at a point  $P$  given the paleolatitude and declination at a reference site  $S$ .

## References

- Arason P, Levi S (1990) Models of inclination shallowing during sediment compaction. *J Geophys Res* 95(B4):4481–4499
- Bartels J (1936) The eccentric dipole approximating the earth's magnetic field. *Terr Magn Atmos Electr* 41:225–250
- Besse J, Courtillot V (1988) Paleogeographic maps of the continents bordering the Indian Ocean since the early Jurassic. *J Geophys Res* 93(B10):11,791–11,808
- Besse J, Courtillot V (1991) Revised and synthetic apparent polar wander paths of the African, Eurasian, North American and Indian plates, and true polar wander since 200 Ma. *J Geophys Res* 96(B3):4029–4050
- Besse J, Courtillot V (2002) Apparent and true polar wander and the geometry of the geomagnetic field over the last 200 Myr. *J Geophys Res* 107(B11):2300
- Butler RF (1992) *Paleomagnetism: magnetic domains to Geologic Terranes*. Blackwell Scientific Publications, Boston, 238 pp
- Cande SC, Stegman DR (2011) Indian and African plate motions driven by the push force of the Réunion plume head. *Nature* 475:47–52. doi:10.1038/nature10174
- Cande SC, LaBrecque JL, Haxby WF (1988) Plate kinematics of the South Atlantic: chron C34 to present. *J Geophys Res* 93(B11):13479–13492
- Chase CG (1979) Asthenospheric counterflow: a kinematic model. *Geophys J R Astron Soc* 56:1–18
- Coupland DH, Van der Voo R (1980) Long-term nondipole components in the geomagnetic field during the last 130 m.y. *J Geophys Res* 85(B7):3529–3548
- Dewey JF, Cande S, Pitman WC (1989) Tectonic evolution of the India/Eurasia collision zone. *Eclogae Geol Helv* 82(3):717–734
- Domeier M, Van der Voo R, Torsvik TH (2012) Paleomagnetism and Pangea: the road to reconciliation. *Tectonophysics* 514:14–43
- Duncan RA (1984) Age progressive volcanism in the New England seamounts and the opening of the central Atlantic Ocean. *J Geophys Res* 89(B12):9980–9990
- Dunlop DJ, Özdemir Ö (1997) *Rock Magnetism*. Cambridge University Press, Cambridge, 573 pp
- Eubank RL (1999) Nonparametric regression and spline smoothing. Marcel Dekker, New York, 337 pp
- Evans DA (2003) True polar wander and supercontinents. *Tectonophysics* 362(1):303–320
- Fisher RA (1953) Dispersion on a sphere. *Proc R Soc London A* 217:295–305
- Frankel HR (2012) The continental drift controversy: Wegener and the early debate. Cambridge University Press, Cambridge, 632 pp
- Gilder S, Chen Y, Cogné JP, Tan X, Courtillot V, Sun D, Li Y (2003) Paleomagnetism of Upper Jurassic to Lower Cretaceous volcanic and sedimentary rocks from the

- western Tarim Basin and implications for inclination shallowing and absolute dating of the M-0 (ISEA?) chron. *Earth Planet Sci Lett* 206(3):587–600
- Gordon RG (1987) Polar wandering and paleomagnetism. *Ann Rev Earth Planet Sci* 15:567–593
- Gordon RG, Cox A, O'Hare S (1984) Paleomagnetic Euler poles and the apparent polar wander and absolute motion of North America since the Carboniferous. *Tectonics* 3(5):499–537
- Harrison CGA, Lindh T (1982) A polar wandering curve for North America during the Mesozoic and Cenozoic. *J Geophys Res* 87(B3):1903–1920
- Huchon P, Le Pichon X, Rangin C (1994) Indochina peninsula and the collision of India and Eurasia. *Geology* 22(1):27–30
- Irving E (2005) The role of latitude in mobilism debates. *Proc Natl Acad Sci U S A* 102(6):1821–1828
- Jupp PE, Kent JT (1987) Fitting smooth paths to spherical data. *Appl Stat* 36(1):34–46
- Jurdy DM, Van der Voo R (1974) A method for the separation of true polar wander and continental drift, including results for the last 55 my. *J Geophys Res* 79(20):2945–2952
- Kachkachi H, Ezzir A, Noguès M, Tronc E (2000) Surface effects in nanoparticles: application to maghemite  $\gamma$ - $\text{Fe}_2\text{O}_3$ . *Eur Phys J B* 14:681–689
- Kent DV, Van Der Voo R (1990) Palaeozoic palaeogeography from palaeomagnetism of the Atlantic-bordering continents. In: McKerrow WS, Scotese CR (eds) *Palaeozoic palaeogeography and biogeography*. Geological Society of London Memoir 12, London, pp 49–56
- Kent DV, Witte WK (1993) Slow apparent polar wander for North America in the Late Triassic and large Colorado Plateau rotation. *Tectonics* 12(1):291–300
- King RF (1955) The remanent magnetism of artificially deposited sediments. *Geophys J R Astron Soc* 7:115–134. doi:10.1111/j.1365-246X.1955.tb06558.x7
- Livermore RA, Vine FJ, Smith AG (1983) Plate motions and the geomagnetic field – I. Quaternary and late Tertiary. *Geophys J R Astron Soc* 73:153–171
- Livermore RA, Vine FJ, Smith AG (1984) Plate motions and the geomagnetic field – II. Jurassic to tertiary. *Geophys J R Astron Soc* 79:939–961
- May SR, Butler RF (1986) North American Jurassic apparent polar wander: implications for plate motion, paleogeography and Cordilleran tectonics. *J Geophys Res* 91(B11):11519–11544
- McElhinny MW, Lock J (1990) Global palaeomagnetic data base project. *Phys Earth Planet Int* 63:1–6
- McElhinny MW, McFadden PL (2000) *Paleomagnetism: continents and oceans*. Academic Press, San Diego, 382 pp
- McFadden PL (1980) The best estimate of Fisher's precision parameter  $\kappa$ . *Geophys J R Astron Soc* 60:397–407
- Morgan WJ (1971) Convection plumes in the lower mantle. *Nature* 230:42–43
- Müller RD, Royer J-Y, Lawver LA (1993) Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* 21(3):275–278
- Musgrave RJ (1989) Aweighted least-squares fit of the Australian apparent polar wander path for the last 100 Myr. *Geophys J* 96:231–243
- Néel L (1949) Théorie du traînage magnétique des ferromagnétiques en grains fins avec applications aux terres cuites. *Ann Géophys* 5(2):99–136
- Parker RL, Denham R (1979) Interpolation of unit vectors. *Geophys J R Astron Soc* 58:685–687
- Pullaiah G, Irving E, Buchan KL, Dunlop DJ (1975) Magnetization changes caused by burial and uplift. *Earth Planet Sci Lett* 28:133–143
- Ruiz-Martínez VC, Torsvik TH, van Hinsbergen DJ, Gaina C (2012) Earth at 200Ma: global palaeogeography refined from CAMP palaeomagnetic data. *Earth Planet Sci Lett* 331:67–79
- Schettino A, Scotese CR (2000) A synthetic APWP for Africa (Jurassic-Present) and global plate tectonic reconstructions. Paper presented at the AGU Spring Meeting, Washington, May 30–June 3
- Schettino A, Scotese CR (2001) New Internet software aids paleomagnetic analysis and plate tectonic reconstructions. *Eos Trans AGU* 82(45):530–536
- Schettino A, Scotese CR (2005) Apparent polar wander paths for the major continents (200Ma–Present Day): a paleomagnetic reference frame for global plate tectonic reconstructions. *Geophys J Int* 163(2):727–759
- Schettino A, Turco E (2011) Tectonic history of the western Tethys since the late Triassic. *GSA Bull* 123(1/2):89–105. doi:10.1130/B30064.1
- Si J, Van der Voo R (2001) Too-low magnetic inclinations in central Asia: an indication of a long-term tertiary non-dipole field? *Terra Nova* 13:471–478
- Steinberger B, Torsvik TH (2008) Absolute plate motions and true polar wander in the absence of hotspot tracks. *Nature* 452:620–623
- Tauxe L, Kent DV (1984) Properties of a detrital remanence carried by hematite from study of modern river deposits and laboratory redeposition experiments. *Geophys J Roy Astron Soc* 77:543–561
- Thompson R, Clark RM (1981) Fitting polar wander paths. *Phys Earth Planet Inter* 27:1–7
- Torsvik TH, Van der Voo R (2002) Refining Gondwana and Pangea palaeogeography: estimates of Phanerozoic non-dipole (octupole) fields. *Geophys J Int* 151:771–794
- Torsvik TH, Smethurst MA, Meert JG, Van der Voo R, McKerrow WS, Brasier MD, Sturt BA, Walderhaug HJ (1996) Continental break-up and collision in the Neoproterozoic and Palaeozoic – a tale of Baltica and Laurentia. *Earth Sci Rev* 40(3):229–258
- Torsvik TH, Van der Voo R, Meert JG, Mosar J, Walderhaug HJ (2001) Reconstructions of the continents around the North Atlantic at about the 60th parallel. *Earth Planet Sci Lett* 187(1):55–69
- Torsvik TH, Müller RD, Van der Voo R, Steinberger B, Gaina C (2008) Global plate motion frames: toward a unified model. *Rev Geophys* 46(3):RG3004. doi:10.1029/2007RG000227
- Torsvik TH, Van der Voo R, Preeden U, Mac Niocaill C, Steinberger B, Doubrovine PV, van Hinsbergen DJJ,

- Domeier M, Gaina C, Tohver E, Meert JG, McCausland PJA, Robin L, Cocks LRM (2012) Phanerozoic polar wander, palaeogeography and dynamics. *Earth Sci Rev* 114(3–4):325–368
- Van der Voo R (1990) The reliability of paleomagnetic data. *Tectonophysics* 184(1):1–9
- Van der Voo R (1993) *Paleomagnetism of the Atlantic, Tethys and Iapetus Oceans*. Cambridge University Press, Cambridge, 411pp
- Van der Voo R, Torsvik TH (2001) Evidence for late Paleozoic and Mesozoic non-dipole fields provides an explanation for the Pangea reconstruction problems. *Earth Planet Sci Lett* 187:71–81
- Van Fossen MC, Kent DV (1992) Paleomagnetism of the Front Range (Colorado) Morrison Formation and an alternative model of Late Jurassic North American apparent polar wander. *Geology* 20(3):223–226
- Wegener A (1912) Die Entstehung der Kontinente. *Geologische Rundschau* 3(4):276–292
- Wilson RL (1970) Permanent aspects of the Earth's non-dipole magnetic field over upper Tertiary times. *Geophys J R Astron Soc* 19:417–437
- Wilson RL, McElhinny MW (1974) Investigation of the large scale palaeomagnetic field over the past 25 million years. Eastward shift of the Icelandic spreading ridge. *Geophys J Roy Astron Soc* 39(3):570–586
- Ziegler AM, Scotese CR, Barrett SF (1983) Mesozoic and Cenozoic paleogeographic maps. In: Brosche P, Sundermann J (eds) *Tidal friction and the Earth's rotation II*. Springer, Berlin, pp 240–252



---

## Part II

# Dynamics of the Lithosphere and the Mantle



**Abstract**

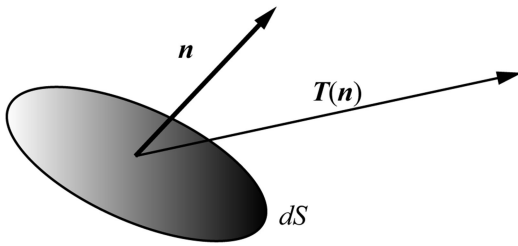
In this chapter, the fundamentals of continuum mechanics are presented, in particular the concepts of stress and strain, and the general equations of motion for a continuum medium. Finally, it is introduced the discipline of rheology and the basic rheological models, which are widely used in geodynamics.

**7.1 The Stress Tensor**

In the first part of this book, we have described the principles of plate kinematics, hence the geometry of plate motions, independently from the active and passive forces that drive the continental drift and the formation of new oceanic crust. Now we are going to consider the system of torques exerted on the lithosphere, and the relation between these torques and deformation processes at different time scales, using the formalism of continuum mechanics (see Sect. 2.1). In the next chapters, we shall prove that both plate motions and deformation of the lithosphere are intimately related to the dynamics of the Earth's mantle. Our starting point will be a review of the relationships between forces and deformation in a continuum body.

In Sect. 2.1, we have introduced the concept of *body forces*, which are external action-at-a-distance force fields, such as gravity or magnetic fields, that operate on each volume element  $dV$  of a body. In addition to these forces, a body may be subject to other interactions with the external

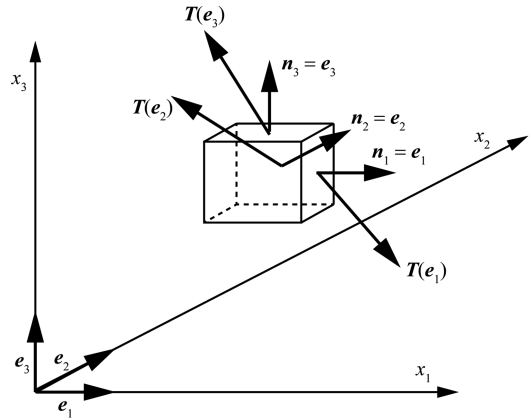
world, by direct *pressure* along its boundary surface. All these forces and force fields may simply produce a change in the state of motion of a body, if it has sufficient *rigidity*. In this instance, the distance between any two points in the region  $\mathbf{R}$  filled by the body does not change with time and the relation between forces and motion is described by Euler's equations (e.g., Goldstein 1980). However, when we consider a tectonic plate this representation is not generally adequate, because in most cases plates deform under the action of external forces. As soon as a body changes its shape, a new force field is generated within the deformed region  $\mathbf{R}'$ , which tries to restore the original undeformed shape  $\mathbf{R}$  or simply opposes deformation. This is a short-range force field between adjacent molecules or atoms of the body that can be represented as a *distributed* interaction across contact surfaces. Therefore, it is usually referred to as a *surface forces* field. A classic example of this class of forces is the resistance opposed by a spring to an external force during its extension or compression. Another example is the pressure field within a compressed fluid, or its resistance to deformations



**Fig. 7.1** The traction across a surface element within a deformed body

that do not involve volume changes. These forces can be modelled assuming that for any arbitrary surface element  $dS = \mathbf{n}dS$ ,  $\mathbf{n}$  being the versor normal to the surface and  $dS$  its infinitesimal area, there exists a force per unit area  $\mathbf{T} = \mathbf{T}(\mathbf{n})$  across the surface element, which depends from its orientation and from the mechanical state of the body (Fig. 7.1). This is a representation of the distributed contact interaction between molecules and atoms across the surface element, so that the effective force exerted across  $dS$  will be the infinitesimal force  $d\mathbf{F}(\mathbf{n}) = \mathbf{T}(\mathbf{n})dS$ . The vector  $\mathbf{T}$ , which has units  $[\text{Nm}^{-2}] \equiv [\text{Pa}]$ , is called *traction*. In equilibrium conditions, the force exerted on the molecules that are close to one of the two sides of the surface element has equal magnitude and opposite direction with respect to the force exerted on the molecules that lie along the opposite side, thereby  $\mathbf{T}(-\mathbf{n}) = -\mathbf{T}(\mathbf{n})$ . The components of  $\mathbf{T}$  that are normal and parallel to the plane of  $dS$  are called, respectively, *normal stress* and *shear stress*. In the case of a liquid substance, the shear stress is always zero, so that  $\mathbf{T} = -P\mathbf{n}$ , where  $P$  is the *hydrostatic pressure*.

Let us consider now the system of surface forces exerted on a volume element  $dV$  within a deformed body. Let us assume that  $dV$  is a small parallelepiped with sides parallel to the coordinate axes of a Cartesian reference frame (Fig. 7.2). In the following, we shall use in most cases index notation ( $x_1 \equiv x$ ,  $x_2 \equiv y$ ,  $x_3 \equiv z$ ) and Einstein's convention summation, which allows to simplify the mathematical notation. In equilibrium conditions, the tractions exerted on opposite sides of  $dV$  have equal magnitude and opposite directions. We shall assume that the unit



**Fig. 7.2** Tractions exerted on a volume element  $dV$ , with sides  $dx_i dx_j$ ,  $i, j = 1, 2, 3$

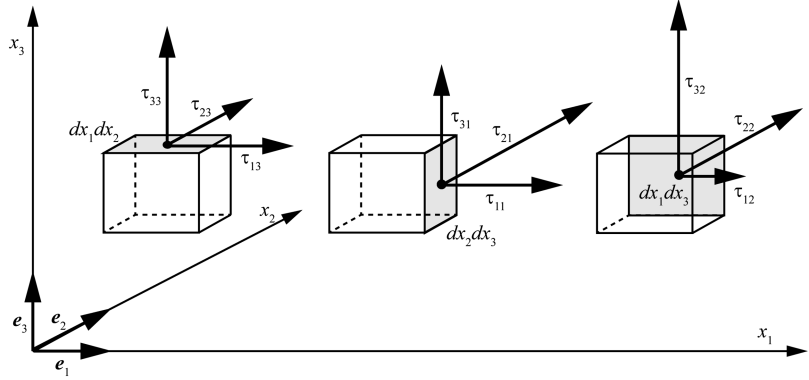
vector representing the orientation of a surface element is always directed *outwards* from closed surfaces. In the case of the volume element  $dV$ , three of its sides have versors  $\mathbf{n}_i$  coinciding with the base versors  $\mathbf{e}_i$  (Fig. 7.2). The components of the three tractions vectors  $\mathbf{T}(\mathbf{e}_i)$  form a rank-two tensor called *stress tensor*:

$$\tau_{ij} = T_i(\mathbf{e}_j) \quad (7.1)$$

With this notation, the second index of the stress tensor indicates the direction of the surface across which the traction is exerted, while the first index identifies the traction component. We shall prove soon that this tensor completely determines the surface force field existing within a deformed body. The sign convention for the stress tensor is that when a diagonal component  $\tau_{ii}$  is *positive*, then it is directed *outwards* the volume element under consideration, while a positive off-diagonal component  $\tau_{ij}$  ( $i \neq j$ ) indicates that the corresponding shear stress component is directed as  $\mathbf{e}_i$ . For example,  $\tau_{32} > 0$  implies that a shear stress is exerted on the surface element  $dx_1 dx_3$  (whose normal versor is  $\mathbf{e}_2$ ) in the direction of  $\mathbf{e}_3$ . This convention is illustrated in Fig. 7.3.

In equilibrium conditions, it is easy to determine the tractions on the opposite faces of the volume element. In fact, we have simply:  $T_i(-\mathbf{e}_j) = -T_i(\mathbf{e}_j) = -\tau_{ij}$ . Therefore, while the total force exerted on the volume element is clearly zero,

**Fig. 7.3** Positive components of the stress tensor



the off-diagonal components of  $\tau$  generate six couples with the corresponding components of traction along the opposite faces. At the equilibrium, both the total force and the total torque must be zero, thereby, two torques in direction  $e_i$  and  $-e_i$  always cancel out. As a consequence, the stress tensor is symmetric and we have only six independent components:

$$\tau_{ij} = \tau_{ji} \quad (7.2)$$

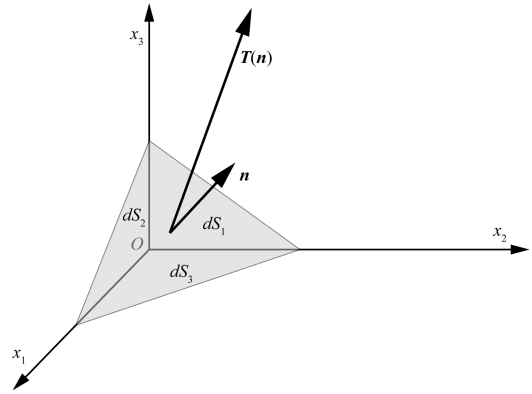
The importance of the stress tensor in the description of the surface force fields within deformed bodies arises from its capability to predict the traction along *any* surface element. This is a consequence of the following theorem:

### Cauchy's Theorem

*In equilibrium conditions, for an arbitrary surface element with normal vector  $\mathbf{n}$ , the components of traction are given by:*

$$T_i = \tau_{ij} n_j \quad (7.3)$$

*Proof* A surface element of arbitrary shape can be always divided into a set of triangles. Therefore, we shall prove the theorem for a triangle having area  $dS$  and orientation  $\mathbf{n}$ . Let us choose a reference frame with the origin  $O$  very close to this triangle, as shown in Fig. 7.4. The triangle forms a tetrahedron with the three coordinate axes, with sides  $dS$ ,  $dS_1$ ,  $dS_2$ , and  $dS_3$ ,  $dS_i$  being orthogonal to  $e_i$ . The unit vectors associated with the four triangles bounding the tetrahedron are,



**Fig. 7.4** The Cauchy tetrahedron

respectively,  $\mathbf{n}$ ,  $-\mathbf{e}_1$ ,  $-\mathbf{e}_2$ , and  $-\mathbf{e}_3$ . It is easy to realize that the areas of triangles  $dS_i$  are given by:  $dS_i = dS \cdot e_i = (\mathbf{n} \cdot \mathbf{e}_i) dS = n_i dS$ . The total force  $\mathbf{F}$  exerted on the tetrahedron is the sum of the forces exerted on the individual faces. At the equilibrium, it must be  $F_i = 0$ . Therefore, taking into account that  $i$ -th component of the force exerted on the surface elements  $dS_j$  is  $-\tau_{ij} dS_j$  (no implicit summation) we have:

$$T_i dS - \tau_{ij} n_j dS = 0 \quad (7.4)$$

Dividing Eq. (7.4) by  $dS$  gives (7.3). This proves Cauchy's theorem. ■

Cauchy's theorem shows that the stress tensor completely determines the surface force field existing within a deformed body. It can be considered as the linear operator that generates a

traction vector  $\mathbf{T}$  from a direction vector  $\mathbf{n}$ . In general, the components of this tensor change with the position within the body. However, its symmetry guarantees that we can always find three directions  $\mathbf{n}^i$  such that the shear stress is zero along surface elements orthogonal to  $\mathbf{n}^i$ . In these conditions,  $\mathbf{n}^i$  and  $\mathbf{T}(\mathbf{n}^i)$  are parallel each other, so that by Cauchy's theorem (7.3) we have:

$$\mathbf{T}(\mathbf{n}^i) = \lambda_i \mathbf{n}^i = \tau \mathbf{n}^i ; i = 1, 2, 3 \quad (7.5)$$

where the quantities  $\lambda_i$  are scalars. To find a vector  $\mathbf{n}^i$  that satisfies (7.5), we solve the following eigenvalue equation:

$$(\tau - \mathbf{I}\lambda_i) \mathbf{n}^i = 0 \quad (7.6)$$

where  $\mathbf{I}$  is the identity matrix of order three. The parameter  $\lambda_i$  that satisfies this equation is the *eigenvalue* of the equation, while  $\mathbf{n}^i$  is the *eigenvector*. Equation 7.6 represents a homogeneous system of three linear equations in the unknown eigenvector components  $n_j^i$ , which has non-trivial solutions only when:

$$\det(\tau - \mathbf{I}\lambda_i) = 0 \quad (7.7)$$

This is a cubic equation, which has three real solutions because of the stress tensor symmetry. To find the eigenvector corresponding to a given eigenvalue  $\lambda_i$ , we insert this value in (7.6) and solve for the eigenvector components  $n_j^i$ , taking into account that only two of the three components are independent. The three eigenvectors  $\mathbf{n}^i$  are orthogonal each other and define a new *local* reference frame that is called the system of the *principal axes* of stress. In this local coordinate system, the stress tensor is represented by a diagonal matrix  $\sigma$  given by:

$$\sigma = \mathbf{N}^T \tau \mathbf{N} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \quad (7.8)$$

where  $\mathbf{N}$  is a matrix formed with the components of the eigenvectors  $\mathbf{n}^i$ :

$$N_{ij} = n_i^j \quad (7.9)$$

By convention, the three stresses  $\sigma_i$  are sorted in such a way that:  $|\sigma_1| > |\sigma_2| > |\sigma_3|$ . When  $\sigma_1 = \sigma_2 = \sigma_3$  we say that the stress field is *hydrostatic*. In liquids, the stress tensor is *always* diagonal and  $\sigma_i \equiv -P$ , where  $P$  is the *pressure*. An important feature of the principal stresses  $\sigma_i$  at a point  $\mathbf{r}$  is that they allow to predict the planes of maximum shear stress at that point. In geology, these are the planes along which the probability of rupture and faulting, for a given stress field, reaches a maximum. Now we are going to show that there are just two orientations of  $\mathbf{n}$  such that the shear component of  $\mathbf{T}(\mathbf{n})$  is maximum, and  $\sigma_1$  and  $\sigma_3$  form angles of  $45^\circ$  with each of these planes. Let us consider an arbitrary plane with normal vector  $\mathbf{n}$  in the principal axes coordinate system. In this instance, by Cauchy's theorem the normal stress along this plane is given by:

$$T_N(\mathbf{n}) = \mathbf{T}(\mathbf{n}) \cdot \mathbf{n} = \sigma_i n_i^2 \quad (7.10)$$

Therefore, the squared shear stress component will be given by:

$$T_S^2(\mathbf{n}) = T^2(\mathbf{n}) - T_N^2(\mathbf{n}) = \sigma_i^2 n_i^2 - (\sigma_i n_i^2)^2 \quad (7.11)$$

To find the directions of maximum shear stress, we must find the solutions of the equations:

$$\frac{\partial T_S^2}{\partial n_j} = 0 \quad (7.12)$$

with the constraint that:

$$n_i n_i = 1 \quad (7.13)$$

This is a classic problem of finding a conditional maximum, which can be solved using Lagrange's multipliers. In this instance, we will search the solutions of the equations:

$$\frac{\partial T_S^2}{\partial n_j} = \lambda n_j ; j = 1, 2, 3 \quad (7.14)$$

where  $\lambda$  is the Lagrange multiplier. Using (7.11), these equations can be rewritten as follows:

$$2n_j \left[ \sigma_j^2 - 2(\sigma_i n_i^2) \sigma_j \right] = \lambda n_j \quad ; \quad j = 1, 2, 3 \quad (7.15)$$

Solving these equations with the constraint (7.13) gives nine solutions, three of which are the base versors  $\mathbf{n}^i$  (for which  $T_S = 0$ ). The remaining six represent the planes where  $T_S$  has local or global maxima:

$$\left( \frac{1}{\sqrt{2}}, \pm \frac{1}{\sqrt{2}}, 0 \right) ; \left( \frac{1}{\sqrt{2}}, 0, \pm \frac{1}{\sqrt{2}} \right) ; \left( 0, \frac{1}{\sqrt{2}}, \pm \frac{1}{\sqrt{2}} \right)$$

It is easy to verify that  $T_S$  attains its maximum for  $\mathbf{n} = (1/\sqrt{2})(\mathbf{n}^1 \pm \mathbf{n}^3)$ . For these directions, it results that the shear and normal components of stress are given by:

$$T_S = \frac{\sigma_1 - \sigma_3}{2} \quad ; \quad T_N = \frac{\sigma_1 + \sigma_3}{2} \quad (7.16)$$

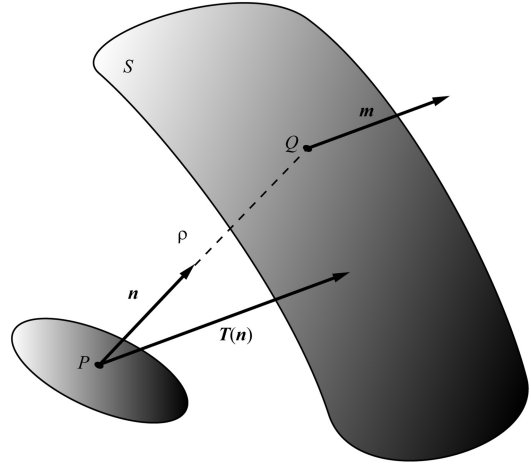
In the case of two-dimensional problems, for example in the study of the state of stress along vertical faults, it is easy to determine direct formulae for the normal and shear components of stress, and for the principal stresses. If  $s$  is the strike of a vertical fault (clockwise angle from the North) and the coordinate axes  $x_1$  and  $x_2$  are directed, respectively, northward and eastward, then:

$$\begin{cases} T_N(s) = \frac{1}{2}(\tau_{11} + \tau_{22}) + \frac{1}{2}(\tau_{11} - \tau_{22}) \cos 2s \\ \quad + \tau_{12} \sin 2s \\ T_S(s) = \frac{1}{2}(\tau_{22} - \tau_{11}) \sin 2s + \tau_{12} \cos 2s \end{cases} \quad (7.17)$$

Then, we see that  $T_S = 0$  for:

$$s = s_1 = \frac{1}{2} \tan^{-1} \left[ \frac{2\tau_{12}}{\tau_{11} - \tau_{22}} \right] \quad (7.18)$$

One principal axis has direction  $\mathbf{n}^1 = (\cos s_1, \sin s_1)$ , the other one has versor  $\mathbf{n}^2 = (\cos s_2, \sin s_2) = (\sin s, \cos s)$ , where  $s_2 = s_1 + \pi/2$ . The principal stresses are obtained substituting  $s_1$  and  $s_2$  into the expression (7.17) for of  $T_N$ .



**Fig. 7.5** Cauchy's stress surface. Versor  $\mathbf{m}$  is parallel to  $\mathbf{T}(\mathbf{n})$  and  $\rho = 1/\sqrt{T_N}$  is the distance  $\overline{PQ}$

We obtain the following direct formula for  $\sigma_1$  and  $\sigma_2$ :

$$\begin{cases} \sigma_1 = \frac{1}{2}(\tau_{11} + \tau_{22}) + \sqrt{\tau_{12}^2 + \frac{1}{4}(\tau_{11} - \tau_{22})^2} \\ \sigma_2 = \frac{1}{2}(\tau_{11} + \tau_{22}) - \sqrt{\tau_{12}^2 + \frac{1}{4}(\tau_{11} - \tau_{22})^2} \end{cases} \quad (7.19)$$

Now we are going to describe the variability of the normal stress  $T_N(\mathbf{n})$  at a point  $P$ , which is determined by (7.10) in the reference frame of the principal axes, as a function of  $\mathbf{n}$ . This can be done through an elegant geometrical method due to Cauchy. For any versor  $\mathbf{n}$  at  $P$ , let us consider the point  $Q$  along the direction of  $\mathbf{n}$ , at distance:

$$\rho(\mathbf{n}) = \frac{1}{\sqrt{T_N(\mathbf{n})}} \quad (7.20)$$

The set of points  $Q$ , whose distance from  $P$  is  $\rho(\mathbf{n})$  for variable  $\mathbf{n}$ , forms a surface  $S$  that traces the variations of normal stress at  $P$ . This surface is known as the *Cauchy stress surface* and has the important property that its normal at  $Q$  has the same direction of  $\mathbf{T}(\mathbf{n})$  at  $P$ , as illustrated in Fig. 7.5.

To prove this assertion, we first note that if  $q_i$  are the Cartesian coordinates of  $Q$  in the principal axes reference frame, which has origin at  $P$ , then

the components of  $\mathbf{n}$  are given by:  $n_i = q_i/\rho$ . Substituting into (7.10) and taking into account of (7.20) gives:

$$T_N(\mathbf{n}) \rho^2 = \pm 1 = \sigma_i q_i^2 \quad (7.21)$$

In this equation, the minus sign is associated with compressive stresses, thereby, (7.10) reduces to a single triaxial ellipsoidal surface with semi-axes  $|\sigma_i|^{-1/2}$ :

$$|\sigma_i| q_i^2 = 1 \quad (7.22)$$

The components of the versors normal to this surface are given by the gradient of (7.22) (see Appendix1):

$$m_i = \kappa \frac{\partial}{\partial q_i} (|\sigma_i| q_i^2) = 2\kappa |\sigma_i| q_i \quad (7.23)$$

where  $\kappa$  is a constant that ensures that  $|\mathbf{m}| = 1$ . Now, in the principal axes coordinate system (7.3) reduces to:

$$T_i(\mathbf{n}) = \sigma_i n_i = \frac{1}{\rho} \sigma_i q_i \quad (7.24)$$

Therefore,

$$m_i = \pm 2\kappa \sigma_i q_i = \pm 2\kappa \rho T_i(\mathbf{n}) \quad (7.25)$$

This expression proves that  $\mathbf{m}$  and  $T(\mathbf{n})$  are parallel. As a consequence, Cauchy's stress ellipsoid completely determines the state of stress at a point. Another important general feature of the stress tensor is that it can be decomposed into an *isotropic stress*,  $\tau_0$ , which describes the hydrostatic component of stress, and a *deviator*,  $\tau'$ , associated essentially but not exclusively with shear:

$$\tau_{ij} = -p\delta_{ij} + \tau'_{ij} \quad (7.26)$$

where  $p = \tau_{kk}/3$  is the mean pressure and  $\tau'_{ij} = \tau_{ij} - \tau_{kk}\delta_{ij}/3$ . This is often a convenient separation, because stresses in the deep Earth are dominated by the large compressive components associated with the hydrostatic pressure, whereas the deviatoric component, associated with

flow and yielding, is relatively small. The last important property of the stress tensor is represented by its *invariants*. The characteristic Eq. (7.7) can be written as follows:

$$\det(\boldsymbol{\tau} - I\lambda_i) = -\lambda^3 + I_1\lambda^2 + I_2\lambda + I_3 = 0 \quad (7.27)$$

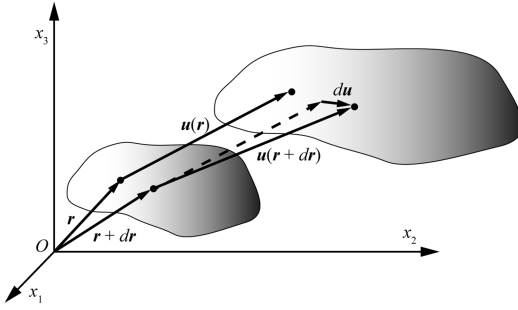
where,

$$\begin{aligned} I_1 &= \tau_{11} + \tau_{22} + \tau_{33} = \text{Tr}(\boldsymbol{\tau}) \\ I_2 &= -(\tau_{11}\tau_{22} - \tau_{12}\tau_{21}) - (\tau_{11}\tau_{33} - \tau_{13}\tau_{31}) \\ &\quad - (\tau_{22}\tau_{33} - \tau_{23}\tau_{32}) \\ I_3 &= \det(\boldsymbol{\tau}) \end{aligned} \quad (7.28)$$

It is possible to prove that the quantities  $I_1$ ,  $I_2$ , and  $I_3$ , which are referred to as the *first*, *second*, and *third invariant*, are unchanged under coordinate transformations, so that they are invariants of the stress tensor. While the first invariant simply states that the hydrostatic pressure is independent from the selected coordinate system, the second one plays an important role in the description of the fluid behaviour of the rocks and non-linear creep.

## 7.2 Displacement Fields and Strain

Let us face now the problem of describing quantitatively the *deformation* of a continuum body under the action of external forces. When a continuum body changes its shape, each point in the undeformed region  $\mathbf{R}$ , with position vector  $\mathbf{r}$ , is subject to a *displacement*  $\mathbf{u}$  that depends from the position. The set of all displacements for each point  $\mathbf{r} \in \mathbf{R}$  forms a vector field  $\mathbf{u} = \mathbf{u}(\mathbf{r})$  (Fig. 7.6). Clearly, the existence of a non-zero displacement field does not imply automatically that a body has been deformed, because rigid body translations and rotations are also associated with displacement fields. In principle, the existence of a deformation can be established observing that two neighbor points  $\mathbf{r}$  and  $\mathbf{r} + d\mathbf{r}$  have experienced *differential* displacement



**Fig. 7.6** Displacement field in a deformed continuum body

$du = \mathbf{u}(\mathbf{r} + d\mathbf{r}) - \mathbf{u}(\mathbf{r})$ , as illustrated in Fig. 7.6. However, even a description of the deformation in terms of differential displacements is not adequate. For example, if a metal bar having length  $l = 1$  m and fixed at one end is uniformly stretched to a length  $l' = 1.2$  m, the deformation could be described saying that the magnitude of the displacement vectors increases uniformly from zero at the fixed end to 0.2 m at the opposite side, thereby, we have differential displacements. However, if the bar were simply rotated about a hinge coinciding with the fixed end, the field of differential displacements would be different from zero anyway, despite this time the body has not changed its shape.

Furthermore, in the example mentioned above of a homogeneously stretched metal bar, a single number would be sufficient to describe this deformation, because the magnitude of the displacement vectors is always  $0.2x/l$ ,  $x$  being the distance from the fixed end. This example suggests that a better description of the deformation should be based on the *relative* variations of the displacement field, rather than on absolute changes. Let us consider a Taylor expansion of the displacement field, stopped at the first order:

$$u_i(\mathbf{r} + d\mathbf{r}) = u_i(\mathbf{r}) + \frac{\partial u_i}{\partial x_j} dx_j \quad (7.29)$$

With this approximation, we are limiting our attention to geologic processes that involve only *infinitesimal deformations*. The phenomena considered in seismology generally conform to this

hypothesis. In this instance, to the first order the components of the differential displacements field are given by:

$$du_i(\mathbf{r}) = \frac{\partial u_i}{\partial x_j} dx_j \quad (7.30)$$

In general, this field does not describe correctly the deformation, because it may include a rotational component not associated with changes of shape. To isolate the component of true deformation in (7.30), let us decompose the Jacobian  $J_{ij} = \partial u_i / \partial x_j$  into symmetric and antisymmetric parts:

$$J_{ij} = \varepsilon_{ij} + \omega_{ij} \quad (7.31)$$

where  $\varepsilon_{ij}$  are the components of a symmetric rank-two tensor,  $\varepsilon_{ij} = \varepsilon_{ji}$ , known as *strain tensor*.

The components of the strain tensor are given by:

$$\varepsilon_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \quad (7.32)$$

while the tensor  $\omega$ , which is antisymmetric ( $\omega_{ij} = -\omega_{ji}$ ), has components:

$$\omega_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i} \right) \quad (7.33)$$

It is not difficult to prove that  $\omega$  describes a rigid rotation without deformation. In fact, by its anti-symmetry,  $\omega$  has a null diagonal, so that there are only three independent components. Let us consider the vector  $\Omega$ , having components:

$$\Omega_k = \frac{1}{2} \varepsilon_{ijk} \omega_{ij} \quad (7.34)$$

where  $\varepsilon_{ijk}$  is the Levi-Civita tensor (see Appendix 1). Using the identity:

$$\varepsilon_{ijk} \varepsilon_{stk} = \varepsilon_{kij} \varepsilon_{kst} = \delta_{is} \delta_{jt} - \delta_{it} \delta_{js} \quad (7.35)$$

we find that:

$$\varepsilon_{ijk} \Omega_k = \frac{1}{2} \varepsilon_{ijk} \varepsilon_{stk} \omega_{st} = \frac{1}{2} (\omega_{ij} - \omega_{ji}) = \omega_{ij} \quad (7.36)$$

Therefore,

$$\omega_{ij} dx_j = \varepsilon_{ijk} \Omega_k dx_j = -(\boldsymbol{\Omega} \times d\mathbf{r})_i \quad (7.37)$$

This result shows that  $\boldsymbol{\omega}$  is a rigid rotation about an axis having the direction of  $\boldsymbol{\Omega}$ , so that no deformation is associated with the anti-symmetric part of the Jacobian matrix. As a consequence, any deformation is described by the strain tensor  $\boldsymbol{\varepsilon}$ . We note that the components of  $\boldsymbol{\varepsilon}$  are non-dimensional quantities that are calculated from partial derivatives of the displacement field. The diagonal components,  $\varepsilon_{kk}$ , represent variations of displacement components along the corresponding directions. For example,  $\varepsilon_{11} = \partial u_1 / \partial x_1$  represents the variation of the  $x$ -component of displacement as we move along the  $x$  axis.

Taking the trace of  $\boldsymbol{\varepsilon}$ , we obtain the divergence of the displacement field:

$$\Delta \equiv \varepsilon_{kk} = \nabla \cdot \mathbf{u} \quad (7.38)$$

This quantity is called *dilatation* and represents the volume change per unit volume during deformation. In fact, assuming that the strain tensor is diagonal, we have that a volume element  $dV = dx_1 dx_2 dx_3$  is changed as follows:

$$\begin{aligned} dV' &= \left(1 + \frac{\partial u_1}{\partial x_1}\right) dx_1 \left(1 + \frac{\partial u_2}{\partial x_2}\right) dx_2 \\ &\quad \times \left(1 + \frac{\partial u_3}{\partial x_3}\right) dx_3 \cong \left(1 + \frac{\partial u_k}{\partial x_k}\right) \\ &\quad \times dx_1 dx_2 dx_3 \\ &= \left(1 + \frac{\partial u_k}{\partial x_k}\right) dV = (1 + \Delta) dV \end{aligned} \quad (7.39)$$

Therefore, the relative variation of volume will be given by:

$$\Delta = \frac{dV' - dV}{dV} \quad (7.40)$$

Whenever it results  $\partial u_i / \partial x_i > 0$ , we have *extension* along the axis  $x_i$ . Conversely, for  $\partial u_i / \partial x_i < 0$  we have contraction. In general, the divergence

of the displacement field plays an important role in the quantitative description of the deformation process. Let us consider now the curl of  $\mathbf{u}$ :

$$\begin{aligned} \nabla \times \mathbf{u} &= \left(\frac{\partial u_3}{\partial x_2} - \frac{\partial u_2}{\partial x_3}\right) \mathbf{e}_1 + \left(\frac{\partial u_1}{\partial x_3} - \frac{\partial u_3}{\partial x_1}\right) \mathbf{e}_2 \\ &\quad + \left(\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2}\right) \mathbf{e}_3 \end{aligned} \quad (7.41)$$

It is not difficult to prove that this quantity is associated only with rigid rotations. This results promptly from a comparison of expressions (7.41) and (7.33). Therefore,  $\nabla \times \mathbf{u}$  is non-zero only when the displacement field includes a component of rigid rotation. A deformation that can be described by a traceless ( $\varepsilon_{kk} = 0$ ) strain tensor leaves invariant the volume of a body and is referred to as a *pure shear*. A simple example of pure shear deformation is illustrated in Fig. 7.7d. Let us consider now the off-diagonal components of  $\boldsymbol{\varepsilon}$ , which arise from variations of displacement components along transversal axes. For example,  $\varepsilon_{12} \neq 0$  when the  $x$  component of  $\mathbf{u}$  varies as we move in the  $y$  direction, or when  $u_y$  changes as we move in the  $x$  direction, both possibilities being admissible at the same time. When in a pure shear deformation the strain tensor diagonal is identically zero and the deformation arises from a single pair of off-diagonal components (e.g.,  $\varepsilon_{12}$  and  $\varepsilon_{21}$ ), a simultaneous rotation about an axis orthogonal to the plane of deformation may lead to the situation illustrated in Fig. 7.7c, which is termed *simple shear*.

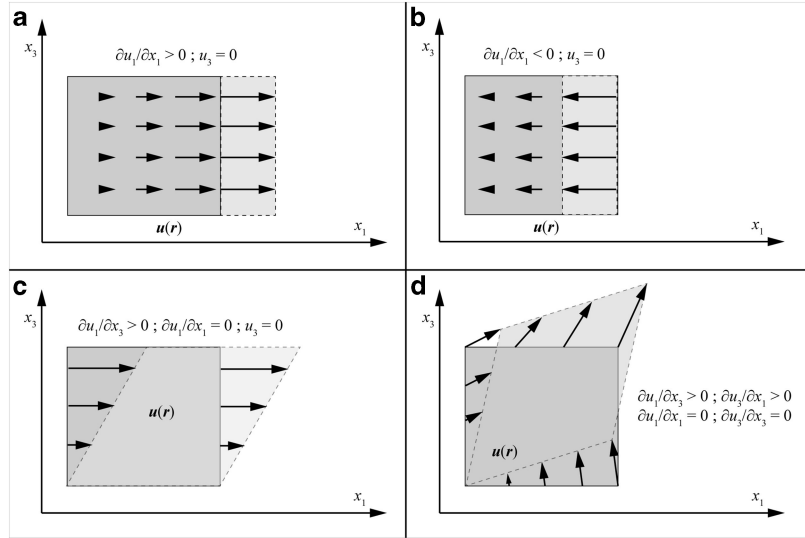
In simple shear, two orthogonal segments with a common fixed point, initially parallel to coordinate axes, assume new orientations after deformation (Fig. 7.7c), with angles  $\phi_1$  and  $\phi_2$  with respect to the coordinate axes. For an infinitesimal deformation in the  $xz$  plane, it results:

$$\phi_1 \cong \tan \phi_1 = \frac{\partial u_1}{\partial x_3} ; \quad \phi_2 \cong \tan \phi_2 = \frac{\partial u_3}{\partial x_1} \quad (7.42)$$

Therefore,

$$\phi_1 + \phi_2 \cong \frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} = 2\varepsilon_{13} \quad (7.43)$$

**Fig. 7.7** Geometry of deformation in two dimensions. **a** Dilatation, **b** Compression, **c** Simple shear, **d** Pure shear



This expression says that the off-diagonal component  $\varepsilon_{13}$  is the average angular variation in the plane  $x_1x_3$ . A similar conclusion can be drawn for the other off-diagonal components. Therefore, (7.43) furnishes a simple intuitive interpretation of the off-diagonal components of the strain tensor.

A consequence of the strain tensor symmetry is that a system of *principal strain axes*  $\mathbf{m}^i$  exists such that the tensor is diagonal. In this frame, and assuming no rotations, the variations of displacement  $du_i$  have the same direction of the variations of position  $dx_i$ :

$$du_i = \varepsilon_{ij} dx_j = \lambda dx_i \quad (7.44)$$

The three eigenvectors of (7.44) are called the *principal strains*  $\mathbf{e}_1$ ,  $\mathbf{e}_2$ , and  $\mathbf{e}_3$ . With the exception of a situation of *hydrostatic strain*, such that  $\mathbf{e}_1 = \mathbf{e}_2 = \mathbf{e}_3$ , some amount of shear strain is always present also in the principal strain axes coordinate system.

### 7.3 Cauchy Momentum Equation

In the previous sections, we have introduced the concepts of traction, stress, displacement, and strain for a continuum body in conditions of static equilibrium. Now we are going to describe the

relation between these variables in a geodynamic context. This can be done adapting the classic Newton's equations of motion (second law of mechanics) to the case of a continuum deformable body. Let us consider the forces exerted on a volume element  $dV = dx_1 dx_2 dx_3$  of the region  $\mathbf{R}$  occupied by the body (Fig. 7.8).

We know that the surface force exerted on a face of  $dV$  is given by the traction on that face, times the area. Therefore, the force  $d\mathbf{F}$  on a face at position  $\mathbf{r} = (x_1, x_2, x_3)$ , with normal  $-\mathbf{e}_j$  and area  $dx_r dx_s$  ( $r, s \neq j$ ), has components:

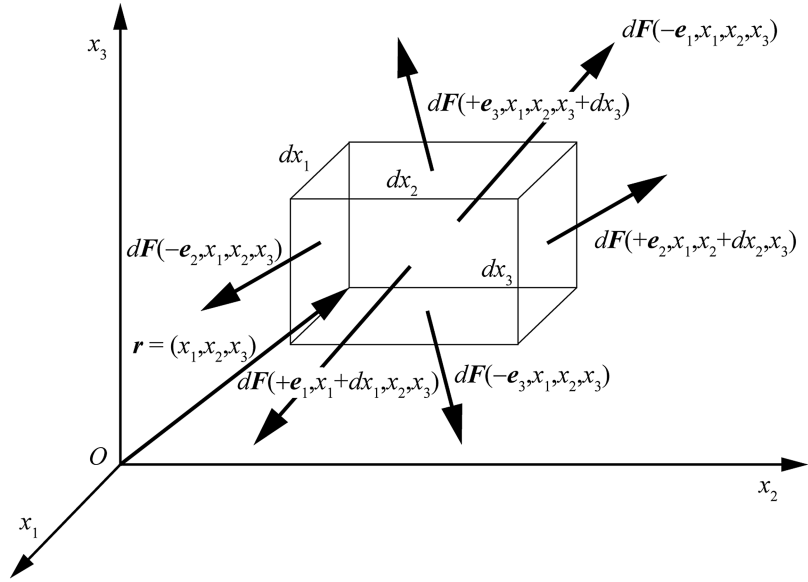
$$\begin{aligned} dF_i(-\mathbf{e}_j, \mathbf{r}) &= T_{ij}(-\mathbf{e}_j) dx_r dx_s \\ &= -\tau_{ij}(\mathbf{r}) dx_r dx_s; \quad r, s \neq j \end{aligned} \quad (7.45)$$

Similarly, the force on a face at position  $\mathbf{r} + \mathbf{e}_j dx_j$ , with normal  $\mathbf{e}_j$  and area  $dx_r dx_s$  ( $r, s \neq j$ ), has components:

$$\begin{aligned} dF_i(\mathbf{e}_j, \mathbf{r} + \mathbf{e}_j dx_j) &= \tau_{ij}(\mathbf{r} + \mathbf{e}_j dx_j) \\ &\quad \times dx_r dx_s; \quad r, s \neq j \end{aligned} \quad (7.46)$$

Clearly, when the stress field is homogeneous the net force exerted on  $dV$  is zero, because the forces on opposite sides of the volume element balance each other:  $d\mathbf{F}(-\mathbf{e}_j, \mathbf{r}) = -d\mathbf{F}(\mathbf{e}_j, \mathbf{r} + \mathbf{e}_j dx_j)$ . Therefore, in

**Fig. 7.8** Forces on a volume element  $dV = dx_1 dx_2 dx_3$  in equilibrium conditions



order to have a net force on  $dV$ , a non-zero spatial gradient of the stress field must exist at the location of  $dV$ . In this instance, to the first order the net force on the faces normal to the  $x_j$  axis is given by:

$$\begin{aligned} \delta F_i(\mathbf{e}_j) &= dF_i(\mathbf{e}_j, \mathbf{r} + \mathbf{e}_j dx_j) + dF_i(-\mathbf{e}_j, \mathbf{r}) \\ &= [\tau_{ij}(\mathbf{r} + \mathbf{e}_j dx_j) - \tau_{ij}(\mathbf{r})] dx_r dx_s \\ &= \left[ \frac{\partial \tau_{ij}}{\partial x_j} dx_j \right] dx_r dx_s = \frac{\partial \tau_{ij}}{\partial x_j} dV; \\ &\quad \times r, s \neq j; \text{no summation on } j \end{aligned} \quad (7.47)$$

Therefore, considering the net force exerted on all pairs of faces, we have that the total surface force *per unit volume* acting on  $dV$  has components:

$$f_i^{surf} = \frac{\partial \tau_{ij}}{\partial x_j} \quad (7.48)$$

Let us introduce now the *body force density*,  $\mathbf{f} = \mathbf{f}(\mathbf{r})$ , exerted on the volume element  $dV$  (see Sect. 2.1), which also gives a contribution to the force on  $dV$ :  $d\mathbf{F} = \mathbf{f}(\mathbf{r})dV$ . Then, the total force per unit volume will be given by:

$$f_i^{tot} = \frac{\partial \tau_{ij}}{\partial x_j} + f_i \quad (7.49)$$

To apply the second law of mechanics, we must balance the total force by an inertial term  $\ddot{u}_i dm$ , where  $dm = \rho dV$  is the mass of the volume element and the second time derivative of the displacement field represents, in the context of continuum mechanics, the analogue of the point mass acceleration. Therefore, considering force densities, the equations of motion can be written as follows:

$$\rho \ddot{u}_i = \frac{\partial \tau_{ij}}{\partial x_j} + f_i \quad (7.50)$$

This is the fundamental equation of dynamics for continuous media. It is often referred to as the *Cauchy momentum equation*. In seismology, it is generally possible to neglect the contribution of body forces in absence of seismic sources. Furthermore, in this context the second time derivative of the displacement can be calculated as a partial derivative, because the location of a volume element does not change significantly during earthquakes (this is not true in fluid dynamics). In this instance, the momentum equation reduces to the following *homogeneous equation of motion*:

$$\rho \frac{\partial^2 u_i}{\partial t^2} - \frac{\partial \tau_{ij}}{\partial x_j} = 0 \quad (7.51)$$

In Chap. 8, we shall see that this is the governing equation for the propagation of seismic waves outside source regions.

## 7.4 Basic Rheological Models and Constitutive Equations

Apart from being in one of the four ordinary *states of matter*, solid, liquid, gas, or plasma, a material has a mechanical behaviour that depends from its *rheology*. By “mechanical behaviour”, we mean a distinct relationship between the stress and strain fields within a body, which can be expressed by a *constitutive equation*:

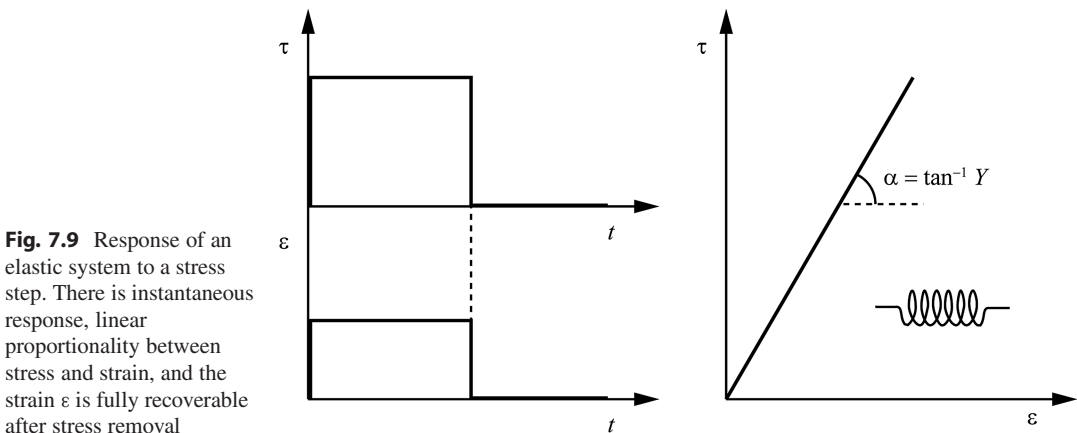
$$R(\boldsymbol{\varepsilon}, \dot{\boldsymbol{\varepsilon}}, \ddot{\boldsymbol{\varepsilon}}, \dots, \boldsymbol{\tau}, \dot{\boldsymbol{\tau}}, \ddot{\boldsymbol{\tau}}, \dots, \times m_1, m_2, \dots, s_1, s_2, \dots) = 0 \quad (7.52)$$

The function  $R$  is referred to as the *rheological function* and usually includes a set of *intrinsic parameters*  $m_i$  that depend from the material (e.g., elastic moduli, viscosity) and a set of *state variables*  $s_j$  describing the microstructural state of the material (e.g., grain size, see Ranalli 1995). The word “Rheology” originates from Greek and means “study of flow”. A good quantitative introduction to this subject can be found in Ranalli (1995). Many materials display proportionality between stress and strain for small deformations. However, *all* materials, independently from their

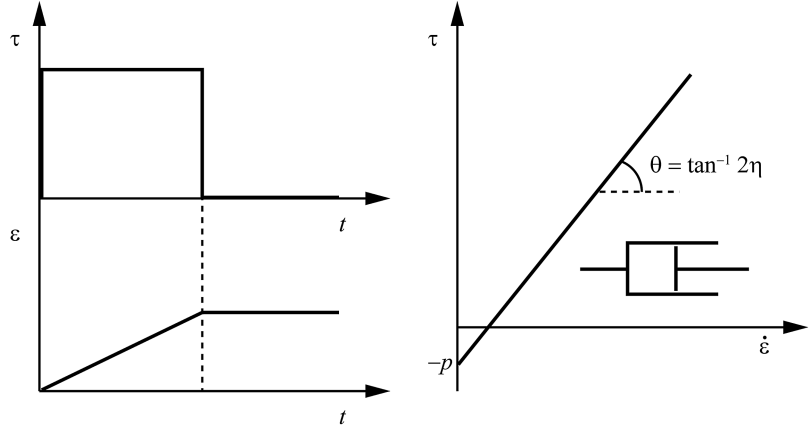
physical state, may *flow* under certain thermo-mechanical conditions, showing a mechanical behaviour typical of liquids, gases and plasma. In this instance, a material is subject to continuous deformation under an applied *shear* stress. An understanding of the different rheological behaviors within the solid Earth is fundamental in geology in general, but particularly in plate tectonics, where the different interacting subsystems of crust and mantle display distinct mechanical behaviors. The simplest constitutive equation is that associated with the *elastic behavior*, which approximates adequately the relationship between stress and strain in solids at low pressure and temperature and for small deformations, but also in non-solids in the case of high-frequency variations of the stress field. For example, the lithosphere can be considered as an elastic body in so far as it is not subducted, but also lower mantle rocks and the outer core, which is liquid, can be modelled as elastic bodies at the frequency of deformations associated with earthquakes. A paradigm for this kind of mechanical behavior is given by *springs*. In this instance, the constitutive equation expressing the relation between stress and strain is *Hooke's law*:

$$\tau(t) = Y\varepsilon(t) \quad (7.53)$$

where  $t$  is the time and the parameter  $Y$ , which depends from the material, is known as the *time-independent elastic modulus*. Figure 7.9 shows the ideal response of an elastic system to a stress step.



**Fig. 7.10** Response of a linear viscous system to a stress step. An analog model for this system is the dashpot. There is continuous deformation at constant rate under constant stress, and the strain  $\epsilon$  is permanent after stress removal

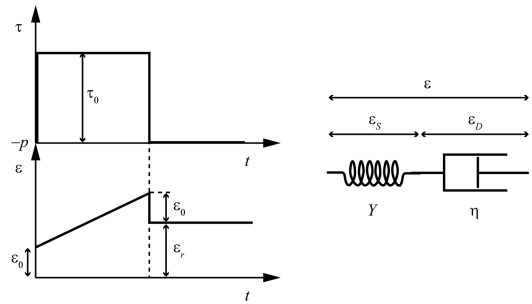


Another theoretical response of a material to stress steps, which can eventually combine with the elastic response, is associated with the *viscous behavior*. In this instance, the constitutive equation relating stress and strain rate expresses proportionality between stress and *strain rate* rather than strain:

$$\tau(t) = 2\eta\dot{\epsilon}(t) - p \quad (7.54)$$

where  $p$  is the hydrostatic pressure and the constant parameter  $\eta$ , which depends from the material and is a decreasing function of temperature, is called *viscosity* and has units [Pa s]. The time derivative of strain,  $\dot{\epsilon}$ , has units [ $s^{-1}$ ] and gives a measure of the rapidity of deformation. In the case of non-solid materials, where  $\epsilon$  represents pure shear strain, the constitutive law (7.54) describes the behavior of *Newtonian fluids*.

The linear viscous behavior is illustrated in Fig. 7.10. These materials do not display strain recovery and deform continuously under constant stress. Just as a spring furnishes an analog model for elastic systems, a *dashpot* can be used as a paradigm for linear viscous materials. However, many materials of geologic interest have a mechanical behavior that can be considered as a combination of linear viscous and elastic rheologies. They are termed *viscoelastic materials*. A simple description of their behavior can be obtained using analog models. These models are built combining springs and dashpots into complex systems. Here we are going



**Fig. 7.11** Analog model for the Maxwell rheology

to describe some simple rheological models commonly used in geodynamic modelling. The first of them is the *Maxwell rheology* model, which is a two-element system consisting of a linear spring element and a dashpot connected in series, as shown in Fig. 7.11.

In this model, the total *strain* is partitioned into a spring strain,  $\epsilon_S$ , and a dashpot strain,  $\epsilon_D$ , so that:

$$\dot{\epsilon}(t) = \dot{\epsilon}_S(t) + \dot{\epsilon}_D(t) \quad (7.55)$$

where  $\epsilon_S$  and  $\epsilon_D$  are related to the applied stress  $\tau$  by the following equations:

$$\tau(t) = Y\epsilon_S(t)$$

$$\tau(t) = 2\eta\dot{\epsilon}_D(t) - p \quad (7.56)$$

Inserting these expressions into (7.55) gives:

$$\dot{\epsilon}(t) = \frac{1}{Y}\dot{\tau}(t) + \frac{1}{2\eta}[\tau(t) + p] \quad (7.57)$$

This is the constitutive equation of the Maxwell rheology model, which can be solved under different stress conditions to determine strain-time behavior. For example, let us assume that a stress step with magnitude  $\tau = \tau_0$  has been applied from  $t = 0$  to  $t = t_0$ , as shown in Fig. 7.11. At the initial time, we have  $\varepsilon(0) = \varepsilon_S = \tau_0/Y$ .

For  $0 \leq t \leq t_0$ , it results  $\dot{\tau} = 0$ , thereby, (7.57) reduces to:

$$\dot{\varepsilon}(t) = \frac{1}{2\eta} [\tau(t) + p] \quad (7.58)$$

Integration of this equation gives:

$$\varepsilon(t) = \begin{cases} \frac{1}{2\eta} \tau_0 t + \frac{\tau_0}{Y}; & 0 \leq t < t_0 \\ \frac{1}{2\eta} \tau_0 t_0; & t \geq t_0 \end{cases} \quad (7.59)$$

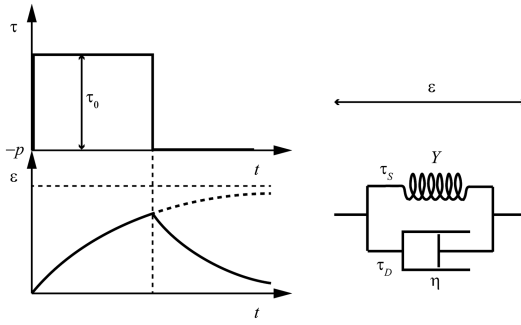
Therefore, this system shows an unrecoverable strain  $\varepsilon_r = \tau_0 t_0 / 2\eta$ . Maxwell's viscoelastic materials have an immediate elastic response, but ultimately behave as linear Newtonian fluids. The first applications of this model in geodynamics go back to the 1970s (e.g., Wang et al. 2012 and references therein). The objective was to give a representation of the Earth's mantle that allowed the elastic transmission of stress associated with earthquakes, yet preserving the fluid behavior required by mantle convection and the delayed response to the removal of surface ice loads (postglacial rebound). The viscous component of a Maxwell viscoelastic model describes the steady-state *creep* of materials under constant stress. In this context, the word "creep" refers to a continuous deformation (flow) of the material without formation of breaks. The Maxwell rheology furnishes a good approximation of the century- to millennium scale glacial isostatic adjustment and of the evolution of the stress and strain fields during the interseismic interval between two earthquakes. An important feature of Maxwell rheological models is a phenomenon known as *stress relaxation*. Let us assume that  $\varepsilon(t) = \varepsilon_0$  for  $t \geq 0$  and  $\tau(0) = \tau_0$ . By (7.57), we have that:

$$\dot{\tau}(t) + \frac{Y}{2\eta} \tau(t) = -\frac{Yp}{2\eta} \quad (7.60)$$

The solution to this equation is an exponential decay curve to the hydrostatic pressure, with relaxation time  $t_R = 2\eta/Y$ :

$$\tau(t) = (\tau_0 - p) e^{-Yt/2\eta} + p \quad (7.61)$$

Therefore, in conditions of constant strain, a Maxwell viscoelastic material exponentially relaxes the internal stress. Studies of a process known as *postglacial rebound* have allowed an estimation of the mantle relaxation time. During the last (Pleistocene) glaciation, the load of continental ice sheets at high latitudes resulted into a downward bending of the lithosphere, accompanied by a peripheral bulge at some distance from the glacier margins. For example, the thick ice sheet that covers Greenland has depressed the surface of this continent several kilometers, below the sea level at some places. The vertical motion associated with such bending was clearly accompanied by corresponding lateral flow in the asthenosphere. The subsequent melting of the ice sheet during the Holocene determined unloading and a gradual restoration of the isostatic equilibrium through upward bending of the lithosphere. A simple mathematical formulation of this process can be found in Turcotte and Schubert (2002). The rate of rebound, which has been determined by the radiocarbon dating of elevated beach terraces along the coastlines of Canada and Scandinavia, and by Uranium-Thorium dating of coral reef deposits, can be used to determine the viscosity of the asthenosphere, hence its relaxation time. Recent estimates constrain the mean viscosity of the asthenosphere and the transition zone to be  $\approx 0.5 \times 10^{21}$  Pa s, the upper 500 km of the lower mantle to be  $\approx 1.6 \times 10^{21}$  Pa s, and the remainder of the mantle to be  $\approx 3.2 \times 10^{21}$  Pa s (Argus and Peltier 2010). Assuming an appropriate elastic modulus  $Y = 70$  GPa (e.g., Schubert et al. 2001), we obtain that an estimate of the relaxation time for the upper mantle is  $t_R \sim 453$  years. This value of  $t_R$  provides an explanation for the elastic response of the mantle to the propagation of seismic waves,



**Fig. 7.12** Analog model for the Kelvin rheology

which involves displacements on a time scale of a few tens seconds, and the simultaneous fluid behavior on a time scale of several tens Myrs in the context of mantle convection. In general, the prevalence of elastic behavior with respect to the viscous response depends from the relative importance of the parameters  $Y$  and  $\eta$ . For large values of the elastic modulus, the material behaves as a viscous fluid over long time intervals, whereas for  $\eta \rightarrow \infty$  it behaves as an elastic solid for short-duration loads.

Another important rheological model is the *Kelvin rheology* model, which is also a two-element system consisting of a linear spring element and a dashpot, but this time the two elements are connected in parallel, as shown in Fig. 7.12. In this instance, the total *stress* is partitioned between the spring and the dashpot, so that:

$$\tau(t) = \tau_S(t) + \tau_D(t) \quad (7.62)$$

where  $\tau_S$  and  $\tau_D$  are related to the common strain  $\epsilon$  by the following equations:

$$\begin{aligned} \tau_S(t) &= Y\epsilon(t) \\ \tau_D(t) &= 2\eta\dot{\epsilon}(t) - p \end{aligned} \quad (7.63)$$

Therefore:

$$\dot{\epsilon}(t) + \frac{Y}{2\eta}\epsilon(t) = \frac{1}{2\eta}[\tau(t) + p] \quad (7.64)$$

This is the constitutive equation associated with the Kelvin rheology model. For a stress

pulse, such that  $\tau(t) + p = \tau_0$  for  $0 \leq t \leq t_0$ , and assuming the initial condition  $\epsilon(0) = 0$ , we have:

$$\epsilon(t) = \frac{\tau_0}{Y} (1 - e^{-Yt/2\eta}) \quad (7.65)$$

Therefore, the total strain increases progressively (at decreasing rate) towards the asymptotic value  $\epsilon(\infty) = \tau_0/Y$  (Fig. 7.12). This behavior implies that the load is initially supported by the viscous element only, which then elongates and transfers an increasing amount of stress to the spring. After a sufficiently long time interval, most of the stress is supported by the elastic element. The strain rate associated with creep is easily obtained by (7.65):

$$\dot{\epsilon}(t) = \frac{\tau_0}{2\eta} e^{-Yt/2\eta} \quad (7.66)$$

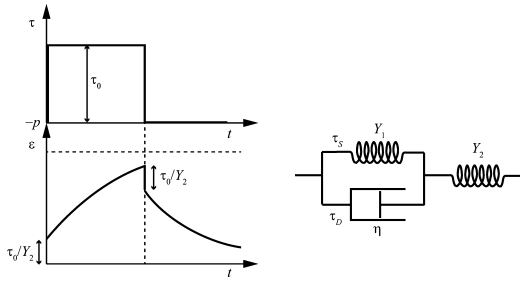
Therefore, the strain rate decays exponentially to zero. Once we have removed the load, at some time  $t = t_0$ , the subsequent evolution of the strain field can be determined using the superposition principle. To set the total stress to zero, we combine the initial stress,  $\tau_0$ , with an additional stress field  $\tau(t) = -\tau_0$ , for  $t \geq t_0$ . The strain associated with the initial stress is given by (7.65), while that resulting from the additional stress will be given by:

$$\epsilon'(t) = \frac{\tau_0}{Y} (e^{-Y(t-t_0)/2\eta} - 1); \quad t \geq t_0 \quad (7.67)$$

Then, for  $t \geq t_0$  the total strain follows an exponential recovery curve:

$$\epsilon(t) = \frac{\tau_0}{Y} (e^{Yt_0/2\eta} - 1) e^{-Yt/2\eta}; \quad t \geq t_0 \quad (7.68)$$

This solution shows that a Kelvin material has full strain recovery after load removal. Neither of the Maxwell and Kelvin models furnishes an accurate description of the viscoelastic rheology. For example, the Kelvin model does not exhibit instantaneous strain after loading or unloading, and does not account for permanent strain after relaxation. Similarly, the Maxwell model does not allow time-dependent recovery and predicts constant strain rate under constant stress, whereas



**Fig. 7.13** Analog model for the standard solid rheology

the creep of most viscoelastic materials exhibits decreasing strain rate for constant load.

An improvement over Maxwell and Kelvin rheologies is given by the *standard solid model* (or *Zener model*) illustrated in Fig. 7.13. This system adds instantaneous elastic response under loading or unloading to a normal Kelvin rheology, but it still lacks of permanent strain after transient creep. A widely used rheological model, which provides a better representation of the viscoelastic behavior, is the *Burgers model*, which puts in series a Kelvin and a Maxwell model (Fig. 7.14).

The total strain associated with the model of Fig. 7.14 is:

$$\varepsilon(t) = \varepsilon_K(t) + \varepsilon_D(t) + \varepsilon_S(t) \quad (7.69)$$

where,

$$\begin{aligned} \varepsilon_S(t) &= \frac{1}{Y_2} \tau(t); \quad \dot{\varepsilon}_D(t) + \frac{1}{2\eta_2} [\tau(t) + p]; \\ \dot{\varepsilon}_K(t) + \frac{Y_1}{2\eta_1} \varepsilon_K(t) &= \frac{1}{2\eta_1} [\tau(t) + p] \end{aligned} \quad (7.70)$$

It is possible to show (e.g., Findley et al. 1989) that the resulting constitutive equation is:

$$\begin{aligned} \tau(t) + p + 2 \left( \frac{\eta_2}{Y_1} + \frac{\eta_2}{Y_2} + \frac{\eta_1}{Y_1} \right) \dot{\tau}(t) + 4 \frac{\eta_1 \eta_2}{Y_1 Y_2} \ddot{\tau}(t) \\ = 2\eta_2 \dot{\varepsilon}(t) + 4 \frac{\eta_1 \eta_2}{Y_1} \ddot{\varepsilon}(t) \end{aligned} \quad (7.71)$$

To determine the creep curve of the Burgers model, we set the initial strain conditions for a stress step such that  $\tau(0) = \tau_0$ :

$$\begin{aligned} \varepsilon(0) = \varepsilon_S(0) &= \frac{\tau_0}{Y_2}; \quad \varepsilon_K(0) = \varepsilon_D(0) = 0 \\ \dot{\varepsilon}(0) &= \frac{1}{2} \left( \frac{1}{\eta_1} + \frac{1}{\eta_2} \right) \tau_0 \end{aligned} \quad (7.72)$$

With these initial conditions, the solution to (7.71) is simply the sum of the strain of a Maxwell rheology component plus the strain of a Kelvin element:

$$\varepsilon(t) = \frac{\tau_0}{Y_2} + \frac{\tau_0}{2\eta_2} t + \frac{\tau_0}{Y_1} (1 - e^{-Y_1 t / 2\eta_1}) \quad (7.73)$$

In this model, when the load is removed there is instantaneous recovery of the elastic strain associated with the Maxwell element. This is followed by transient creep strain, but permanent residual deformation results for  $t \rightarrow \infty$ . In general, the Burgers body exhibits instantaneous elastic response after loading or unloading, permanent strain after relaxation, time-dependent recovery, and decreasing transient strain rate under constant stress. Over the long timescale, it exhibits linear viscous behaviour. Therefore, it provides a unifying model for the rheology of the Earth's mantle over the whole frequency spectrum. Differentiating (7.73) gives the strain rate under constant load:

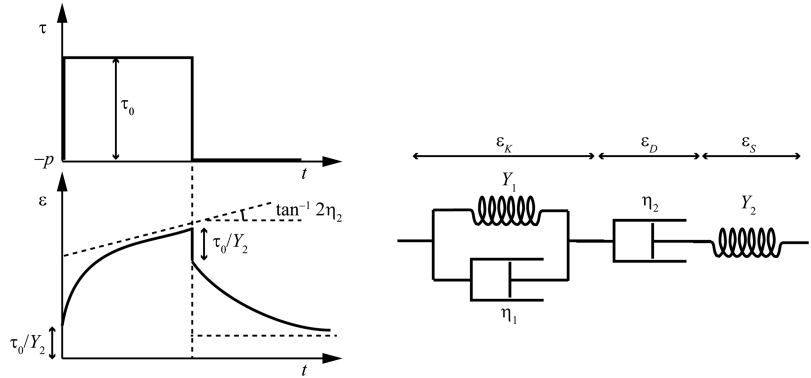
$$\dot{\varepsilon}(t) = \frac{\tau_0}{2} \left( \frac{1}{\eta_2} + \frac{1}{\eta_1} e^{-Y_1 t / 2\eta_1} \right) \quad (7.74)$$

Therefore, the initial rate of creep for  $t = 0^+$  and the asymptotic value for  $t \rightarrow \infty$  are:

$$\dot{\varepsilon}(0^+) = \frac{\tau_0}{2} \left( \frac{1}{\eta_2} + \frac{1}{\eta_1} \right); \quad \dot{\varepsilon}(\infty) = \frac{\tau_0}{2\eta_2} \quad (7.75)$$

The delayed response of Burgers systems (but also of Kelvin and Zener models) to a pulse train is usually indicated as *anelastic behaviour*. In this instance, part of the elastic strain energy is dissipated as heat. Microscopically, the anelasticity is associated with slipping along grain boundaries and internal friction. In the frequency domain, transient creep is partly responsible for the attenuation (i.e., the damping) of seismic waves. The magnitude of this attenuation depends

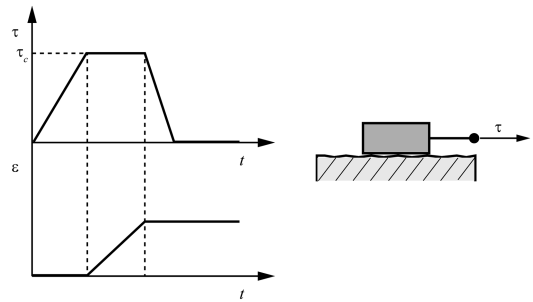
**Fig. 7.14** Analog model for the burgers rheology



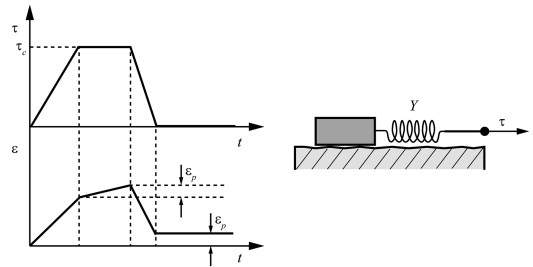
from environmental parameters such as temperature, pressure, and the frequency of the propagating wave. Therefore, the study of anelasticity gives valuable information about the physical conditions and the rheological parameters of the Earth's mantle.

We conclude this review of rheological models with a short mention to the *plastic rheology* of solids. Plasticity occurs when a body does not deform until the stress has attained a critical value  $\tau_c$ , which is referred to as the *yield stress*. Starting from this point, a plastic body deforms continuously under constant stress, thereby, we can say that it cannot support a stress greater than the critical value  $\tau_c$ . At the yield stress, deformation is permanent and irreversible and proceeds at constant rate. Therefore, ideally the amount of deformation is unlimited as long as the yield stress is maintained, and the strain rate is independent of stress. Plastic strain is a form of shear strain, thereby, it can only be associated with shear stress.

An analog model for the plastic rheology is given by a mass at rest on a flat and rough surface (*Saint-Venant body*) (Fig. 7.15). Static friction prevents displacement of the mass if the applied stress  $\tau$  is less than the frictional resistance. As soon as  $\tau$  overcomes the frictional resistance, the mass starts moving. A key feature of plastic deformation is that it implies *strain localization*, as strain may take place only locally, in regions where the critical value of stress is reached. Therefore, it is possible to use the plastic rheology to build analog models describing the brittle and ductile behaviours of the upper crust and the



**Fig. 7.15** Response of a plastic body to a stress step. An analog model for this system is a mass at rest on a rough surface. There is continuous deformation at constant rate for  $\tau = \tau_c$



**Fig. 7.16** Analog model for the visco-plastic rheology.  $\epsilon_p$  is the amount of plastic strain

upper part of subducting slabs, where deformation occurs either by seismic slip (associated with faulting) or strain localization without loss of continuity. In the simplest model, we combine the plastic behaviour with an instantaneous elastic response by the introduction of a spring element, as shown in Fig. 7.16. Such a rheological model is called the *Prandtl body*, and the corresponding mechanical behavior is referred to as the

*elasto-plastic rheology*. When the stress is below the yield strength, the system stretches the spring and there is no slip along the contact surface. In these conditions, the material deforms elastically. As soon as the stress attains the critical value, the mass is pulled at constant stress and deformation proceeds plastically until the load falls below the yield stress  $\tau_c$ . Apparently, this behaviour is quite similar to the viscous flow. However, differently from visco-elastic deformation, plastic flow occurs only when the yield stress is attained. It is termed *ductile deformation* when the continuity within the material persists despite strain localization. Conversely, when the material loses its cohesion through the development of fractures or faults we will use the term *brittle deformation*.

The mode of failure of Earth's rocks mainly depends from the confining hydrostatic pressure and from temperature, while the yield point is a decreasing function of the strain rate. The effect of increased hydrostatic pressure is to inhibit fracturing and cracking. Therefore, in so far as the hydrostatic pressure is increased, the mechanical behaviour changes suddenly from brittle to ductile. The range of depths at which this transition occurs is called the *brittle-ductile transition zone* and depends not only from  $(P, T)$  conditions but also from the strain rate and the presence of water. In general, the rheology of Earth's lithosphere rocks changes with increasing depth from brittle to ductile to visco-elastic. The mechanical behaviour is usually represented graphically through strength (yield stress) versus depth profiles that are called *rheological profiles* or *yield-strength envelopes*. Now we are going to describe the general methods adopted to build these profiles.

The static frictional resistance along fault planes is generally constant, so that by *Amonton's law* the coefficient of static friction,  $\mu_s$ , determines the shear stress required to have seismic slip:

$$\tau_s = \mu_s \sigma_N \quad (7.76)$$

where  $\tau_s$  and  $\sigma_N$  are respectively the shear stress and the normal stress along the fault plane. Amonton's law implies that the shear stress

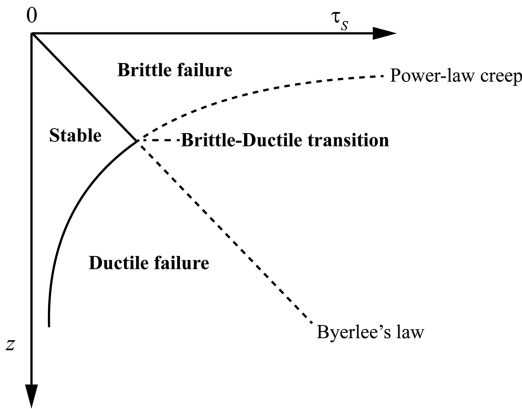
required for seismic slip is independent from the contact area and increases linearly with the confining pressure. Byerlee (1968) determined experimentally the following linear relations over a range of normal stresses from 3 MPa to 1.7 GPa:

$$\tau_s = \begin{cases} 0.85\sigma_N; & 3 < \sigma_N < 200 \text{ MPa} \\ 60 \pm 10 + 0.6\sigma_N; & \sigma_N > 200 \text{ MPa} \end{cases} \quad (7.77)$$

These simple relations hold for all geologic materials except certain clay minerals (e.g., Brace and Kohlstedt 1980). They can be used to build the upper part of a rheological profile, assuming that  $\sigma_N$  corresponds to the hydrostatic pressure. In this instance, the strength increases linearly with depth and is independent from the specific kind of rocks. The lower part of a rheological profile for the lithosphere is built taking into account that Earth's rocks exhibit non-linear viscous behaviour at relatively low temperatures. In fact, the linear visco-elastic rheological models discussed above provide good approximations of the real mechanical behaviour of Earth's rocks only at lower lithosphere and asthenosphere conditions, that is, temperatures between 1,000 °C and 1,500 °C and slow strain rates ( $10^{-12}$ – $10^{-14}$  s $^{-1}$ ). Conversely, in the case of upper lithospheric mantle and crustal rocks, we observe steady-state flow even at small stresses, but the relation between stress and strain rate is nonlinear. In this instance, the viscosity is a decreasing exponential function of temperature through an Arrhenius relationship and the stress raised to some power (usually between 3 and 5) is proportional to the strain rate. The most important empirical non-linear relation between stress and strain rate is known as the *power-law creep* or *Dorn equation*:

$$\dot{\epsilon} = A \tau_s^n e^{-E/RT} \quad (7.78)$$

where  $n$  is determined experimentally and depends from the material,  $A$  is a constant depending from the material and the  $(P, T)$  conditions,  $T$  is the temperature,  $R$  is the gas constant, and the *thermal activation energy*  $E$  is of the order of 100–600 kJ mol $^{-1}$  depending from the material.



**Fig. 7.17** A simple rheological profile, illustrating the variations of yield stress with depth (*solid line*)

This non-Newtonian flow behaviour is very common in silicate polycrystals at high temperature and low stresses ( $T > T_s/2$  and  $\tau_s$  between 10 and 100 MPa). The lower part of a rheological profile for the lithosphere shows the rock strength in a context of ductile deformation. It is largely insensitive to hydrostatic pressure variations but decreases exponentially with depth due to thermal softening.

At any given depth, the *strength* is defined as the *lowest* between brittle and ductile yield stresses. Therefore, inverting (7.78) for  $\tau_s$  and assuming a temperature versus depth relation (see Chap. 12) allows to build a lithospheric strength profile like that illustrated in the example of Fig. 7.17. More complex profiles can be built for different tectonic environments assuming specific crustal compositions, lithospheric layering, and temperature field (e.g., Ranalli and Murphy 1987).

## Problems

1. Find the principal axes for the 2-D stress tensor:  $\tau = \begin{bmatrix} 20 & \sqrt{125} \\ \sqrt{125} & 40 \end{bmatrix}$  MPa and determine: (1) the components of traction on a vertical fault oriented E–W; (2) the plane of maximum shear stress and the shear and normal stresses along this plane;

2. Determine the body force field in conditions of static equilibrium for the following stress tensor:

$$\tau(x, y, z)$$

$$= \begin{bmatrix} -3x^2 + 5y - 2z^3 & 4x + 3xy & 3xz \\ 4x + 3xy & 2x^3 - 4y & 0 \\ 3xz & 0 & -\frac{3}{2}z^2 \end{bmatrix}$$

and comment the result;

3. Consider the following transformation of the stress tensor:

$$\tau_{ij} \rightarrow \tau_{ij} + a\delta_{ij}$$

where  $a$  is a constant. What is the effect on the principal axes?

4. What is the relation between the principal axes of the stress tensor and those of the deviator?
5. Find a constitutive equation for the standard solid rheology and determine the creep and relaxation curves.

## References

- Argus DF, Peltier WR (2010) Constraining models of postglacial rebound using space geodesy: a detailed assessment of model ICE-5G (VM2) and its relatives. *Geophys J Int* 181(2):697–723
- Brace WF, Kohlstedt DL (1980) Limits on lithospheric stress imposed by laboratory experiments. *J Geophys Res* 85(B11):6248–6252
- Byerlee JD (1968) Brittle-ductile transition in rocks. *J Geophys Res* 73(14):4741–4750
- Findley WN, Lai JS, Onaran K (1989) Creep and relaxation of nonlinear viscoelastic materials. Dover Books, London, p 371
- Goldstein H (1980) Classical mechanics, 2nd edn. Addison-Wesley, Reading, p 672
- Ranalli G (1995) Rheology of the earth, 2nd edn. Chapman & Hall, London, p 413
- Ranalli G, Murphy DC (1987) Rheological stratification of the lithosphere. *Tectonophysics* 132(4): 281–295
- Schubert G, Turcotte D, Olson P (2001) Mantle convection in the earth and planets. Cambridge University Press, Cambridge, p 940
- Turcotte DL, Schubert G (2002) Geodynamics, 2nd edn. Cambridge University Press, Cambridge, p 848
- Wang K, Hu Y, He J (2012) Deformation cycles of subduction earthquakes in a viscoelastic earth. *Nature* 484(7394):327–332

## Abstract

An important source of data in plate tectonics comes from seismology. This is the first of three chapters devoted to the fundamental laws of propagation of seismic waves. In this chapter, I describe the elastic response of Earth's rocks to deformation, which is quantified by Hooke's law, the seismic wave equation, and the concept of seismic energy.

## 8.1 Hooke's Law

In this chapter, we consider the elastic behaviour of Earth's rocks in response to dynamic loads at a time scale of a few tens seconds. In the context of plate tectonics, slip along plate boundaries is the ultimate cause of most earthquakes, but are the elastic properties of the rocks to determine the mode of propagation of seismic waves through the Earth. Therefore, these properties constitute the "hard background" for the study of seismology. The basic formulation of this subject assumes that the elasticity of rocks does not depend on direction, so that the material is *isotropic*. In this instance, it is possible to show that the number of independent parameters necessary to describe the elastic behaviour of a rock body reduces to three scalar quantities, and the three-dimensional time-dependent constitutive equation describing the relation between stress and strain is a tensor form of Hooke's law (7.53):

$$\tau_{ij}(\mathbf{r}, t) = \lambda(\mathbf{r}) \delta_{ij} \varepsilon_{kk}(\mathbf{r}, t) + 2\mu(\mathbf{r}) \varepsilon_{ij}(\mathbf{r}, t) \quad (8.1)$$

The scalar fields  $\lambda = \lambda(\mathbf{r})$  and  $\mu = \mu(\mathbf{r})$  are called *Lamé parameters* and represent two of the three quantities determining the local elastic properties of a rock, the third one being the *density*. In the case of pure shear deformation, the dilatation  $\Delta = \varepsilon_{kk}$  is zero, so that (8.1) reduces, component-by-component, to (7.53) with  $Y = 2\mu$ . The quantity  $\mu$  is termed *shear modulus* or *rigidity modulus* and is a measure of the resistance of rocks to shear deformation. Values of  $\mu$  for the main crustal and mantle minerals are listed in Table 1.1. This parameter is always zero in the case of liquids, gases, and plasma, because matter in the fluid states does not exhibit resistance to finite strain. Conversely, the parameter  $\lambda$  does not have a simple physical interpretation, thereby it is often convenient to introduce some additional, more descriptive, elastic parameters. For example, we can separate in (8.1) isotropic and deviatoric components. If we define the mean normal stress  $\tau_0 = -p \equiv \tau_{kk}/3$ , then taking the trace in (8.1) gives:

$$\tau_0 = \left( \lambda + \frac{2}{3}\mu \right) \varepsilon_{kk} \equiv \kappa \Delta \quad (8.2)$$

where  $\Delta$  is the dilatation (Eq. 7.38) and  $\kappa$ , which is called *bulk modulus*, is a measure of the incompressibility of the material. In fact, by (8.2) we see that this quantity represents the ratio of hydrostatic pressure to the resulting volume change. Values of the *adiabatic* bulk modulus for the main crustal and mantle minerals are listed in Table 1.1. Two other common parameters, especially in engineering and applied geology, are the *Young modulus*,  $Y$ , and *Poisson's ratio*  $\nu$ . The Young modulus is defined as the ratio between extensional stress and resulting extensional strain for a cylinder that is pulled by both ends. It can be shown (e.g., Ranalli 1995) that is given by:

$$Y = \frac{(3\lambda + 2\mu)\mu}{\lambda + \mu} \quad (8.3)$$

Finally, the non-dimensional Poisson's ratio is the ratio between the lateral contraction of a cylinder that is pulled by both ends and its longitudinal extension. It is given by:

$$\nu = \frac{\lambda}{2(\lambda + \mu)} \quad (8.4)$$

This parameter varies between  $-1$  and a maximum value  $\nu = 0.5$  in the case of a liquid ( $\mu = 0$ ). A *Poisson solid* is a material such that  $\lambda = \mu$ , so that  $\nu = 0.25$ . In seismology, crustal rocks are often approximated as Poisson solids in the estimation of seismic velocities. In general,  $0.25 \leq \nu \leq 0.30$  for most crustal rocks. An interesting property of isotropic media is that in this instance the principal axes of stress coincide with the principal axes of strain. To prove this assertion, let us assume that:

$$\begin{cases} \tau_{ij} n_j = \lambda n_i \\ \varepsilon_{ij} n'_j = \lambda' n'_i \end{cases} \quad (8.5)$$

where  $\lambda$  and  $\lambda'$  are eigenvalues.

Substituting Hooke's law (8.1) in the first of these equations gives:

$$\begin{aligned} \tau_{ij} n_j &= (\lambda \delta_{ij} \Delta + 2\mu \varepsilon_{ij}) n_j \\ &= \lambda n_i \Delta + 2\mu \varepsilon_{ij} n_j = \lambda n_i \end{aligned} \quad (8.6)$$

Therefore,

$$\varepsilon_{ij} n_j = \frac{\lambda(1 - \Delta)}{2\mu} n_i \quad (8.7)$$

As a consequence, a principal axis of stress,  $\mathbf{n}$ , with eigenvalue  $\lambda$  is also a principal axis of strain with eigenvalue:

$$\lambda' = \frac{\lambda(1 - \Delta)}{2\mu} \quad (8.8)$$

## 8.2 Equations of Motion for Elastic Media

Now we are going to search a solution to the homogeneous version (7.51) of Cauchy's momentum equation, which links the second time derivatives of the displacement field (inertial term) to the spatial variations of the stress tensor (surface forces field). To this end, we use Hooke's law (8.1) to write the components of the stress tensor in terms of displacement field. Substituting the expression (7.32) for the strain tensor into Hooke's law gives:

$$\tau_{ij} = \lambda \delta_{ij} \nabla \cdot \mathbf{u} + \mu \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \quad (8.9)$$

This equation is a version of Hooke's law that makes explicit the dependence of the stress field from the displacements. It is important to note that in this relation the variables are evaluated at a given time  $t$  and position  $\mathbf{r}$ , and the small displacement  $u_i(\mathbf{r}, t)$  associated with the elastic response *instantaneously* accompanies the variations of stress in time. However, the momentum equation tells us that it is the spatial variability of the stress tensor to drive changes in the displacement field, which occur with velocity and acceleration that are given by the first and second time derivatives of  $u_i$ . Therefore, while (8.9) links the local components of stress at  $\mathbf{r}$  to the *spatial* variations of displacement in a neighbor of  $\mathbf{r}$ , the momentum Eq. (7.51) relates the *spatial* variability of the stress field to the *temporal* changes in the local displacement.

Substituting (8.9) into the momentum equation gives a differential equation in terms of displacements only:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \lambda}{\partial x_i} \Delta + (\lambda + \mu) \frac{\partial \Delta}{\partial x_i} + \frac{\partial \mu}{\partial x_j} \times \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \mu \frac{\partial^2 u_i}{\partial x_j^2} \quad (8.10)$$

This second-order partial derivatives equation is the *seismic wave equation*. It can be solved numerically when the three scalar fields  $\rho = \rho(\mathbf{r})$ ,  $\lambda = \lambda(\mathbf{r})$ , and  $\mu = \mu(\mathbf{r})$  are known. Sometimes this technique is used to create computer simulations of seismic wave propagation, following a hypothetical earthquake with assigned source parameters. The corresponding theoretical ground motion that would be observed at a station is called a *synthetic seismogram*, and it is also possible to predict the associated damage to human structures. The terms that include gradients of the Lamé parameters at the right-hand side of (8.10) are zero in the case of a homogeneous material. This is often a useful approximation in the study of seismic wave propagation within small quasi-homogeneous regions. In this instance, Eq. (8.10) reduces to:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = (\lambda + \mu) \frac{\partial \Delta}{\partial x_i} + \mu \frac{\partial^2 u_i}{\partial x_j^2} \quad (8.11)$$

There are two ways for using this equation in appropriate manner. First, it is possible to assume that the Earth is composed by a sequence of quasi-homogeneous layers with variable thickness, so that (8.11) is applied independently within each layer and the local solutions are linked together a posteriori. For example, the propagation of seismic waves through the oceanic crust is often modelled this way. In general, this approach assumes that the lateral variations of the mechanical parameters are negligible. Second, when both  $\mu$  and  $\lambda$  vary smoothly, it is possible to find an approximate solution to (8.10) considering the material as formed by small homogeneous regions where (8.11) holds, and assuming smooth variations of the parameters  $\mu$  and  $\lambda$  from each

region to its neighboring areas. Clearly, this approach is effective when one takes into account of the presence of discontinuities in the mechanical properties at some boundaries, for example at the Moho, the CMB, etc. We shall explore this technique, which is referred to as the *seismic ray method*, in the next chapter. We will prove that the method can be used to describe the propagation of high-frequency seismic waves, thereby the Earth's free oscillations (with frequencies of the order of one mHz or less) are excluded from a description in terms of seismic rays.

### 8.3 Seismic Waves

Although the seismic wave Eq. (8.10) can be solved numerically or assuming specific distributions of the elastic parameters, it is possible to gain some insight about the physics of seismic waves propagation from the more simple homogeneous version (8.11). Taking the divergence of this equation gives:

$$\rho \frac{\partial^2}{\partial t^2} \frac{\partial u_i}{\partial x_i} = (\lambda + \mu) \frac{\partial^2 \Delta}{\partial x_i^2} + \mu \frac{\partial}{\partial x_i} \frac{\partial^2 u_i}{\partial x_j^2}$$

Therefore,

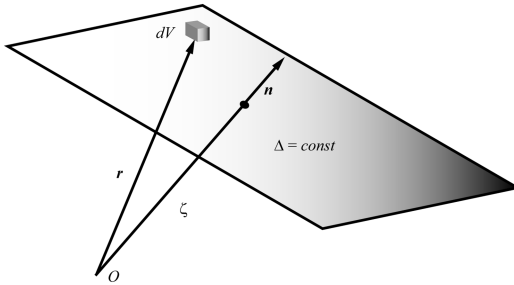
$$\nabla^2 \Delta - \frac{\rho}{\lambda + 2\mu} \frac{\partial^2 \Delta}{\partial t^2} = 0 \quad (8.12)$$

This is a standard *wave equation* (or *D'Alembert's equation*). It implies that in isotropic homogeneous media a volume perturbation propagates with velocity:

$$\alpha \equiv \sqrt{\frac{\lambda + 2\mu}{\rho}} \quad (8.13)$$

To prove this, let us assume that at any time  $t$  the dilatation  $\Delta = \Delta(\mathbf{r}, t)$  is constant along a plane having distance  $\zeta$  from the origin, as shown in Fig. 8.1.

Although this assumption is a good approximation of reality only at great distance from the source, it is useful to understand the physics of



**Fig. 8.1** Plane waves propagation. At any time  $t$ , the dilatation is a function of the distance  $\zeta = \zeta(t)$  of a plane from the origin  $O$

waves. For example, if we consider the propagation of a circular wave along the surface of a lake after having thrown a stone, we note that at sufficient distance from the source the wavefront is approximately flat, so that the geometry of the wavefront can be approximated by a plane rather than a circle. If  $\mathbf{n}$  is a versor normal to the plane, then at any point  $\mathbf{r}$  we have that:

$$\Delta(\mathbf{r}, t) = \varphi(\mathbf{r} \cdot \mathbf{n}, t) = \varphi(\zeta, t) \quad (8.14)$$

In this instance, the gradient of the scalar field  $\Delta = \Delta(\mathbf{r}, t)$  can be written as follows:

$$\nabla \Delta(\mathbf{r}, t) = \mathbf{n} \frac{\partial \varphi}{\partial \zeta} \quad (8.15)$$

Using (8.15), we see that the wave Eq. (8.12) reduces to:

$$\frac{\partial^2 \varphi}{\partial \zeta^2} - \frac{1}{\alpha^2} \frac{\partial^2 \varphi}{\partial t^2} = 0 \quad (8.16)$$

This is the *plane waves equation*, which describes the propagation of unidimensional waves (or *plane waves*). This equation can be solved analytically. To this purpose, let us rewrite (8.16) as follows:

$$\left( \frac{\partial}{\partial t} - \alpha \frac{\partial}{\partial \zeta} \right) \left( \frac{\partial}{\partial t} + \alpha \frac{\partial}{\partial \zeta} \right) \varphi = 0 \quad (8.17)$$

Now we perform a change of variables introducing two new variables  $\eta$  and  $\xi$ :

$$\eta = t - \frac{\zeta}{\alpha}; \quad \xi = t + \frac{\zeta}{\alpha} \quad (8.18)$$

The inverse transformation from  $(\eta, \xi)$  to  $(t, \zeta)$  is the following one:

$$t = \frac{1}{2} (\eta + \xi); \quad \zeta = \frac{\alpha}{2} (\xi - \eta) \quad (8.19)$$

Therefore, the derivatives are changed as follows:

$$\frac{\partial}{\partial \eta} = \frac{1}{2} \left( \frac{\partial}{\partial t} - \alpha \frac{\partial}{\partial \zeta} \right); \quad \frac{\partial}{\partial \xi} = \frac{1}{2} \left( \frac{\partial}{\partial t} + \alpha \frac{\partial}{\partial \zeta} \right) \quad (8.20)$$

If we substitute these expressions into (8.17), we see that the plane wave equation assumes the simple form:

$$\frac{\partial^2 \varphi}{\partial \eta \partial \xi} = 0 \quad (8.21)$$

The solutions of this equation have the form:

$$\varphi = \varphi_1(\eta) + \varphi_2(\xi) \quad (8.22)$$

where  $\varphi_1$  and  $\varphi_2$  are arbitrary functions. Therefore, the general solution to (8.16) has the form:

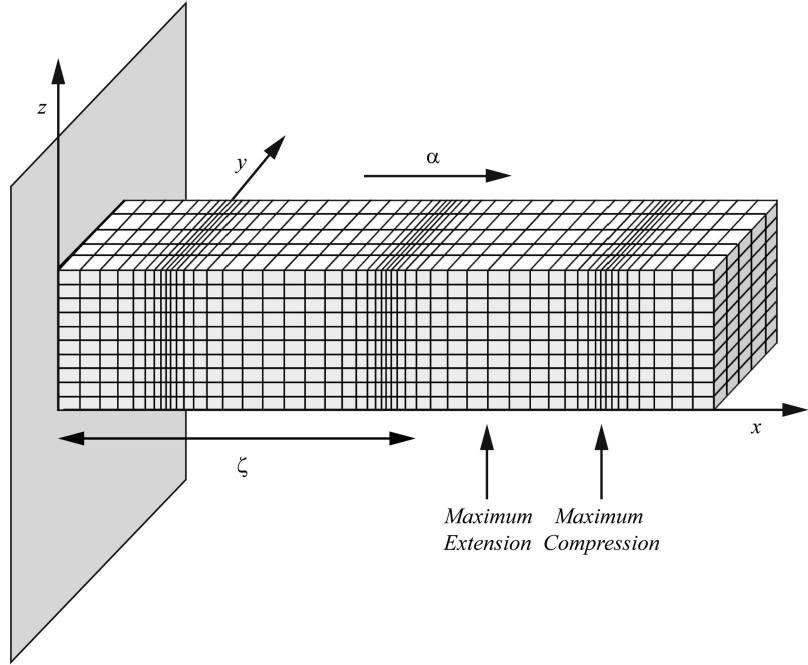
$$\varphi = \varphi_1 \left( t - \frac{\zeta}{\alpha} \right) + \varphi_2 \left( t + \frac{\zeta}{\alpha} \right) \quad (8.23)$$

To understand the meaning of this solution, let us assume, for example, that  $\varphi_2 = 0$ , so that  $\varphi = \varphi_1(t - \zeta/\alpha)$ . Along any plane defined by some value  $\zeta = \text{const}$ , the field  $\varphi$  clearly varies with time. Similarly, at any given time  $t$  the field assumes different values along distinct planes  $\zeta$ . However, from (8.23) we see that  $\varphi$  assumes the same value for the set of all pairs  $(\zeta, t)$  such that  $t - \zeta/\alpha = \text{const}$ , hence for:

$$\zeta = \text{const} + \alpha t \quad (8.24)$$

Therefore, if for  $t = t_0$  the field  $\varphi$  had some value  $\varphi = \varphi_0$  along the plane  $\zeta = \zeta_0$ , it will acquire the same value after a time interval  $\Delta t$  along a plane having distance  $\alpha \Delta t$  from the plane  $\zeta = \zeta_0$ . Consequently, we can say that the field *propagates* through the material along the direction  $\mathbf{n}$  with velocity  $\alpha$ , so that the function  $\varphi_1(t - \zeta/\alpha)$  represents a longitudinal plane wave

**Fig. 8.2** Pattern of deformation associated with the propagation of  $P$  waves



propagating in the direction  $\mathbf{n}$ . Similarly, it can be shown that the function  $\varphi_2(t + \zeta/\alpha)$  represents a plane wave propagating in the direction  $-\mathbf{n}$ . The waves associated with the propagation of volume variations, with velocity  $\alpha$  given by (8.13), are called  $P$  waves and represent the first arrivals to seismic stations after an earthquake. In the case of plane waves, the volume variations at any given distance  $\zeta$  from the source occur in the direction of wave propagation and are alternatively longitudinal shortening or dilatation, as illustrated in Fig. 8.2.

We say that the displacement field of  $P$ -waves is *longitudinal*. Now we will take the curl of the homogeneous wave Eq. (8.11). This gives:

$$\begin{aligned} \rho \frac{\partial^2}{\partial t^2} \left( \varepsilon_{ijk} \frac{\partial u_k}{\partial x_j} \right) &= (\lambda + \mu) \varepsilon_{ijk} \frac{\partial^2 \Delta}{\partial x_j \partial x_k} \\ + \mu \nabla^2 \left( \varepsilon_{ijk} \frac{\partial u_k}{\partial x_j} \right) &= \mu \nabla^2 \left( \varepsilon_{ijk} \frac{\partial u_k}{\partial x_j} \right) \end{aligned} \quad (8.25)$$

where  $\varepsilon_{ijk}$  is the Levi-Civita tensor (see Appendix 1). It is useful at this point to introduce a new vector field, the curl of the displacement field:

$$\mathbf{\Gamma} = \nabla \times \mathbf{u} \quad (8.26)$$

With this definition, the wave Eq. (8.25) can be rewritten as follows:

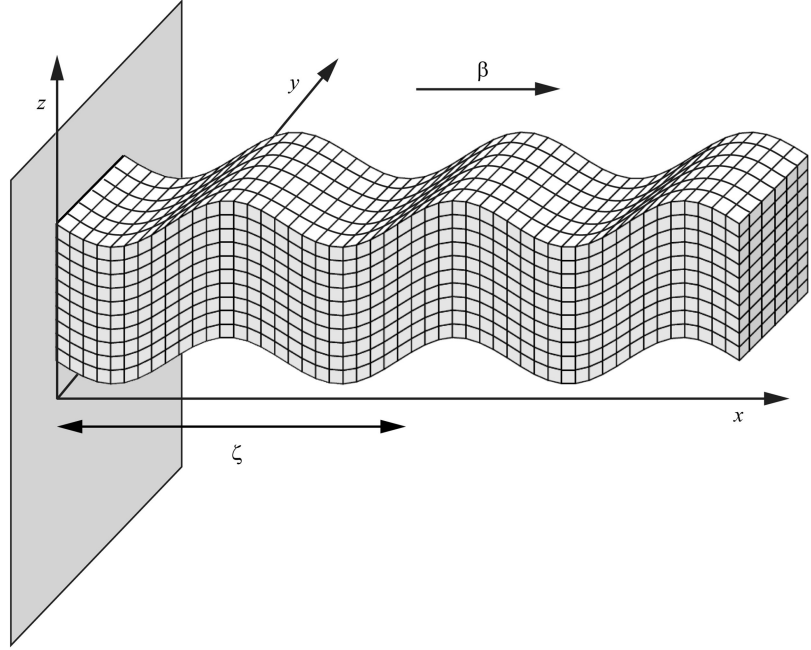
$$\nabla^2 \mathbf{\Gamma} - \frac{1}{\beta^2} \frac{\partial^2 \mathbf{\Gamma}}{\partial t^2} = 0 \quad (8.27)$$

Again, this is a standard wave equation, with velocity of propagation given by:

$$\beta = \sqrt{\frac{\mu}{\rho}} \quad (8.28)$$

A comparison of (8.28) with (8.13) shows that we always have  $\alpha > \beta$ , because the Lamé parameters are both positive. The waves associated with (8.27) are called  $S$  waves (Fig. 8.3) and represent a later arrival to seismic stations after an earthquake (in fact, “ $S$ ” stands for «*Secundae*», while “ $P$ ” stands for «*Primae*», from Latin). We also note the  $S$  waves cannot propagate in fluid state matter, because in this case we would have  $\mu = 0$ , so that (8.28) implies  $\beta = 0$ . Therefore, these waves do not cross the external Earth’s core as well as the oceans.

**Fig. 8.3** Pattern of deformation associated with the propagation of  $S$  waves



In the case of a plane wave that is propagating in the  $x$  direction, we have that:

$$\Gamma_i = \Gamma_i \left( t - \frac{x}{\beta} \right) = \varepsilon_{ijk} \frac{\partial u_k}{\partial x_j} = f_i(x) \quad (8.29)$$

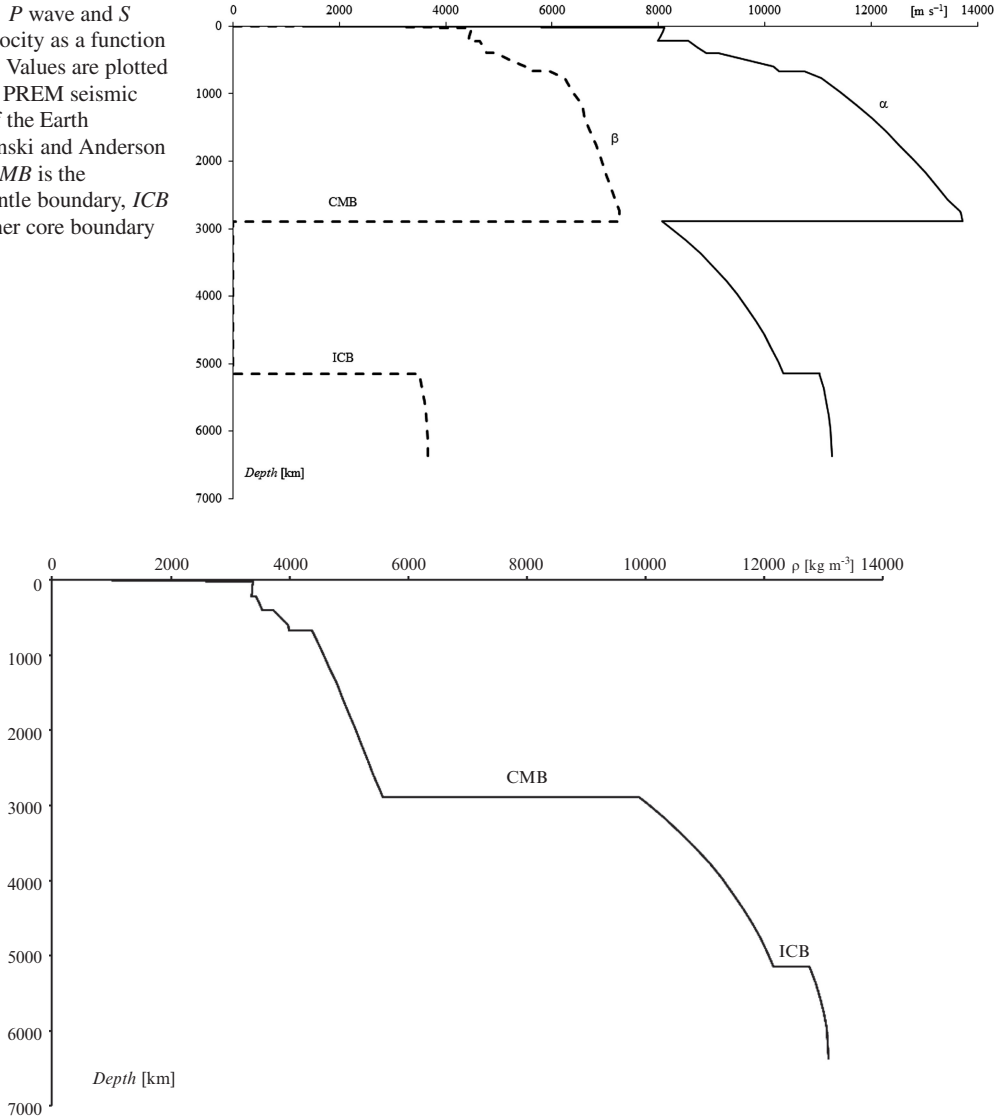
Therefore, as for  $P$  waves, the derivatives of displacement components in the  $y$  and  $z$  directions must be zero, because otherwise the displacement would increase or decrease indefinitely:

$$\frac{\partial u_z}{\partial y} = \frac{\partial u_y}{\partial z} = 0$$

Consequently,  $\Gamma_x = 0$  and  $\mathbf{u}$  lies in the  $yz$  plane ( $u_x = 0$ ). These waves are said *transverse waves*, to express the fact that displacements always occur in a direction normal to the propagation direction, as illustrated in the example of Fig. 8.3. When the displacement occurs along the  $z$  axis, we say that the wave is polarized vertically (*SV waves*, Fig. 8.3), while for a motion in the  $y$  direction, parallel to the Earth's surface, we say that the wave is polarized horizontally (*SH waves*). The typical velocities of  $P$  and  $S$  waves in the crust are easily calculated assuming that the crust is a Poisson solid with  $\mu = \lambda$

$\approx 3 \times 10^{10}$  Pa and density  $\rho = 3 \times 10^3$  Kg/m<sup>3</sup>. In this instance, we have average velocities  $\alpha \approx 5.5$  km s<sup>-1</sup> and  $\beta \approx 3.2$  km s<sup>-1</sup>. The trends of the seismic velocities and the density in the *PREM* (Preliminary Reference Earth Model, Dziewonski and Anderson 1981) global seismic model of the Earth are illustrated in Figs. 8.4 and 8.5. We note an approximately linear increase of  $\alpha$ ,  $\beta$ , and  $\rho$  in the lower mantle, with gradients:  $\partial\alpha/\partial z = 1.35 \pm 0.04$  m s<sup>-1</sup> km<sup>-1</sup>,  $\partial\beta/\partial z = 0.56 \pm 0.03$  m s<sup>-1</sup> km<sup>-1</sup>, and  $\partial\rho/\partial z = 0.54 \pm 0.01$  kg m<sup>-3</sup> km<sup>-1</sup>, and a roughly linear increase in the transition zone, with higher gradients:  $\partial\alpha/\partial z = 4.47 \pm 0.48$  m s<sup>-1</sup> km<sup>-1</sup>,  $\partial\beta/\partial z = 2.52 \pm 0.29$  m s<sup>-1</sup> km<sup>-1</sup>, and  $\partial\rho/\partial z = 1.07 \pm 0.14$  kg m<sup>-3</sup> km<sup>-1</sup>. Conversely, the trend is discontinuous in the asthenosphere. The linear increase in the lower mantle and in the transition zone is in agreement with what we expect from the changes in pressure and temperature on rocks of uniform composition and crystal structure. We also note the dramatic  $\sim 5.7$  km s<sup>-1</sup> drop of  $\alpha$  at the CMB, which accompanies the disappearance of  $S$  waves in the liquid external core. Another important discontinuity for  $\alpha$ ,  $\beta$ , and  $\rho$  is the inner-core boundary (ICB) at  $\sim 5,150$  km, which separates the liquid external core from the solid iron internal core (Figs. 8.4 and 8.5).

**Fig. 8.4**  $P$  wave and  $S$  wave velocity as a function of depth. Values are plotted from the PREM seismic model of the Earth (Dziewonski and Anderson 1981).  $CMB$  is the core–mantle boundary,  $ICB$  is the inner core boundary



**Fig. 8.5** Rock density as a function of depth. Values are plotted from the PREM seismic model of the Earth (Dziewonski and Anderson 1981).  $CMB$  is the core–mantle boundary,  $ICB$  is the inner core boundary

An important class of solutions to the plane waves Eq. (8.12) is represented by periodic functions having the form:

$$\Delta(\mathbf{r}, t) = \Delta_0(\mathbf{r}) \cos(\omega t + \vartheta) \quad (8.30)$$

where  $\vartheta$  is the *phase*,  $\omega$  is the angular frequency, and  $\nu = \omega/2\pi$  is the wave frequency. Solutions of the form (8.30) are called *monochromatic waves*. Substituting (8.30) into (8.12) gives a time-independent equation that determines

the spatial distribution of the monochromatic wave:

$$\nabla^2 \Delta_0 + \frac{\omega^2}{\alpha^2} \Delta_0 = 0 \quad (8.31)$$

This is a scalar Helmholtz's equation, similar to that encountered in Sect. 4.8. We know that the solutions to this equation are combinations of sines and cosines:

$$\Delta_0(\mathbf{r}) = e^{i \mathbf{k} \cdot \mathbf{r}} \quad (8.32)$$

where  $\mathbf{k}$  is the wave vector. In the case of a *monochromatic plane wave* that is propagating in a unique direction, the dilatation must also be a function of  $(t - \zeta/\alpha)$ . Therefore we have that,

$$\omega t + \vartheta = \omega \left( t - \frac{\zeta}{\alpha} \right) \equiv \omega t - k\zeta \quad (8.33)$$

If  $\mathbf{n}$  is the versor in the direction of propagation and  $\lambda = \alpha/v$  is the wavelength, then the wave vector is given by:

$$\mathbf{k} = \frac{\omega}{\alpha} \mathbf{n} = \frac{2\pi v}{\alpha} \mathbf{n} = \frac{2\pi}{\lambda} \mathbf{n} \quad (8.34)$$

The magnitude  $k$  of this vector represents the number of oscillations in a segment of length  $2\pi$  (wavenumber). Using a complex notation, the general plane wave monochromatic solution to (8.12) will have the following form:

$$\Delta(\mathbf{r}, t) = A e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \quad (8.35)$$

where  $A$  is the amplitude. A similar formula can be written for the transverse waves. Of course, only the real part of (8.35) has physical significance. In the crust, where  $\alpha \approx 5.5 \text{ km s}^{-1}$ , a  $P$  wave that is propagating with period  $T = 1/v = 5 \text{ s}$  will have wavelength  $\lambda \approx 27.5 \text{ km}$ . In general, seismic waves associated with earthquakes have periods between 1 and 10 s, that is, frequency  $0.1 \text{ Hz} \leq v \leq 1 \text{ Hz}$ . Therefore, typical crustal wavelengths of  $P$  waves range between 5.5 and 55 km. Other kinds of seismic waves have different periods. For example, in the case surface waves  $10 \text{ s} \leq T \leq 100 \text{ s}$ , while free oscillations of the Earth occur with periods  $100 \text{ s} \leq T \leq 1,000 \text{ s}$ . Conversely, artificial waves generated in exploration geophysics have very short periods between  $10^{-4} \text{ s}$  and  $10^{-3} \text{ s}$ .

More realistic solutions to the wave Eqs. (8.12) and (8.27) can be found assuming a spherical rather than planar symmetry. In this instance, it is necessary to represent the equations in spherical coordinates (Eq. 2.28) and use the spherical version (4.73) of the gradient in a local reference frame having the origin at the seismic source. In this frame, (8.12) assumes the form:

$$\begin{aligned} \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \Delta}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \Delta}{\partial \theta} \right) \\ + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \Delta}{\partial \phi^2} - \frac{1}{\alpha^2} \frac{\partial^2 \Delta}{\partial t^2} = 0 \end{aligned} \quad (8.36)$$

In this equation, the spherical coordinates  $(\theta, \phi)$  do not represent global coordinates of colatitude and longitude; they are simply *local* angular coordinates. Assuming spherical symmetry implies that  $\Delta$  does not depend from these variables. Therefore, the wave equation for the dilatation can be rewritten as follows:

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \Delta}{\partial r} \right) - \frac{1}{\alpha^2} \frac{\partial^2 \Delta}{\partial t^2} = 0 \quad (8.37)$$

To solve this equation, we set:

$$\Delta(r, t) = \xi(r, t) / r \quad (8.38)$$

Then, substituting into (8.37):

$$\frac{1}{r} \left( \frac{\partial^2 \xi}{\partial r^2} - \frac{1}{\alpha^2} \frac{\partial^2 \xi}{\partial t^2} \right) = 0 \quad (8.39)$$

For  $r \neq 0$  this reduces to a classic plane waves Eq. (8.16). Therefore, the solution for  $\Delta$  has the form:

$$\Delta(r, t) = \frac{f(t \pm r/\alpha)}{r} \quad (8.40)$$

This is the spherically symmetric solution to the wave equation for the dilatation. A similar solution can be written for the components of the vector field  $\mathbf{\Gamma}$ . The solution describes wave fronts that are spherical surfaces centered about the origin  $\mathbf{r} = 0$  and having amplitude that is inversely proportional to the distance from the origin. For  $r = 0$ , (8.40) is *not* a solution to the wave equation. However, it is possible to show that it is a solution of the following non-homogeneous wave equation (e.g., Aki and Richards 2002):

$$\nabla^2 \Delta - \frac{1}{\alpha^2} \frac{\partial^2 \Delta}{\partial t^2} = -4\pi \delta(\mathbf{r}) f(t) \quad (8.41)$$

where  $\delta(\mathbf{r})$  is the Dirac delta function centered at the origin. The term at the right-hand side of

(8.41) represents a point source located at the origin and having time-varying magnitude  $= f(t)$ .

## 8.4 Seismic Energy

Seismic waves carry energy both in the form of kinetic energy, associated with the motion of volume elements, and potential energy related to deformation. If  $K$  and  $U$  are respectively the kinetic and the potential energy per unit volume, then the total energy density  $E$  in a material during the travel of a seismic wave is given by:

$$E = K + U \quad (8.42)$$

The kinetic energy density is clearly determined by the local velocity of displacement, so that:

$$K = \frac{1}{2} \rho \dot{u}^2 = \frac{1}{2} \rho \left[ \left( \frac{\partial u_1}{\partial t} \right)^2 + \left( \frac{\partial u_2}{\partial t} \right)^2 + \left( \frac{\partial u_3}{\partial t} \right)^2 \right] \quad (8.43)$$

Regarding the potential energy, let us assume that an elastic body starts deforming under the action of external forces. The rate of mechanical work  $P(t)$  depends from both internal body forces operating in the region  $\mathbf{R}$  and surface forces exerted along the boundary,  $\Gamma(\mathbf{R})$ , of  $\mathbf{R}$ :

$$P(t) = \oint_{\Gamma(\mathbf{R})} T_i \dot{u}_i dS + \int_{\mathbf{R}} f_i \dot{u}_i dV \quad (8.44)$$

where  $\mathbf{T} = \mathbf{T}(\mathbf{n})$  is the traction exerted along the surface element  $d\mathbf{S} = \mathbf{n}dS$  and  $\mathbf{f} = \mathbf{f}(\mathbf{r})$  is the body forces field. Using Cauchy's theorem (7.3) and Gauss' divergence theorem (see Appendix 1) in (8.44) gives:

$$\begin{aligned} P(t) &= \oint_{\Gamma(\mathbf{R})} \tau_{ij} n_j \dot{u}_i dS + \int_{\mathbf{R}} f_i \dot{u}_i dV \\ &= \oint_{\Gamma(\mathbf{R})} (\dot{u}_i \tau_{ij}) n_j dS + \int_{\mathbf{R}} f_i \dot{u}_i dV \end{aligned}$$

$$\begin{aligned} &= \int_{\mathbf{R}} \left[ \frac{\partial}{\partial x_j} (\dot{u}_i \tau_{ij}) + f_i \dot{u}_i \right] dV \\ &= \int_{\mathbf{R}} \left[ \frac{\partial \tau_{ij}}{\partial x_j} + f_i \right] \dot{u}_i dV + \int_{\mathbf{R}} \tau_{ij} \frac{\partial \dot{u}_i}{\partial x_j} dV \\ &= \int_{\mathbf{R}} \left[ \rho \frac{\partial \dot{u}_i}{\partial t} \dot{u}_i + \tau_{ij} \frac{\partial \dot{u}_i}{\partial x_j} \right] dV \\ &= \int_{\mathbf{R}} \left[ \rho \frac{\partial \dot{u}_i}{\partial t} \dot{u}_i + \frac{1}{2} \tau_{ij} \frac{\partial \dot{u}_i}{\partial x_j} + \frac{1}{2} \tau_{ji} \frac{\partial \dot{u}_j}{\partial x_i} \right] dV \\ &= \int_{\mathbf{R}} \left[ \rho \frac{\partial \dot{u}_i}{\partial t} \dot{u}_i + \tau_{ij} \dot{\epsilon}_{ij} \right] dV \\ &= \frac{1}{2} \frac{\partial}{\partial t} \int_{\mathbf{R}} \rho \dot{u}_i \dot{u}_i dV + \int_{\mathbf{R}} \tau_{ij} \dot{\epsilon}_{ij} dV \\ &= \frac{\partial K}{\partial t} + \int_{\mathbf{R}} \tau_{ij} \dot{\epsilon}_{ij} dV \quad (8.45) \end{aligned}$$

where we have used the momentum Eq. (7.50), the symmetry of the stress tensor, and the definition (8.44) of kinetic energy density. It is useful at this point to introduce a function  $W$  of the strain components, which allows to generate the corresponding stress tensor by its derivatives:

$$\tau_{ij} = \frac{\partial W}{\partial \epsilon_{ij}} \quad (8.46)$$

The function  $W$  is called the *strain-energy function* and has the dimensions of an energy density [ $\text{J m}^{-3}$ ]. The most general form of Hooke's law, which expresses the linear dependence of stress from strain, can be written as follows:

$$\tau_{ij} = C_{ijhk} \epsilon_{hk} \quad (8.47)$$

This constitutive equation is more general than (8.1), because it also holds in the case of non-isotropic materials. The tensor  $C_{ijhk}$  at the right-hand side is referred to as the *elastic tensor*. Substituting (8.47) into (8.46) gives:

$$\frac{\partial W}{\partial \epsilon_{ij}} = C_{ijhk} \epsilon_{hk} \quad (8.48)$$

This equation says that all the derivatives of  $W$  are linear functions of the strain tensor components. Therefore, considering that this function must be zero in the undeformed state of equilibrium, we have that  $W$  must be a homogeneous bilinear function of the strain components:

$$W = D_{ijhk} \varepsilon_{ij} \varepsilon_{hk} \quad (8.49)$$

Taking the first derivative of this expression gives:

$$\begin{aligned} \frac{\partial W}{\partial \varepsilon_{rs}} &= D_{rshk} \varepsilon_{hk} + D_{ijrs} \varepsilon_{ij} = D_{rshk} \varepsilon_{hk} + D_{hkrs} \varepsilon_{hk} \\ &= (D_{rshk} + D_{hkrs}) \varepsilon_{hk} \end{aligned}$$

A comparison with (8.48) allows to express the elastic tensor in terms of components of  $D$ :

$$C_{hkrs} = D_{rshk} + D_{hkrs}$$

Therefore, using (8.47) we see that (8.49) can be rewritten as follows:

$$\begin{aligned} W &= D_{ijhk} \varepsilon_{ij} \varepsilon_{hk} = \frac{1}{2} (D_{ijhk} + D_{hkij}) \varepsilon_{ij} \varepsilon_{hk} \\ &= \frac{1}{2} C_{ijhk} \varepsilon_{ij} \varepsilon_{hk} = \frac{1}{2} \tau_{ij} \varepsilon_{ij} \end{aligned} \quad (8.50)$$

In order to give a physical significance to the strain-energy function  $W$ , we first calculate its time derivative:

$$\begin{aligned} \dot{W} &= \frac{1}{2} C_{ijhk} (\dot{\varepsilon}_{ij} \varepsilon_{hk} + \varepsilon_{ij} \dot{\varepsilon}_{hk}) \\ &= \frac{1}{2} \tau_{ij} \dot{\varepsilon}_{ij} + \frac{1}{2} \tau_{hk} \dot{\varepsilon}_{hk} = \tau_{ij} \dot{\varepsilon}_{ij} \end{aligned} \quad (8.51)$$

where we have used an obvious symmetry property of the elastic tensor:  $C_{ijhk} = C_{hkij}$ . A comparison of this expression with (8.45) shows that  $W$  represents the potential energy density of the material in the deformed state. Let us consider now a monochromatic  $SH$  plane wave that is propagating in the  $x$  direction. In this instance, the displacements occur in the  $y$  direction at any time, so that:

$$\begin{cases} u_y = A \sin(\omega t - kx) \\ \frac{\partial u_y}{\partial t} = A \omega \cos(\omega t - kx) \end{cases} \quad (8.52)$$

where  $A$  is the wave amplitude,  $\omega$  is the angular frequency, and  $k = \omega/\beta$ . The density of kinetic energy can be calculated using (8.43):

$$K = \frac{1}{2} \rho A^2 \omega^2 \cos^2(\omega t - kx) \quad (8.53)$$

Therefore, the average kinetic energy density over a wavelength is:

$$\langle K \rangle = \frac{1}{4} \rho A^2 \omega^2 \quad (8.54)$$

Let us consider now the potential energy density. In this example, the only non-zero components of the strain tensor are:

$$\varepsilon_{xy} = \varepsilon_{yx} = \frac{1}{2} \frac{\partial u_y}{\partial x} = -\frac{1}{2} A k \cos(\omega t - kx) \quad (8.55)$$

Therefore, using Hooke's law (8.1) we have that the non-zero components of the stress tensor are:

$$\tau_{xy} = \tau_{yx} = 2\mu \varepsilon_{xy} = -Ak\mu \cos(\omega t - kx) \quad (8.56)$$

Substituting these expressions into (8.50) gives the strain-energy function associated with this wave:

$$W = \frac{1}{2} A^2 k^2 \mu \cos^2(\omega t - kx) \quad (8.57)$$

Also in this case it is useful to consider the spatial average over a wavelength. It will be given by:

$$\langle W \rangle = \frac{1}{4} A^2 k^2 \mu = \frac{1}{4} \rho A^2 \omega^2 = \langle K \rangle \quad (8.58)$$

Therefore, the average kinetic energy coincides with the average potential energy. The same result can be easily obtained in the case of monochromatic  $P$  plane wave.

As a result, the average total energy density is proportional to the squared wave amplitude and to the square of angular frequency:

$$\langle E \rangle = \langle K \rangle + \langle W \rangle = \frac{1}{2} \rho A^2 \omega^2 \quad (8.59)$$

This result implies that for any given amplitude high-frequency waves carry a greater quantity of energy per wavelength. More information about this topic can be found in the book of Shearer (2009).

---

## Problems

1. Download the MS Excel file PREM.xlsx, containing the PREM model, from the supplemental material web site. Use this file to determine how long time takes a downward directed  $P$  wave generated at the Earth's surface to reach the antipodal point;
2. Determine an expression for the strain tensor of a monochromatic wave travelling in the  $x$  direction;
3. Find a test to determine if a material is a Poisson solid from measurements of seismic velocity;
4. Find an expression for determining the Poisson ratio from measurements of seismic velocity;
5. Determine the relation between elastic moduli, Lamé parameters, and the velocities  $\alpha$  and  $\beta$  when  $\nu = -1$  and  $\nu = 0$ ;
6. Estimate the  $\alpha/\beta$  ratio from a set of arrival times of  $P$  and  $S$  waves at seismic stations.

---

## References

- Aki K, Richards PG (2002) Quantitative seismology, 2nd edn. University Science Books, Sausalito, 700 pp
- Dziewonski AM, Anderson DL (1981) Preliminary reference Earth model. *Phys Earth Planet Inter* 25: 297–356
- Ranalli G (1995) Rheology of the Earth, 2nd edn. Chapman & Hall, London, 413 pp
- Shearer PM (2009) Introduction to seismology, 2nd edn. Cambridge University Press, Cambridge, UK, 396 pp



## Abstract

In this second chapter on seismology, I introduce seismic ray theory starting from the Eikonal equation. The classic concepts about travel-time curves are discussed, as well as the seismic phase's nomenclature at regional and global scale.

## 9.1 The Eikonal Equation

Seismic rays are the continuum mechanics analogous of the usual light rays of geometrical optics. They have been used since the dawn of seismological science in the interpretation of earthquake data and still furnish the simplest computational approach to a wide class of problems, including the localization of earthquake foci, the determination of focal mechanisms, and seismic tomography. However, we shall see that the range of applicability of this approach is limited to the propagation of high-frequency waves. Furthermore, seismic rays do not adequately describe non-geometrical phenomena such as diffraction. The starting point of seismic ray theory is the so-called *eikonal equation*, which determines the relation between the geometry of a wavefront and the velocity fields  $\alpha = \alpha(\mathbf{r})$  and  $\beta = \beta(\mathbf{r})$ . We know that these fields depend in turn from the mechanical parameters of the rocks via Eqs. (8.13) and (8.28). Let us assume that a non-homogeneous but isotropic material can be approximated by a regular grid of small homogeneous regions and that the mechanical

parameters vary smoothly from a grid cell to its neighboring cells. In this instance, at each grid center we can write the wave equations in the form:

$$\nabla^2 \Delta - \frac{1}{\alpha^2(\mathbf{r})} \frac{\partial^2 \Delta}{\partial t^2} = 0 \quad (9.1)$$

$$\nabla^2 \Gamma - \frac{1}{\beta^2(\mathbf{r})} \frac{\partial^2 \Gamma}{\partial t^2} = 0 \quad (9.2)$$

These equations can be solved locally to determine the mode of propagation of seismic waves within specific grid cells. Let  $\varphi = \varphi(\mathbf{r}, t)$  any of the field components  $\Delta(\mathbf{r}, t)$  or  $\Gamma_i(\mathbf{r}, t)$ . The form of Eqs. (9.1) and (9.2) suggests a monochromatic solution with amplitude and phase that depend from the position.

Therefore, we make the ansatz:

$$\varphi(\mathbf{r}, t) = A(\mathbf{r}) e^{i\omega(t - T(\mathbf{r}))} \quad (9.3)$$

where  $T = T(\mathbf{r})$  is a phase factor and  $A = A(\mathbf{r})$  is the local wave amplitude. Taking the gradient and then the divergence of (9.3) gives:

$$\nabla \varphi = [\nabla A(\mathbf{r}) - i\omega A(\mathbf{r}) \nabla T(\mathbf{r})] e^{i\omega(t-T(\mathbf{r}))} \quad (9.4)$$

$$\begin{aligned} \nabla^2 \varphi = & [\nabla^2 A(\mathbf{r}) - 2i\omega \nabla A(\mathbf{r}) \cdot \nabla T(\mathbf{r}) \\ & - i\omega A(\mathbf{r}) \nabla^2 T(\mathbf{r}) \\ & - \omega^2 A(\mathbf{r}) \nabla T(\mathbf{r}) \cdot \nabla T(\mathbf{r})] e^{i\omega(t-T(\mathbf{r}))} \end{aligned} \quad (9.5)$$

Regarding the time derivative, we easily obtain:

$$\frac{\partial^2 \varphi}{\partial t^2} = -A(\mathbf{r}) \omega^2 e^{i\omega(t-T(\mathbf{r}))} \quad (9.6)$$

Substituting these expressions in one of the seismic wave Eqs. (9.1) or (9.2) gives:

$$\begin{aligned} \nabla^2 A(\mathbf{r}) - 2i\omega \nabla A(\mathbf{r}) \cdot \nabla T(\mathbf{r}) \\ - i\omega A(\mathbf{r}) \nabla^2 T(\mathbf{r}) - \omega^2 A(\mathbf{r}) |\nabla T(\mathbf{r})|^2 \\ = -\frac{A(\mathbf{r}) \omega^2}{c^2(\mathbf{r})} \end{aligned} \quad (9.7)$$

where  $c = c(\mathbf{r})$  is either the  $P$ -wave velocity  $\alpha$  or the  $S$ -wave velocity  $\beta$ . Now we decompose this equation into the real and imaginary parts. The resulting equations are:

$$\nabla^2 A(\mathbf{r}) - \omega^2 A(\mathbf{r}) |\nabla T(\mathbf{r})|^2 = -\frac{A(\mathbf{r}) \omega^2}{c^2(\mathbf{r})} \quad (9.8)$$

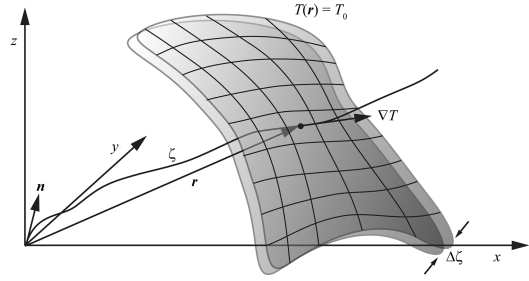
$$2\nabla A(\mathbf{r}) \cdot \nabla T(\mathbf{r}) + A(\mathbf{r}) \nabla^2 T(\mathbf{r}) = 0 \quad (9.9)$$

The second of these equations is called the *amplitude transport equation*. For the moment, we shall focus on the first equation only. Dividing both sides by  $A\omega^2$  gives:

$$|\nabla T(\mathbf{r})|^2 - \frac{1}{c^2(\mathbf{r})} = \frac{\nabla^2 A(\mathbf{r})}{A(\mathbf{r}) \omega^2} \quad (9.10)$$

When the angular frequency of the wave is sufficiently high, hence in the limit  $\omega \rightarrow \infty$ , the term at the right-hand side can be ignored. Therefore:

$$|\nabla T(\mathbf{r})|^2 = \frac{1}{c^2(\mathbf{r})} \quad (9.11)$$



**Fig. 9.1** Wavefront propagation. At any arrival time  $t = T_0$ , the set of points  $T(\mathbf{r}) = T_0$  is a 3-D surface and the gradient of  $T$  is locally normal to the surface. The distance  $\zeta = \zeta(t)$  from the source, placed at the origin, can be used to build a parametric seismic ray equation  $\mathbf{r} = \mathbf{r}(\zeta)$

This is known as the *eikonal equation*. Let us introduce now the *slowness*  $s$ , which is simply the reciprocal of a seismic velocity  $\alpha$  or  $\beta$ :

$$s(\mathbf{r}) \equiv \frac{1}{c(\mathbf{r})} \quad (9.12)$$

Substituting into (9.11) allows to express the eikonal equation in the following alternative form:

$$|\nabla T(\mathbf{r})|^2 = s^2(\mathbf{r}) \quad (9.13)$$

The phase factor  $T = T(\mathbf{r})$  is a scalar field having the dimensions of a time. According to the eikonal equation, the amplitude of its gradient locally coincides with the slowness  $s$ . The physical interpretation of an arbitrary surface  $T(\mathbf{r}) = T_0$  is simple. From (9.3), we see that for  $t = T_0$  the propagating anomaly (either in  $\Delta$  or  $\Gamma$ ) associated with the wave assumes a unique value along this surface. Therefore, a surface  $T(\mathbf{r}) = T_0$  can be interpreted as a *wavefront* and  $T$  assumes the significance of *travel time* necessary to the propagating wavefront to reach that location. We know that the gradient of a scalar field is a vector that is always normal to its iso-surfaces (see Appendix 1). Therefore, at any location  $\mathbf{r}$ ,  $\nabla T$  is normal to the wavefront passing through  $\mathbf{r}$  (Fig. 9.1). Starting from a seismic source located at  $\mathbf{r}_0$  and considering a sequence of wavefronts with arrival times  $T = k\Delta T$  ( $k = 1, 2, \dots$ ) for some small time interval  $\Delta T$ , we can build a

line  $\mathbf{r} = \mathbf{r}(\zeta)$  in the 3-D space by the following procedure. We select an initial arbitrary direction  $\mathbf{n}_0 = \mathbf{n}$ . Then, we link the source point  $\mathbf{r}_0$  to the point  $\mathbf{r}_1 = \mathbf{r}_0 + \mathbf{n}_0 \Delta\zeta_0 = \mathbf{r}_0 + \mathbf{n}_0 c(\mathbf{r}_0) \Delta T$ . At the next step, we consider the versor  $\mathbf{n}_1 = \nabla T / |\nabla T|$  and link  $\mathbf{r}_1$  to a new point  $\mathbf{r}_2 = \mathbf{r}_1 + \mathbf{n}_1 \Delta\zeta_1 = \mathbf{r}_1 + \mathbf{n}_1 c(\mathbf{r}_1) \Delta T$ . At any successive step, we set  $\mathbf{n}_k = \nabla T / |\nabla T|$  and link the current point  $\mathbf{r}_k$  to a new point  $\mathbf{r}_{k+1} = \mathbf{r}_k + \mathbf{n}_k \Delta\zeta_k = \mathbf{r}_k + \mathbf{n}_k c(\mathbf{r}_k) \Delta T$ . For  $\Delta T$  sufficiently small, this algorithm generates a (generally) curved line that is called a *seismic ray* (Fig. 9.1). Of course, we can build infinitely many rays starting from  $\mathbf{r}_0$  simply changing the initial arbitrary direction  $\mathbf{n}$ .

At any step, the parameter  $\zeta = \Delta\zeta_0 + \Delta\zeta_1 + \dots$  represents the total distance from the source, so that the position vector  $\mathbf{r}$  can be considered as a function of the parameter  $\zeta$ . We can also build a *slowness vector*  $\mathbf{s}$  such that  $\hat{\mathbf{s}} = \nabla T / |\nabla T|$  and with magnitude  $s = 1/c$ . Using this notation, we have that the eikonal equation assumes the form:

$$\nabla T(\mathbf{r}) = s(\mathbf{r}) \quad (9.14)$$

Assuming that the position along a seismic ray is parametrized by an equation  $\mathbf{r} = \mathbf{r}(\zeta)$ , where  $\zeta$  is distance from the source, then the infinitesimal variation of  $\mathbf{r}$  along the ray will be given by:

$$d\mathbf{r} = \frac{\nabla T(\mathbf{r})}{s} d\zeta \quad (9.15)$$

In fact,  $c\nabla T(\mathbf{r}) = (1/s)\nabla T(\mathbf{r})$  is always a versor normal to the wavefront. To determine the variation of arrival time along the ray, corresponding to an infinitesimal variation of  $\mathbf{r}$ , we will use the directional derivative of  $T$  along the tangent to the ray at  $\mathbf{r}$ :

$$\frac{dT}{d\zeta} = \nabla T(\mathbf{r}) \cdot \frac{d\mathbf{r}}{d\zeta} = \nabla T(\mathbf{r}) \cdot \frac{\nabla T(\mathbf{r})}{s} = s \quad (9.16)$$

This equation confirms our previous interpretation of  $T$  as a travel time from the seismic source to the wavefront. Now we want to find

an equation that allows to determine  $\mathbf{r} = \mathbf{r}(\zeta)$ , hence the seismic ray geometry, directly from the slowness field  $s = s(\mathbf{r})$ . By (9.15) we have that:

$$\frac{ds}{d\zeta} = \frac{d}{d\zeta} \left( s(\mathbf{r}) \frac{d\mathbf{r}}{d\zeta} \right) = \frac{d}{d\zeta} \nabla T(\mathbf{r}) = \nabla \left( \frac{dT}{d\zeta} \right) \quad (9.17)$$

Finally, using (9.16) we obtain a differential equation for the ray that does not depend from the travel time  $T$ :

$$\frac{d}{d\zeta} \left( s(\mathbf{r}) \frac{d\mathbf{r}}{d\zeta} \right) = \nabla s(\mathbf{r}) \quad (9.18)$$

This equation, which allows to determine the seismic ray geometry as a function of the slowness field, is called the *seismic ray equation*. It can be solved easily by finite differences to obtain the function  $\mathbf{r} = \mathbf{r}(\zeta)$  given an initial direction  $\mathbf{n}$ , granted that the variables  $s$  and  $\nabla s$  are known at any point. In the case of a homogeneous region, (9.18) reduces to  $d^2\mathbf{r}/d\zeta^2 = \mathbf{0}$ , which has the general solution:  $\mathbf{r}(\zeta) = \mathbf{a}\zeta + \mathbf{b}$ ,  $\mathbf{a}$  and  $\mathbf{b}$  being constant vectors. This is clearly a straight line in the direction  $\mathbf{a}$  and passing through the point  $\mathbf{r}_0 = \mathbf{b}$ .

## 9.2 Geometrical Spreading

Now let us turn our attention to the amplitude transport Eq. (9.9). We are going to prove that it determines how the amplitude  $A$  is transported along a seismic ray. Substituting in the first term at the left-hand side the gradient of  $T$  by the slowness vector  $\mathbf{s}$  (Eq. 9.14) we have that:

$$2\nabla A(\mathbf{r}) \cdot \mathbf{s}(\mathbf{r}) + A(\mathbf{r}) \nabla^2 T(\mathbf{r}) = 0 \quad (9.19)$$

We note that  $\nabla A(\mathbf{r}) \cdot \mathbf{s}(\mathbf{r})$  is at any point  $\mathbf{r}$  proportional to the directional derivative of  $A$  in the direction of  $\nabla T$ , thereby (9.19) can be considered as an ordinary differential equation along the curved line representing the seismic ray. If rays are described by parametric equations  $\mathbf{r} = \mathbf{r}(\zeta)$ , then the variation of amplitude along a seismic ray can be expressed as a function of the parameter  $\zeta$ .

We have:

$$\frac{dA}{d\zeta} = \frac{\partial A}{\partial x_i} \frac{dx_i}{d\zeta} = \nabla A \cdot \frac{d\mathbf{r}}{d\zeta} = \nabla A \cdot \frac{\nabla T}{s} = \mathbf{k} \cdot \nabla A \quad (9.20)$$

where  $\mathbf{k} = \mathbf{k}(\mathbf{r}) = \nabla T/s$  is the versor tangent to the seismic ray at  $\mathbf{r}$ . Therefore:

$$s \cdot \nabla A = s \mathbf{k} \cdot \nabla A = s \frac{dA}{d\zeta} \quad (9.21)$$

Using (9.16), we can convert the derivative of  $A$  with respect to  $\zeta$  into a derivative with respect to the travel time  $T$ . We have:

$$s \cdot \nabla A = s \frac{dA}{dT} \frac{dT}{d\zeta} = s^2 \frac{dA}{dT} \quad (9.22)$$

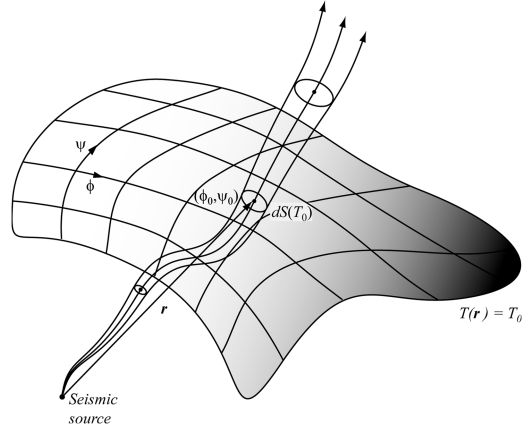
Substituting this result into (9.19) and expressing  $A$  as a function of  $T$  we obtain a first-order ordinary differential equation along the ray:

$$2s^2(\mathbf{r}) \frac{dA}{dT} + A(T) \nabla^2 T(\mathbf{r}) = 0 \quad (9.23)$$

To solve (9.23), it is necessary to know the Laplacian of the travel time  $T$  along the seismic ray. Let us consider the surface of a wavefront  $T(\mathbf{r}) = T_0$  (Fig. 9.2). The position of a point on this surface can be specified through two curvilinear coordinates  $(\phi, \psi)$ :

$$\begin{cases} x = x(\phi, \psi) \\ y = y(\phi, \psi) \\ z = z(\phi, \psi) \end{cases} \quad (9.24)$$

For example, in the case of a spherical wavefront propagating from the Earth's center, we could use spherical coordinates of longitude and latitude to identify the intersection of a seismic ray with the propagating wavefront. A set of pairs  $(\phi, \psi)$  defines the intersection of a family of rays and the *eikonal surface*  $T(\mathbf{r}) = T_0$ . Let  $(\phi_0, \psi_0)$  be the coordinates of a point along this surface, and consider the set of seismic rays that intersects the wavefront in a neighbor  $\Phi$  of this point.



**Fig. 9.2** Curvilinear coordinates on a wavefront and ray tube at a point  $(\phi_0, \psi_0)$

Let us assume that for any eikonal surface  $\Phi$  is defined through *fixed* displacements  $d\phi$  and  $d\psi$  from  $(\phi_0, \psi_0)$ :

$$\begin{aligned} \Phi(\phi_0, \psi_0) = \{(\phi, \psi) : \phi_0 \leq \phi \leq \phi_0 + d\phi ; \\ \psi_0 \leq \psi \leq \psi_0 + d\psi\} \end{aligned} \quad (9.25)$$

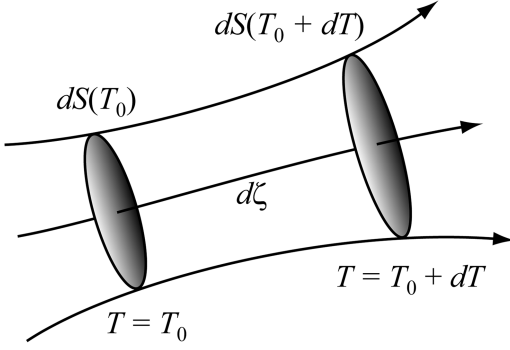
The set of rays crossing the wavefront through  $\Phi$  forms what is known as a *ray tube* (Fig. 9.2). The area  $dS$  of  $\Phi$  is clearly a function of  $T_0$ , as shown in Fig. 9.2. It can be calculated as follows:

$$\begin{aligned} dS(T_0) &= \left| \frac{d\mathbf{r}}{d\phi} d\phi \times \frac{d\mathbf{r}}{d\psi} d\psi \right| \\ &= \left| \frac{d\mathbf{r}}{d\phi} \times \frac{d\mathbf{r}}{d\psi} \right| d\phi d\psi \\ &\equiv J(T_0, \phi_0, \psi_0) d\phi d\psi \end{aligned} \quad (9.26)$$

The quantity:

$$J(T_0, \phi_0, \psi_0) = \left| \frac{d\mathbf{r}}{d\phi} \times \frac{d\mathbf{r}}{d\psi} \right| \quad (9.27)$$

is called *geometrical spreading* of the tube and describes the focusing and defocusing of seismic rays. It increases when the distance between rays raises with the distance from the source. Conversely, if the rays converge to a point, then  $J$



**Fig. 9.3** The volume element formed by two eikonal surfaces and a tube

decreases progressively to zero at the point of convergence. Let us consider now the volume element  $dV$  formed by a ray tube and two closely spaced wavefronts at  $T = T_0$  and  $T = T_0 + dT$ , as shown in Fig. 9.3.

To evaluate  $dV$ , we observe that the triple product  $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$  of three vectors represents the volume of the parallelepiped formed by the three vectors. Therefore:

$$\begin{aligned}
 dV &= \left| \frac{d\mathbf{r}}{d\zeta} d\zeta \cdot \left( \frac{d\mathbf{r}}{d\phi} d\phi \times \frac{d\mathbf{r}}{d\psi} d\psi \right) \right| \\
 &= \left| \frac{d\mathbf{r}}{d\zeta} \cdot \left( \frac{d\mathbf{r}}{d\phi} \times \frac{d\mathbf{r}}{d\psi} \right) \right| d\zeta d\phi d\psi \\
 &= \left| \begin{vmatrix} \frac{\partial x}{\partial \zeta} & \frac{\partial y}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \\ \frac{\partial x}{\partial \phi} & \frac{\partial y}{\partial \phi} & \frac{\partial z}{\partial \phi} \\ \frac{\partial x}{\partial \psi} & \frac{\partial y}{\partial \psi} & \frac{\partial z}{\partial \psi} \end{vmatrix} \right| d\zeta d\phi d\psi \\
 &\equiv D(\zeta, \phi, \psi) d\zeta d\phi d\psi \\
 &= D(T, \phi, \psi) dT d\phi d\psi
 \end{aligned} \tag{9.28}$$

where we have used (9.16) to transform derivatives with respect to  $\zeta$  into derivatives with respect to  $T$ . On the other hand, from (9.26) we have that:

$$\begin{aligned}
 dV &= dS d\zeta = J(T, \phi, \psi) d\phi d\psi d\zeta \\
 &= \frac{1}{s(T)} J(T, \phi, \psi) d\phi d\psi dT
 \end{aligned} \tag{9.29}$$

where the slowness  $s$  has been expressed as a function of  $T$  through  $\mathbf{r}$ . Finally, comparing (9.28) with (9.29) gives the following expression for the geometrical spreading at  $T$ :

$$J(T, \phi, \psi) = s D(T, \phi, \psi) \tag{9.30}$$

Let us consider now the integral of the vector field  $\nabla T$  over the surface of  $dV$ . It is given by:

$$\begin{aligned}
 \oint_{dV} \nabla T \cdot d\mathbf{S} &= s(T + dT) dS(T + dT) \\
 &- s(T) dS(T) = [s(T + dT) J(T + dT) \\
 &- s(T) J(T)] d\phi d\psi
 \end{aligned}$$

Therefore, using the definition of divergence (see Appendix 1) and (9.29), we have:

$$\begin{aligned}
 \nabla^2 T &= \nabla \cdot \nabla T = \lim_{dV \rightarrow 0} \frac{1}{dV} \oint_{dV} \nabla T \cdot d\mathbf{S} \\
 &= \lim_{dV \rightarrow 0} \frac{[s(T + dT) J(T + dT) - s(T) J(T)] d\phi d\psi}{\frac{1}{s(T)} J(T) d\phi d\psi dT} = \\
 &= \lim_{dV \rightarrow 0} \frac{s(T) [s(T + dT) J(T + dT) - s(T) J(T)]}{J(T) dT} \\
 &= \frac{s(T)}{J(T)} \frac{d}{dT} (s(T) J(T))
 \end{aligned} \tag{9.31}$$

Substituting this expression into the transport equation (9.23) and dividing by  $s$  gives:

$$2s(T) \frac{dA}{dT} + \frac{1}{J(T)} A(T) \frac{d}{dT} (s(T) J(T)) = 0 \tag{9.32}$$

This is a first order differential equation with separable variables, with solution:

$$A(T) = \frac{c(\phi, \psi)}{\sqrt{s(T) J(T)}} \tag{9.33}$$

where  $c$  is a constant not depending from  $T$ , which can be expressed in terms of seismic ray parameters  $\phi$  and  $\psi$ . This solution shows that the amplitude decreases when the geometrical spreading increases. More precisely, the

amplitude depends from the inverse square root of the ray tube cross-section. To determine the constant  $c$ , let us consider again the transport equation in the form (9.9). Multiplying this equation by  $A(\mathbf{r})$ , it is possible to rewrite it in the form:

$$\nabla \cdot (A^2(\mathbf{r}) \nabla T(\mathbf{r})) = 0 \quad (9.34)$$

Let us consider two points  $\mathbf{r}'$  and  $\mathbf{r}''$  of a seismic ray and the corresponding wavefronts  $T = T'$  and  $T = T''$ . A tube segment between  $\mathbf{r}'$  and  $\mathbf{r}''$  is a region  $\mathbf{R}$  bounded by the two wavefronts and by the tube walls. We know that the gradients of  $T$  are always parallel to the lateral surface of the tube and perpendicular to its end faces. Therefore, integrating (9.34) over the region  $\mathbf{R}$  we have:

$$\begin{aligned} 0 &= \int_{\mathbf{R}} \nabla \cdot (A^2(\mathbf{r}) \nabla T(\mathbf{r})) dV \\ &= \oint_{\Gamma(\mathbf{R})} A^2(\mathbf{r}) \nabla T(\mathbf{r}) \cdot d\mathbf{S} = \oint_{\Gamma(\mathbf{R})} A^2(\mathbf{r}) s(\mathbf{r}) dS \\ &= A^2(\mathbf{r}'') s(\mathbf{r}'') dS(\mathbf{r}'') - A^2(\mathbf{r}') s(\mathbf{r}') dS(\mathbf{r}') \end{aligned} \quad (9.35)$$

Hence:

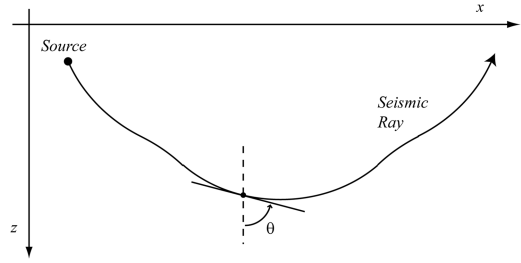
$$A^2(\mathbf{r}') s(\mathbf{r}') dS(\mathbf{r}') = A^2(\mathbf{r}'') s(\mathbf{r}'') dS(\mathbf{r}'') \quad (9.36)$$

By (9.26), we also know that the cross-sections of a tube can be expressed in terms of geometrical spreading  $J$ . Therefore, if we have determined the quantities  $A$  and  $J$  at some point  $\mathbf{r} = \mathbf{r}_0$ , then we can determine the amplitude at any other point rewriting (9.33) as follows:

$$A(\mathbf{r}) = A(\mathbf{r}_0) \sqrt{\frac{J(\mathbf{r}_0) s(\mathbf{r}_0)}{J(\mathbf{r}) s(\mathbf{r})}} \quad (9.37)$$

### 9.3 Snell's Law

Let us consider now a material such that the velocity  $c$  (either  $\alpha$  or  $\beta$ ) depends only from the depth  $z$ , so that  $s = s(z)$ . In this instance, the quantity:



**Fig. 9.4** Typical geometry of a seismic ray in a material where the velocity  $c$  depends only from depth.  $\theta$  is the incidence angle

$$\mathbf{p} = \mathbf{k} \times s(z) \frac{d\mathbf{r}}{d\zeta} \quad (9.38)$$

is an *invariant* of the seismic ray. In fact, by the ray Eq. (9.18) we have:

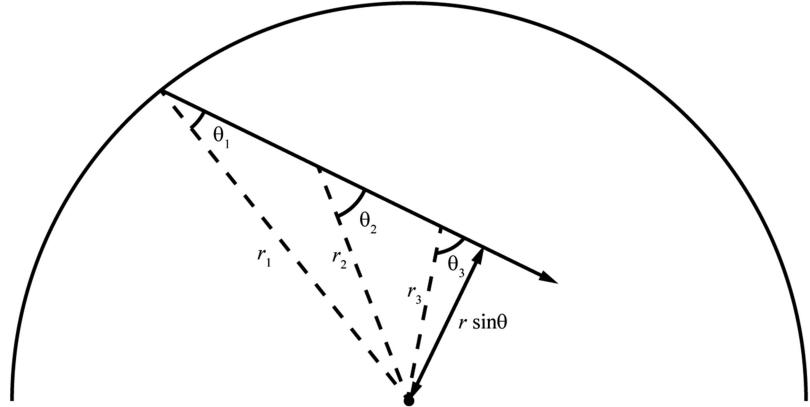
$$\frac{d\mathbf{p}}{d\zeta} = \mathbf{k} \times \frac{d}{d\zeta} \left( s(z) \frac{d\mathbf{r}}{d\zeta} \right) = \mathbf{k} \times \nabla s(z) = \mathbf{0} \quad (9.39)$$

The definition (9.38) implies that  $\mathbf{p}$  is normal to the direction of propagation, thereby any ray will lie in a vertical plane. Furthermore, the conservation of  $\mathbf{p}$  requires that also its magnitude must be constant. Let us define the *incidence angle*  $\theta = \theta(z)$  as the angle that a seismic ray forms with the vertical at any point (Fig. 9.4). The magnitude of  $\mathbf{p}$  is called the *ray parameter* and is a fundamental invariant in seismology, because it is uniquely associated with a seismic ray. By (9.38) this quantity satisfies the following simple law:

$$p = s(z) \sin \theta(z) \quad (9.40)$$

The conservation law (9.40) for  $p$  along a seismic ray is known as *Snell's law*. It allows to determine the geometry of any seismic ray given a velocity function  $c = c(z)$  and a *take-off* angle  $\theta_0$ , which is the incidence angle at the source point. Clearly, (9.40) results from a flat-Earth approximation that is adequate only in the case of shallow propagation, say up to 30 km depth. In a spherical Earth with radial symmetry, the slowness  $s$  and the incidence angle  $\theta$  depend

**Fig. 9.5** In a spherical homogeneous Earth, seismic rays are straight lines but the incidence angle changes with depth. In this instance, the quantity  $r \sin\theta(r)$  is invariant



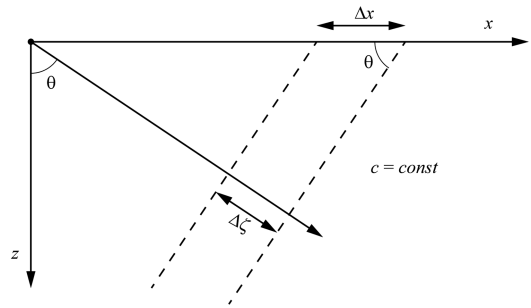
only upon the distance  $r$  from the Earth's center, so that  $s = s(r)$  and  $\theta = \theta(r)$ . In this instance, the vector invariant assumes the form:

$$\mathbf{p} = \mathbf{r} \times s(r) \frac{d\mathbf{r}}{d\zeta} \quad (9.41)$$

Also in this case the seismic rays lie on vertical planes, and it is easy to prove that Snell's law assumes the form:

$$p = s(r)r \sin\theta(r) \quad (9.42)$$

In fact, in this instance  $\sin\theta$  is not anymore invariant when the slowness is constant (Fig. 9.5). Conversely, the example of Fig. 9.5 shows that in the spherical context the quantity  $r \sin\theta$  will be conserved for  $s(r) = \text{const}$ . It is important to note that both the flat approximation (9.40) and the spherical conservation law (9.42) do not take into account of *lateral inhomogeneities*, so that they are only applicable when the seismic velocity depends from a unique parameter ( $z$  or  $r$ ).



**Fig. 9.6** A downward propagating plane wave from a seismic source at the Earth's surface. The material is homogeneous, so that the seismic rays are straight lines and the incidence angle  $\theta$  is constant

$$\Delta\zeta = \frac{\Delta T}{s} = \Delta x \sin\theta \quad (9.43)$$

Therefore, by Snell's law:

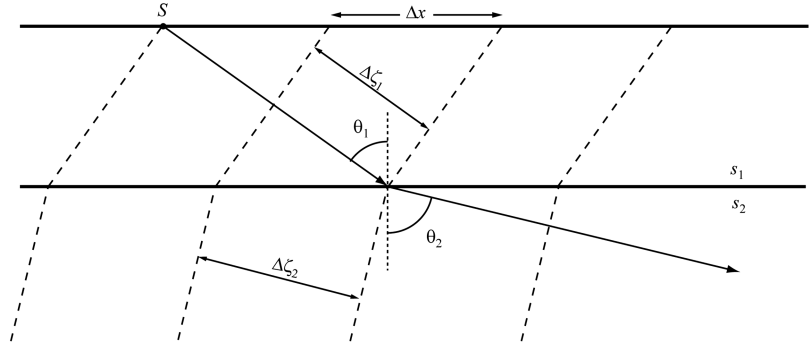
$$\frac{\Delta T}{\Delta x} = s \sin\theta = p \quad (9.44)$$

This relation implies that a measurement of arrival times at different seismic stations allows to determine the seismic parameter  $p$ . It also implies that  $p$  is the *apparent* slowness of the wavefront in the horizontal direction. Therefore,  $p$  is sometimes referred to as the *horizontal slowness* of the seismic ray. Let us assume now that the wave strikes a horizontal discontinuity in the elastic parameters, so that we can separate the material into an upper layer with velocity  $c_1$  and a lower layer with higher velocity  $c_2 > c_1$  (Fig. 9.7). For  $s_2 < s_1$ , Snell's law predicts an *increase*

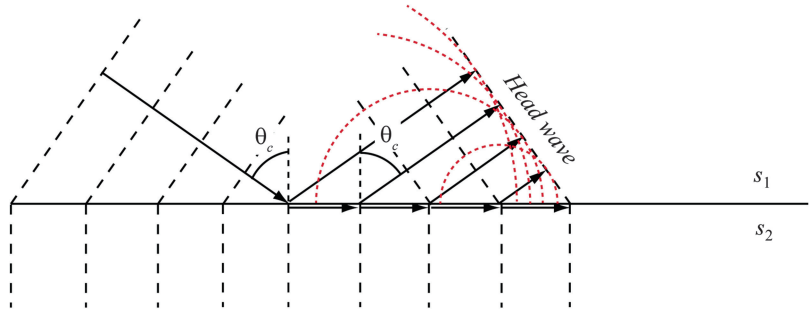
## 9.4 1-D Velocity Models

Let us consider now a plane wave that propagates through a homogeneous material with constant slowness  $s$  (Fig. 9.6). The wavefronts at time  $T$  and  $T + \Delta T$  are separated by a distance  $\Delta\zeta$  along a seismic ray. At the Earth's surface, the wavefront arrivals have a different separation  $\Delta x$ , which depends from both  $\Delta\zeta$  and the slowness  $s$ :

**Fig. 9.7** Wavefront propagation (dashed lines) and ray refraction through a seismic discontinuity separating two homogeneous media. Higher velocity in the lower layer determines an increase of spacing  $\Delta\zeta$



**Fig. 9.8** Formation of a head wave at the turning point interface. This is formed by the envelope of secondary spherical waves generated along the discontinuity (red lines)



of both the spacing between wavefronts and the incidence angle, so that  $\Delta\zeta_2 > \Delta\zeta_1$  and  $\theta_2 > \theta_1$ . Conversely, for  $s_2 > s_1$  we would have downward bending of the seismic rays. This phenomenon is termed *seismic refraction* and is analogous to the refraction in optics. From (9.40) we have:

$$p = s_1 \sin \theta_1 = s_2 \sin \theta_2 \quad (9.45)$$

Therefore, the incidence angle of the transmitted wave will be given by:

$$\theta_2 = \arcsin \left( \frac{s_1}{s_2} \sin \theta_1 \right) \quad (9.46)$$

A *critical incidence angle*,  $\theta_c$ , is defined as an incidence angle such that the transmitted ray has  $\theta_2 = 90^\circ$ . In these case, we say that the ray is at its *turning point*. From (9.46) we easily obtain:

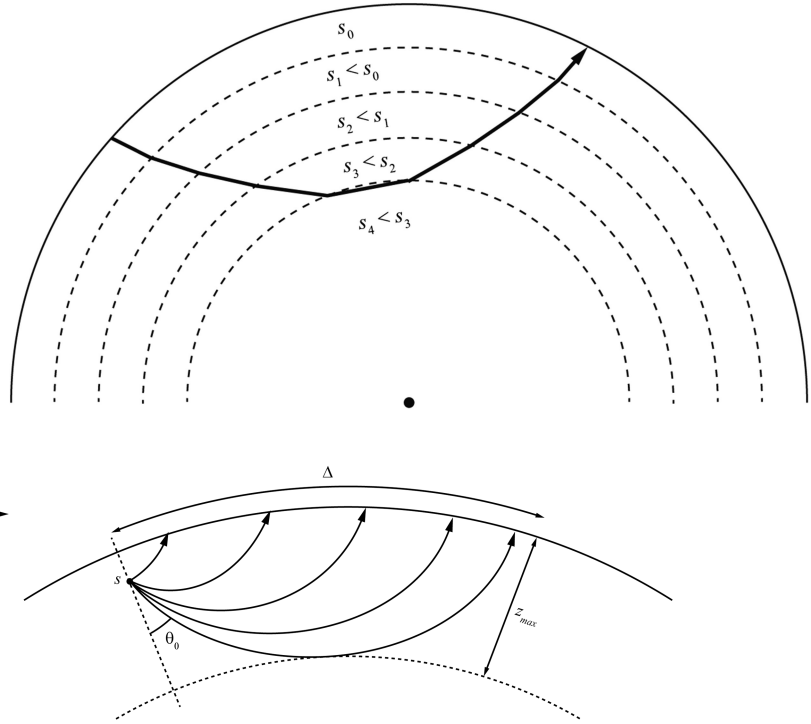
$$\theta_c = \arcsin \left( \frac{s_2}{s_1} \right) \quad (9.47)$$

At the turning point, we have that the slowness coincides with the ray parameter and the ray direction becomes horizontal. The corresponding

wave propagates horizontally (in the ideal limit of plane waves) through the lower layer with velocity  $1/s_2$  as shown in Fig. 9.8.

During its propagation, the transmitted wave excites the interface surface between the upper and lower layer, determining by Huygens's principle the formation of a *head wave* (sometimes called a *bow wave*) that travels upwards in the direction of the Earth's surface. The seismic rays associated with this secondary source have the same incidence angle,  $\theta_c$ , as the downgoing rays (Fig. 9.8). A complete description of the head waves is not possible in the context of ray theory and requires an analysis in terms of waves rather than seismic rays. It is important to note that the velocity of propagation of the head wave in Fig. 9.8 is  $1/s_1$ , while the boundary perturbation always moves at higher velocity  $1/s_2$  along the interface. Therefore, any spherical wavefront generated along the discontinuity surface and propagating through the upper layer will be overtaken by its source, so that each new wavefront will start beyond the last one. The resulting envelope wave is V-shaped and tangent to all the emitted spherical wave fronts, as shown in Fig. 9.8. Let us consider now a layered material, such that the

**Fig. 9.9** Seismic ray geometry for a layered spherical Earth model, such that the seismic wave velocity  $c$  increases with the layer depth. The existence of a turning point does not always depend from the achievement of a critical incidence angle (Eq. 9.47). In fact, in this example  $\theta_4 < 90^\circ$



**Fig. 9.10** Seismic rays from a source point  $S$  for a spherical Earth model such that the seismic wave velocity  $c$  increases linearly with depth. Both the *range* (horizontal

angular distance travelled by the ray) and the turning point depth increase as the take-off angle decreases

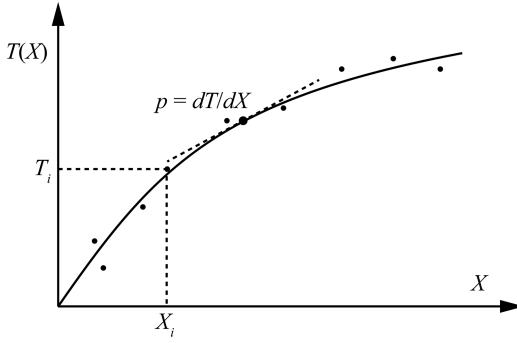
seismic velocity is constant within each layer  $L_k$  ( $k = 1, 2, \dots, n$ ) but increases progressively traveling into deeper layers. In this instance, the seismic rays will be broken lines that bend *upwards* until the critical angle is reached. According to Snell's law (9.40) we have that:

$$p = s_1 \sin \theta_1 = s_2 \sin \theta_2 = \dots = s_n \sin \theta_c = s_{n+1} \quad (9.48)$$

where  $\theta_1 < \theta_2 < \dots < \theta_c$ . Therefore, if  $s_n$  is the slowness within the layer where the incidence angle assumes the critical value  $\theta_c$ , then the slowness in the underlying layer will coincide with the ray parameter  $p$ . In the limit case that the velocity increases continuously with depth, we still have a turning point at some depth  $z = z_{max}$  and it results:  $s(z_{max}) = p$ . Head waves form an important class of seismic waves in exploration geophysics, where the depths reached by artificial waves do not exceed a few tens km and the flat Earth approximation is effective.

In a spherical Earth, the upward propagation of seismic waves and the existence of turning points is a consequence of the spherical Snell's law (9.42). In this instance, for a velocity  $c$  that continuously raises with depth, the downgoing segment of a seismic ray will bend progressively through increasing incidence angles, as far as a turning point depth  $z_{max}$  is reached. This is illustrated in Fig. 9.9 and does not require the formation of head waves. In general, the maximum depth  $z_{max}$  of a seismic ray and the maximum angular distance,  $\Delta$ , from the source depend from the take-off angle  $\theta_0$  (Fig. 9.10).

We shall face now the problem of determining the range and the arrival time of a seismic ray. Let us consider first the flat Earth approximation and assume to have placed a number of receivers (i.e., seismometers) at distances  $X_1, X_2, \dots, X_n$  from a seismic source  $S$ . If the velocity  $c = c(z)$  is a monotonically increasing function of depth, then a measurement of the corresponding arrival times  $T_1, T_2, \dots, T_n$  allows to fit these data by a



**Fig. 9.11** Travel-time fitting curve for a monotonically increasing velocity function. Dots are observed travel times  $T_i$  at locations  $X_i$ . The tangent to the regression curve (dashed line) represents the ray parameter of the seismic ray arriving at distance  $X$  from the source

continuous curve  $T = T(X)$  of the arrival time as a function of the distance  $X$  from the seismic source (that is, the *range*). By (9.44) we have that the slope of this curve will be given by:

$$\frac{dT}{dX} = p = s(z_{\max}) \quad (9.49)$$

Clearly, the derivative  $dT/dX$  is a decreasing function of the range, because the ray parameter (hence the slowness at the turning point) decreases with increasing range (see Fig. 9.8). Therefore, a plot of the *travel time curve*  $T = T(X)$  has the typical shape illustrated in Fig. 9.11. In this plot, each point of the regression curve represents a different seismic ray, whose parameter is given by the slope of the curve. When the seismic velocity is a known function of  $z$ , it is possible to determine analytically both the travel-time  $T$  and the range  $X$  for any value of the parameter  $p$ , hence for any seismic ray from a known source. We know that the horizontal component of the slowness vector  $s = s(z)$  coincides with  $p$  (Eq. 9.44):  $s_x = p$ . Regarding the vertical component,  $s_z$ , it is given by:

$$s_z(z) = s(z) \cos \theta(z) = \sqrt{s^2(z) - p^2} \quad (9.50)$$

At the turning point, we have that  $p = s$  and  $s_z = 0$ . It is easy to find expressions for  $X = X(p)$  and  $T = T(p)$ . Let us consider an infinitesimal ray path segment  $d\zeta$ . By Snell's law (9.40) we have that:

$$\begin{aligned} \frac{dx}{d\zeta} &= \sin \theta = \frac{p}{s}; \quad \frac{dz}{d\zeta} = \cos \theta \\ &= \sqrt{1 - \sin^2 \theta} = \sqrt{1 - p^2/s^2} = \frac{1}{s} \sqrt{s^2 - p^2} \end{aligned} \quad (9.51)$$

Therefore,

$$\frac{dx}{dz} = \frac{dx}{d\zeta} \frac{d\zeta}{dz} = \frac{p}{\sqrt{s^2 - p^2}} \quad (9.52)$$

This expression can be easily integrated to determine the  $x$  component of points along the seismic ray with parameter  $p$ , assuming a source located at the origin of the reference frame. To determine the range  $X$  associated with  $p$ , we must take into account that in laterally homogeneous models the distance at which the upgoing segment of the seismic ray reaches the Earth's surface is twice the horizontal distance between the seismic source and the turning point. Therefore,

$$X(p) = 2p \int_0^{z_{\max}} \frac{dz}{\sqrt{s^2(z) - p^2}} \quad (9.53)$$

A similar procedure allows to determine the arrival time  $T$ . Because  $dT = sd\zeta$  (Eq. 9.16), we have:

$$\frac{dT}{dz} = \frac{dT}{d\zeta} \frac{d\zeta}{dz} = \frac{s^2}{\sqrt{s^2 - p^2}} \quad (9.54)$$

Integrating from the Earth's surface to  $z_{\max}$  and multiplying by 2, we easily obtain the arrival time and distance  $X$ :

$$T(p) = 2 \int_0^{z_{\max}} \frac{s^2(z) dz}{\sqrt{s^2(z) - p^2}} \quad (9.55)$$

When the transmitting domain can be considered as a stack of homogeneous layers, the integral solutions (9.53) and (9.55) are substituted by sums.

We have:

$$X(p) = 2p \sum_i \frac{\Delta h_i}{\sqrt{s_i^2 - p^2}}; \quad s_i > p \quad (9.56)$$

$$T(p) = 2 \sum_i \frac{s_i^2 \Delta h_i}{\sqrt{s_i^2 - p^2}}; s_i > p \quad (9.57)$$

where  $\Delta h_i$  is the thickness of the  $i$ -th layer. When the velocity model has continuous gradients, the approximate solutions (9.56) and (9.57) are inadequate, because of the large number of quasi-homogeneous layers that would be needed to obtain reliable results. In this instance, a better approach is to assume a velocity model such that  $c = c(z)$  is a broken line and evaluate the expressions (9.53) and (9.55) for each segment. For example, if  $c_i$  ( $i = 0, 2, \dots, n$ ) are  $n + 1$  representative control points at depths  $z_i$  along an experimental velocity profile, we can assume that the velocity varies linearly between any pair of successive control points, so that:

$$c(z) = \frac{c_{i+1} - c_i}{z_{i+1} - z_i}(z - z_i) + c_i \equiv m_i(z - z_i) + c_i; \\ \text{for } z_i \leq z \leq z_{i+1}; i = 0, 1, \dots, n - 1 \quad (9.58)$$

In this case, the variation of slowness with depth will be given by:

$$ds(z) = \frac{ds}{dc} dc(z) = -\frac{1}{c^2(z)} dc(z) \\ = -\frac{m_i}{c^2(z)} dz = -m_i s^2(z) dz; \\ \text{for } z_i < z < z_{i+1}; i = 0, 1, \dots, n - 1 \quad (9.59)$$

If we substitute the functions (9.58) into the integrals (9.53) and (9.55) and change the integration variable to  $ds$ , we have that each segment gives a contribution  $X_i$  to the total range  $X$  that can be expressed as follows:

$$X_i(p) = p \int_{z_i}^{z_{i+1}} \frac{dz}{\sqrt{s^2(z) - p^2}} \\ = -\frac{p}{m_i} \int_{s_i}^{s_{i+1}} \frac{ds}{s^2 \sqrt{s^2 - p^2}} = -\frac{\sqrt{s^2 - p^2}}{m_i p s} \Big|_{s_i}^{s_{i+1}} \\ = -\frac{\sqrt{1 - p^2[m_i(z - z_i) + c_i]^2}}{p m_i} \Big|_{z_i}^{z_{i+1}}$$

$$= -\frac{1}{p m_i} \left( \sqrt{1 - p^2 c_{i+1}^2} - \sqrt{1 - p^2 c_i^2} \right); \\ i = 0, 1, \dots, n - 1 \quad (9.60)$$

Similarly,

$$T_i(p) = \int_{z_i}^{z_{i+1}} \frac{s^2(z) dz}{\sqrt{s^2(z) - p^2}} = -\frac{1}{m_i} \int_{s_i}^{s_{i+1}} \frac{ds}{\sqrt{s^2 - p^2}} \\ = -\frac{1}{m_i} \ln \left( s + \sqrt{s^2 - p^2} \right) \Big|_{s_i}^{s_{i+1}} \\ = -\frac{1}{m_i} \ln \left[ \frac{1}{m_i(z - z_i) + c_i} + \sqrt{\frac{1}{[m_i(z - z_i) + c_i]^2} - p^2} \right] \Big|_{z_i}^{z_{i+1}} \\ = -\frac{1}{m_i} \left\{ \ln \left[ \frac{1}{v_{i+1}} \left( 1 + \sqrt{1 - p^2 c_{i+1}^2} \right) \right] \right. \\ \left. - \ln \left[ \frac{1}{v_i} \left( 1 + \sqrt{1 - p^2 c_i^2} \right) \right] \right\} \quad (9.61)$$

The final arrival time  $T(p)$  and range  $X(p)$  are then obtained summing these contributions over the index  $i$  and multiplying the result by two. Let us consider now the generalization of the previous solutions to the spherical Earth. We know that in this instance the horizontal distance  $X$  is substituted by the angular distance  $\Delta$ , so that  $dX = R d\Delta$ ,  $R$  being the Earth's radius. In this case, the ray parameter coincides with the slowness at the turning point only when  $c$  is a continuous and mostly increasing function of the depth. To generalize (9.49), we start as before from (9.43). However, in this case Snell's law in the form (9.40) must be substituted by (9.42). Therefore,

$$\frac{dT}{d\Delta} = R \frac{dT}{dX} = R \frac{p}{R} = p \quad (9.62)$$

This is the generalization of (9.49) to a spherical Earth. Solutions for the arrival time  $T(p)$  and the range  $X(p)$  can be found easily using the same approach discussed previously for the flat Earth approximation. We obtain:

$$\Delta(p) = 2p \int_{r_{\min}}^R \frac{1}{r \sqrt{r^2 s^2(r) - p^2}} dr \quad (9.63)$$

$$T(p) = 2 \int_{r_{\min}}^R \frac{rs^2(r)}{\sqrt{r^2s^2(r) - p^2}} dr \quad (9.64)$$

where  $r_{\min}$  is the distance from the Earth's center at the turning point.

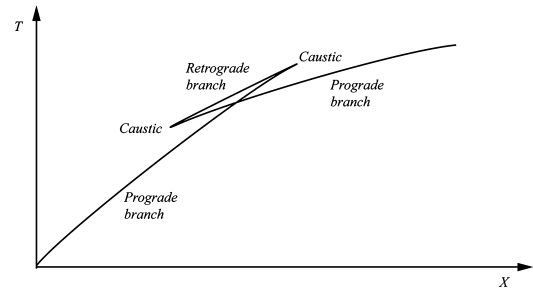
## 9.5 Travel Time Curves

We have seen that the range  $X(p)$  is a decreasing function of the ray parameter  $p$ , as shown in Fig. 9.10. Therefore, we expect that in normal conditions  $dX/dp < 0$ . However, when the velocity profile includes segments characterized by elevated gradient, the behaviour of seismic rays becomes more complicate. In the case of rays with a turning point below or above the high-gradient layer, the corresponding tracts of travel time curve will have a trend similar to that illustrated in Fig. 9.11, thereby the arrival time increases and the parameter decreases for increasing distance from the source. These portions of travel–time curve are called *prograde*.

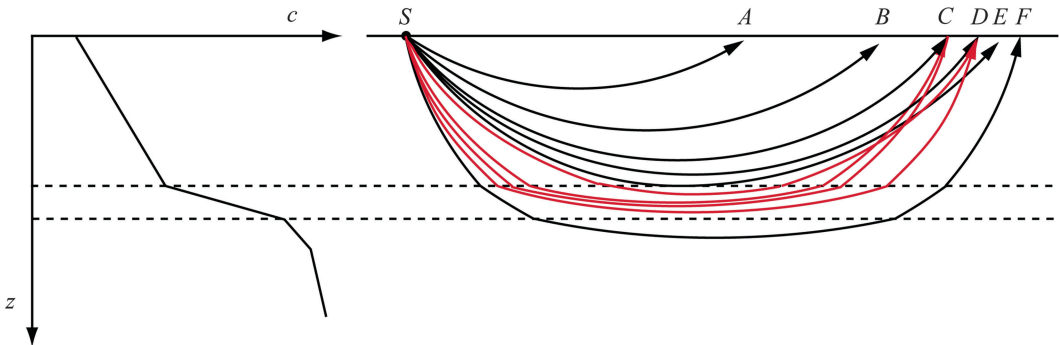
When seismic rays bottom in the high velocity gradient region, the increased bending determines a decrease of range, as illustrated in Fig. 9.12. In this instance, some stations will experience more than one arrival for the same kind of seismic wave. The corresponding segment of travel time curve is called *retrograde*. When a travel time curve includes a retrograde segment, we say that

it contains a *triplication* (Fig. 9.13), while the corresponding cusps at the end points of the retrograde tract are called *caustics*. Caustics are hit by two distinct seismic rays, while any other point in the range of the retrograde branch will be hit by three rays. At the caustics, we have that  $dX/dp = 0$ , because these points separate branches for which  $T$  increases as the parameter  $p$  decreases (prograde branches) by tracts such that the travel time decreases for decreasing values of  $p$  (retrograde tract). The triplication range is always associated with greater energy delivery, because of ray focusing (see Sect. 9.2).

In particular, ray theory predicts infinite amplitude of the seismic waves at the caustics. To prove this, let us consider the distribution of seismic energy along a wavefront in the simple case of a laterally homogeneous model. In Sect. 8.4 we have shown that the average energy density is proportional to the wave amplitude  $A$  and to the angular frequency  $\omega$  (Eq. 8.59). Therefore, we

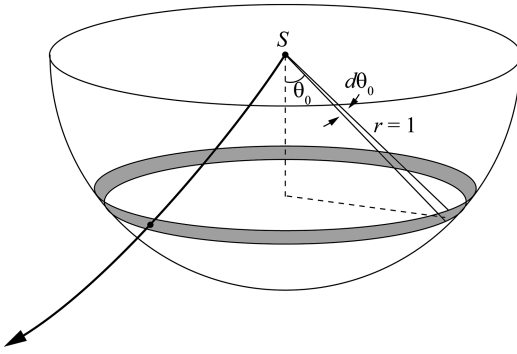


**Fig. 9.13** A travel–time curve containing a triplication



**Fig. 9.12** Normal prograde rays (black lines) and retrograde rays (red lines) associated with high velocity gradient. Seismic rays arriving at stations A–F have decreasing

take-off angles, but rays that bottom in the high-gradient layer will have a shorter range. A single ray incides on A, B, E, and F, while three distinct rays incide on C and D



**Fig. 9.14** Lower focal hemisphere about a seismic source  $S$  and band through which the energy associated with rays having take-off angle between  $\theta_0$  and  $\theta_0 + d\theta_0$  is irradiated

expect that focusing and defocusing of seismic rays associated with geometrical spreading determine the distribution of seismic energy along the wavefronts. Let us assume that a source at some depth below the Earth's surface irradiates isotropically a seismic energy  $E$ , and consider the rays with take-off angle between  $\theta_0$  e  $\theta_0 + d\theta_0$  (Fig. 9.14).

These rays leave a unit sphere about the source through a horizontal band having area  $2\pi \sin\theta_0 d\theta_0$  (Fig. 9.14). Because the total area of the unit sphere is  $4\pi$  and the radiation of seismic energy is isotropic, the energy transmitted through the band is:

$$dE(\theta_0) = \frac{1}{2} E \sin \theta_0 d\theta_0 \quad (9.65)$$

The corresponding rays strike the Earth's surface through a ring belt having area  $2\pi X(\theta_0) dX$ , while the wavefront portion will have area  $2\pi X(\theta_0) \cos\theta_0 dX$ , because the incidence angle at  $X$  coincides with the take-off angle. Therefore, the energy distributed along this wavefront portion will be given by:

$$dE(X) = 2\pi \langle E(X) \rangle X(\theta_0) \cos \theta_0 dX \quad (9.66)$$

where  $\langle E(X) \rangle$  is the average energy density at  $X$ . By the law of conservation of energy, we must have:  $dE(\theta_0) = dE(X)$ . Therefore,

$$\langle E(x) \rangle = \frac{E}{4\pi X(\theta_0)} \tan \theta_0 \left| \frac{d\theta_0}{dX} \right| \quad (9.67)$$

By Snell's law we have that  $s_0 \sin\theta_0 = p$ , thereby  $dp(\theta_0) = s_0 \cos\theta_0 d\theta_0$ . Furthermore, it results:  $\tan \theta_0 = p/(s_0^2 - p^2)^{1/2}$  and  $d\theta_0/dX = (s_0^2 - p^2)^{-1/2} dp/dX$ . Therefore, we obtain the following expression for the average energy density at distance  $X$ :

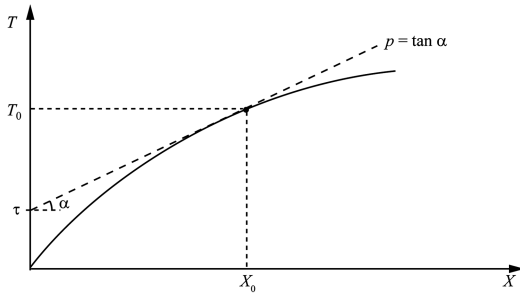
$$\begin{aligned} \langle E(X) \rangle &= \frac{pE}{4\pi X (s_0^2 - p^2)} \left| \frac{dp}{dX} \right| \\ &= \frac{pE}{4\pi X (s_0^2 - p^2) |dX/dp|} \quad (9.68) \end{aligned}$$

The role of the term  $|dX/dp|$  in expression (9.68) is quite intuitive. When this quantity is small, a large number of seismic rays with different parameters strike the Earth's surface at comparable distance  $X$ . In this instance, the energy density and the amplitude of seismic waves attain elevated values. By contrast, when  $|dX/dp|$  is high, rays with comparable parameters distribute on a large area, thereby the density of energy and the amplitude are small. Finally, for  $|dX/dp| = 0$ , hence at the caustics, the energy density (9.68) is infinite. Clearly, this result is valid only in the infinite frequency limit of ray theory; real amplitudes of the seismic waves and the energy density are elevated but finite at the caustics.

So far, we have represented travel time curves in terms of  $(T, X)$  pairs. In this context,  $T = T(X)$  is *not* generally a single-value function because of triplications. An alternative representation, which does not suffer the problems associated with triplication, is based on pairs  $(\tau, p)$ , where the quantity  $\tau$  is called *delay time* and can be calculated easily taking the intercept of a tangent to the travel time curve:

$$\tau(p) = T(X) - pX(p) \quad (9.69)$$

In fact, any point belonging to a travel time curve can be uniquely identified by an intercept along the vertical axis and a straight line with appropriate slope  $p$  as illustrated in Fig. 9.15. Although less intuitive than the previous one, this representation allows an elegant solution to the inversion of seismic data. In the case of a laterally homogeneous material, we can substitute the



**Fig. 9.15** A point  $(X_0, T_0)$  on the travel time curve can be uniquely identified by an intercept  $\tau$  and the slope  $p$  of the tangent to the curve at the point

$$\frac{d^2\tau}{dp^2} = -\frac{dX}{dp} \quad (9.73)$$

Therefore, the curve  $\tau = \tau(p)$  is concave upward in the case of prograde segments and concave downward in the case of retrograde branches. Differently from the curve  $T = T(X)$ , this curve is *always* a single-value function, thereby it can be used more easily to determine velocity models from travel time data.

solutions (9.53) and (9.55) into (9.69) to obtain an expression for the delay time  $\tau$ :

$$\begin{aligned} \tau(p) &= 2 \int_0^{z_{\max}} \left[ \frac{s^2(z)}{\sqrt{s^2(z) - p^2}} - \frac{p^2}{\sqrt{s^2(z) - p^2}} \right] dz \\ &= 2 \int_0^{z_{\max}} \sqrt{s^2(z) - p^2} dz = 2 \int_0^{z_{\max}} s_z(z) dz \end{aligned} \quad (9.70)$$

where we have used the expression (9.50) for the vertical slowness. In the case of a material formed by a stack of homogeneous layers, this expression assumes the form:

$$\tau(p) = 2 \sum_i \sqrt{s_i^2 - p^2} \Delta z_i = 2 \sum_i \eta_i \Delta z_i; s_i > p \quad (9.71)$$

The slope of the curve  $\tau = \tau(p)$  is given by:

$$\begin{aligned} \frac{d\tau}{dp} &= 2 \frac{d}{dp} \int_0^{z_{\max}} \sqrt{s^2(z) - p^2} dz \\ &= -2p \int_0^{z_{\max}} \frac{1}{\sqrt{s^2(z) - p^2}} dz = -X(p) \end{aligned} \quad (9.72)$$

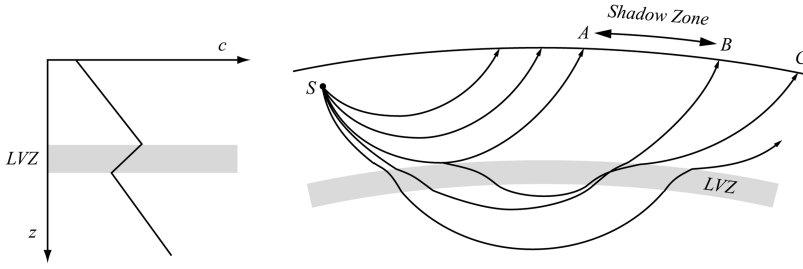
Therefore,  $X(p) \geq 0$  implies  $d\tau/dp < 0$ , and the curve  $\tau = \tau(p)$  is monotonically decreasing also in the case of triplications. Now let us take the second derivative of (9.70). We have:

## 9.6 Low-Velocity Zones

*Low-velocity zones (LVZ)* are regions where the seismic velocity decreases with depth, in contrast to a general trend characterized by positive gradients (hence by increasing velocities). According to Snell's law, we have that seismic rays are bent downwards within an LVZ, as illustrated in Fig. 9.16. Clearly, no seismic ray can bottom in an LVZ, and those that penetrate one of these regions will turn at greater depth, where the velocity  $c$  has raised above any velocity in the LVZ. Note that the seismic ray emerging at point C in Fig. 9.16 has the same parameter of the ray emerging at A. In fact, the ray with turning point at the upper boundary of the LVZ will bifurcate at its bottom, generating an additional ray that travels through the LVZ. Seismic rays with a slightly less parameter will have a *decreasing* range, back to a caustic point at the far end of the *shadow zone*.

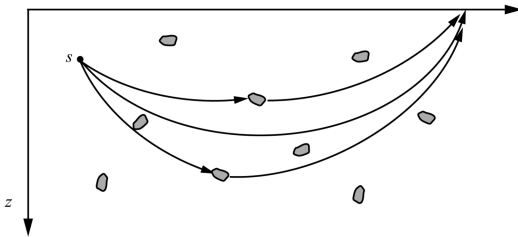
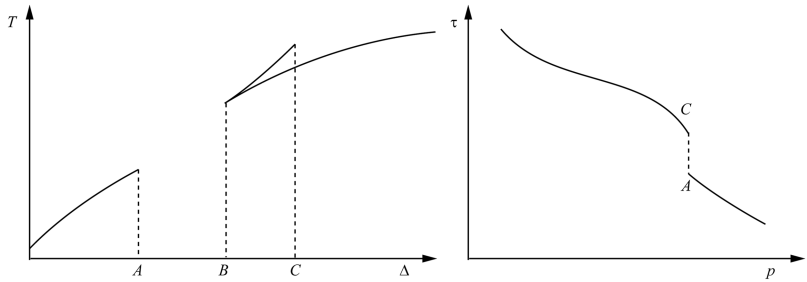
In presence of an LVZ, the travel-time curve,  $T = T(x)$ , and the delay time curve,  $\tau = \tau(p)$ , show a gap respectively for a range of distances corresponding to the shadow zone at the Earth's surface and for a range of values of the ray parameter  $p$  (Fig. 9.17). At global scale, the most important example of LVZ is a region of asthenosphere or lower lithosphere between  $100 \pm 20$  km and  $\sim 220$  km depth, which is characterized by an average 1.08 % drop in  $\alpha$  and 1.13 % drop in  $\beta$  in the PREM model (Dziewonski and Anderson 1981) (see Fig. 8.4).

This upper mantle LVZ has greater thickness and velocity reduction beneath the oceans, and the velocity drop can reach 5 % in some regions



**Fig. 9.16** A low-velocity zone at planetary scale (gray region). Seismic rays penetrating the LVZ are bent downwards, determining the formation of a shadow zone (A,B) and a duplication zone (B,C) at the Earth's surface

**Fig. 9.17** Travel-time curve (left) and delay time curve (right) in presence of an LVZ



**Fig. 9.18** Deflection of seismic rays from a random distribution of scatterers may determine multiple arrivals at a location

(Thybo 2006). In general, *S*-waves are strongly attenuated within the LVZ, especially under oceanic regions. The approximately constant depth of the top boundary suggests that the upper mantle LVZ arises from metamorphic or other pressure-driven transformations, because none of the known phase transformations of peridotite occurs at the lithostatic pressure of 100 km depth ( $\sim 3.1$  GPa). Another feature of this region is the presence *seismic wave scatterers* on a scale of  $\sim 7 \times 3$  km (Thybo 2006). *Scattering* of seismic waves consists into the deviation of seismic rays associated with high-frequency waves from small random heterogeneities (Fig. 9.18).

The most evident phenomenon associated with scattering from small scale heterogeneities is the

appearance of continuous wave trains with slowly decreasing amplitude, following the direct arrival of *S* waves or *P* waves. Aki (1969) named these wave trains *coda* and proposed that they were formed by composition of incoherent waves scattered by distributed heterogeneities in the lithosphere. In particular, he observed that while the main phase amplitude decreased with increasing epicentral distance, the coda amplitude was almost independent from the range. An in-depth treatment of this interesting topic is beyond the scope of this book and can be found in Sato et al. (2012) and in Shearer (2007). Here we are concerned only with the observation that seismic waves are strongly scattered by the LVZ, which requires the existence of smallscale heterogeneities within this layer. The scattering is observed as a pronounced coda behind the first arrivals in the offset interval of  $\sim 500$ – $1,400$  km (Thybo 2006), with more than 7 s duration at short offset and more than 2–3 s at far offsets.

The gap in the travel time curve associated with the upper mantle LVZ is observed between 800 and 1,000 km ( $\sim 8^\circ$  from the epicenter). Seismologists, petrologists, and the general geodynamics community have wondered about the origin of this region since the end of the 1950s. Some of these scientists argued that it originates

from the presence of small amounts of partial melt (e.g., Lambert and Wyllie 1970; Anderson and Spetzler 1970), while others proposed that its existence can be explained by intrinsic properties of peridotite close to its solidus. For example, Karato and Jung (1998) observed that experimental studies failed to prove significant effects of partial melting on the physical properties of peridotite for a range of melt fractions expected over most upper mantle conditions ( $<1\text{--}3\%$ ). Conversely, they pointed out that in presence of partial melting seismic wave velocities should *increase*. In fact, water dissolves in melts much more than in crystal lattices, thereby in presence of partial melting it would be removed from minerals such as olivine determining an increase of mechanical strength. These authors suggested that no significant partial melting occurs in the asthenospheric LVZ, so that the decrease of seismic velocities would result exclusively from the high water content of this layer. Similarly, Stixrude and Lithgow-Bertelloni (2005) built an upper mantle elastic isotropic and homogeneous model in conditions of thermodynamic equilibrium. These authors proved that an LVZ could be explained by the model even excluding the presence of melts. However, Hirschmann (2010) has recently showed that melts should exist anyway in the LVZ. Therefore, the open problem is not understanding whether or not partial melting may occur in the LVZ, but if this melting would effectively influence or determine the observed decrease of seismic velocities.

## 9.7 Seismic Phases Nomenclature

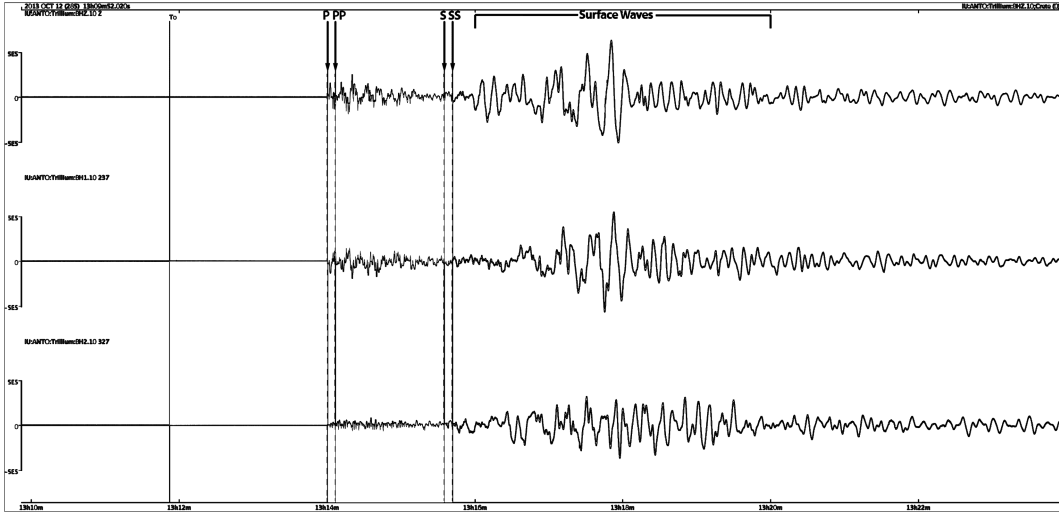
The variations of seismic velocity within the Earth and the presence of discontinuities determine the formation of several classes of ray paths after an earthquake. The corresponding arrivals at recording devices are named *seismic phases*. In general, *seismograms* result from the superposition of distinct seismic phases associated with a unique event. The possibility to observe a seismic phase at a particular station depends not only from its amplitude, but also from its polarization

and frequency spectrum. Seismic stations usually record three components of ground velocity or acceleration: a vertical component, Z, and N–S and E–W components. The horizontal recordings are then rotated to have the  $x$  axis along the radial direction to the earthquake epicenter. Therefore, if  $\gamma$  is the azimuth to the source, then the radial and transverse components of velocity are calculated by the following transformation:

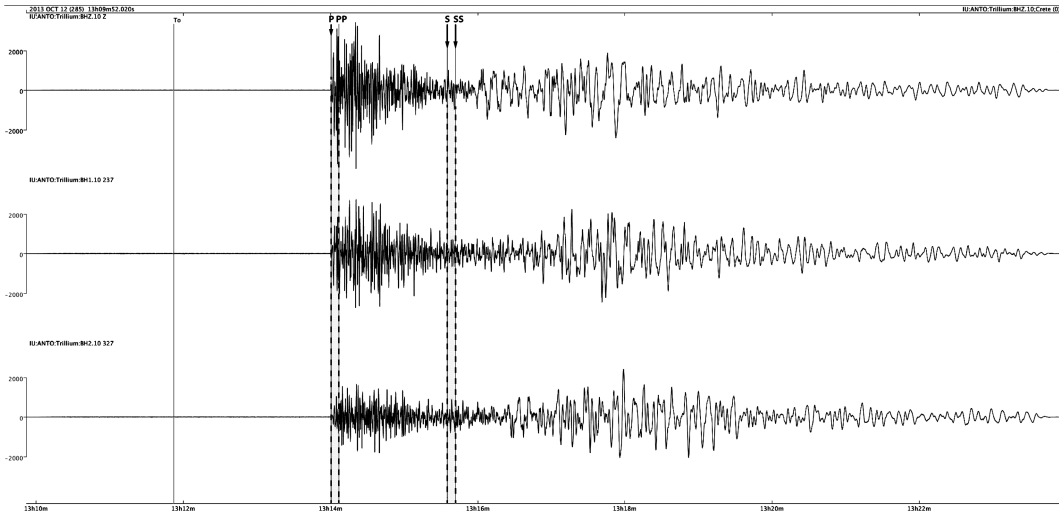
$$\begin{bmatrix} \dot{u}_R \\ \dot{u}_T \end{bmatrix} = \begin{bmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} \dot{u}_{EW} \\ \dot{u}_{NS} \end{bmatrix} \quad (9.74)$$

where  $\psi = 3\pi/2 - \gamma$ . Figure 9.19 shows an example of three-component seismogram, recorded at a station in the range of distances known as *near field* ( $\Delta < 20^\circ$ ). The three traces illustrate the main features of a seismogram in the near field, which includes  $P$  and  $S$  phases and surface wave arrivals. The identification of seismic phases on seismograms is a difficult task that can be performed either manually by a skilled seismologist or automatically by specialized computer algorithms. In general, a seismic phase is identified by a change of both amplitude and dominant frequency. For seismograms in the near field, phase identification is usually difficult in the sub-range of angular distances associated with triplication from the Moho discontinuity (see Fig. 9.13), but the correct *picking* of the arrival time of any seismic phase can be hindered anyway by background noise or by the complexity of the field of seismic velocities along the raypath. This is why the interpretation of broad-band seismograms at local or regional scale is often preceded by high-pass filtering (with typical low cut-off frequency of 1 Hz) to remove low-frequency noise, occasionally by low-pass or band-pass filtering to cut high-frequency noise.

Sometimes, filtering can be useful for identifying the precise onset of some phases, for example  $S$  waves. In this instance, it is possible to apply a filter that simulates the Wood–Anderson torsion seismometer (high-pass 2-poles Butterworth filtering with low-corner frequency of 2 Hz, followed by integration from velocity to displacement). The effect of this kind of filtering on the seismogram in Fig. 9.19 is



**Fig. 9.19** Unfiltered vertical, radial, and transverse components of ground velocity, recorded at station IRIS/ANTO (Ankara, Turkey), for an earthquake with epicenter in Crete,  $M_w$  6.7, which occurred on October 12, 2013 h13:11:52.4000. Epicentral distance is  $\Delta = 8.72^\circ$

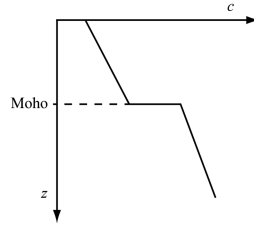


**Fig. 9.20** The same seismogram as Fig. 9.19 after application of a Wood-Anderson simulation filter.  $S$  wave arrival on the vertical trace is enhanced by this operation

illustrated in Fig. 9.20. Both these seismograms have been analyzed using *SeisGram2K*, a public-domain software tool for the visualization of seismic traces designed by Antony Lomax (<http://alomax.free.fr/software.html>). The nomenclature of seismic phases in the near field includes both *crustal phases* and *surface waves* (Fig. 9.21).

The first arrival of  $P$  waves in the near field is either a  $P_g$  or  $P_n$  phase. A  $P_g$  phase is a  $P$  wave having its turning point in the crust, while a  $P_n$  phase is a low-amplitude phase associated with a refracted wave, generated at sub-critical incidence angle and travelling almost horizontally just below the Moho. Finally, a reflection from the Moho discontinuity is called a  $PmP$

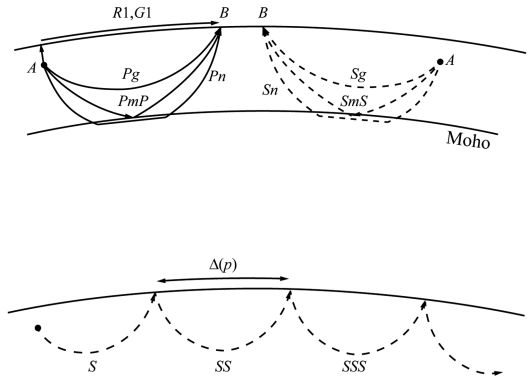
**Fig. 9.21** Near field phases from a seismic source  $A$  to a receiver  $B$ . For clarity,  $S$  wave arrivals (dashed lines) are shown separately



phase. The first arrival at a seismic station can be either a  $Pg$  phase or a  $Pn$  phase, depending on the epicentral distance  $\Delta$ , the Moho depth,  $z_m$ , and the focal depth  $z_f$ . Close to the epicenter, the first arrival is always  $Pg$ .

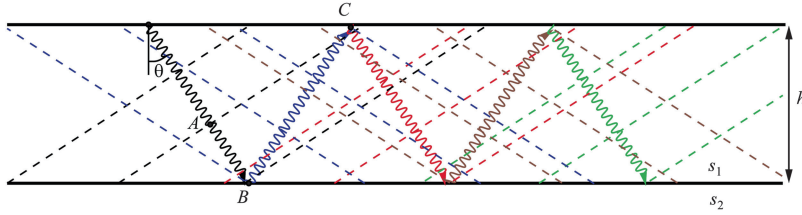
However, at a certain angular distance  $\Delta_c$  the  $Pn$  phase takes over  $Pg$ . The distance where the first arrival changes from  $Pg$  to  $Pn$  is called the *crossover point*. This location clearly depends from the crustal thickness and is  $\sim 0.27^\circ$  from the epicenter in the oceans and  $\sim 1.35^\circ$  from the epicenter on the continents. In general, for shallow earthquakes it results  $\Delta_c \approx 900z_m/\pi R$ , where  $R$  is the Earth's radius, which corresponds to a great circle distance  $X_c \approx 5z_m$  km. The reflection  $PmP$  always arrives after  $Pg$  and  $Pn$ , although its amplitude can be dominant in the coda. This phase follows very closely  $Pg$ , with delay less than 2 s beyond the crossover point. As shown in Fig. 9.21, an equivalent nomenclature exists for the  $S$  wave arrivals. Furthermore, it is possible to have *converted phases*, such as  $PmS$ ,  $SmP$ , or  $PS$ , which arise from the conversion of a  $P$  wave into an  $S$  wave or vice versa after a reflection at the Moho discontinuity or at the Earth's surface.

The last high-amplitude arrival is represented by *surface waves* travelling near the Earth's surface. In the case of spherical body waves, we know that the energy density decays as  $1/r^2$ , because it depends from the squared wave amplitude (Eq. 8.59), which decreases as  $1/r$  (Eq. 8.40). Conversely, geometrical spreading of surface waves determines a two-dimensional spread of energy, thereby energy in this instance decays as  $1/r$  and not as  $1/r^2$ . Consequently, at large distances from the source, surface waves dominate the seismograms. These phases, which are easily recognized in the case of shallow earthquakes, are of two kinds. *Rayleigh waves* are radially polarized phases resulting from



**Fig. 9.22** Multiple reflections of  $S$  waves at the Earth's surface

superposition of  $P$  and  $SV$  waves, with period less than 3 s, group velocity 3.0 km/s, which are absent if the focal depth exceeds 3 km. They determine retrograde elliptical trajectories of ground in the radial vertical plane. It can be shown that at the top of a homogeneous Poisson solid their velocity is  $0.92\beta$ , slightly less than the  $S$  wave velocity (e.g., Stein and Wysession 2003; Shearer 2009). *Love waves* are transversely polarized surface waves that form by constructive interference of high-order  $SH$  multiples, that is  $SH$  wave reflections at the Earth's surface (Fig. 9.22). The multiples are usually indicated as  $SS$ ,  $SSS$ ,  $SSSS$ ,  $SSSSS$ , etc., as illustrated in Fig. 9.22. Love waves, just as Rayleigh waves, do not form an independent class of seismic waves, because they represent an interference phenomena of normal  $S$  waves. Therefore, in principle it is possible to build them by superposition of body waves. A simplified model of formation for Love waves considers the propagation of monochromatic plane waves through a *homogeneous* layer overlying a homogeneous half-space with different mechanical properties (Fig. 9.23). It is assumed that the interference is associated with the superposition of multiple reflections trapped between the discontinuity plane and the Earth's surface. In fact, if the incidence angle of the  $SH$  waves exceeds the critical angle  $\theta_c = \arcsin(\beta_1/\beta_2)$  (Eq. 9.47), then the waves are totally reflected both at the interface and at the Earth's surface. Therefore, they are trapped in the upper layer.



**Fig. 9.23** Multiple reflections of  $S$  waves in a flat Earth model. The upper layer is homogeneous with slowness  $s_1$ . The underlying half-space has slowness  $s_2 < s_1$ . Wavefronts are shown as *dashed lines*

Let us consider the three points  $A$ ,  $B$ , and  $C$  in Fig. 9.23. The downgoing wavefronts at  $A$  and  $C$  are in phase and are subject to constructive interference only if the phase of the  $SH$  wave changes by  $2n\pi$  (for some integer  $n$ ) after the two reflections. Such phase variation includes two contributions, one associated with the reflections,  $\delta\vartheta_R$ , and one due to propagation,  $\delta\vartheta_P$ :

$$\delta\vartheta = \delta\vartheta_R + \delta\vartheta_P \quad (9.75)$$

It is possible to prove (e.g., Stein and Wysession 2003) that reflection at postcritical incidence angle determines a phase shift given by:

$$\delta\vartheta_R = 2 \arctan \left[ \frac{\mu_2 \left( \frac{p^2 - s_2^2}{s_1^2 - p^2} \right)^{1/2}}{\mu_1} \right] \quad (9.76)$$

where  $\mu_1$  and  $\mu_2$  are the rigidity moduli of the upper layer and the half-space, respectively. Regarding  $\delta\vartheta_P$ , it is determined by the path length, which is  $\overline{AB} + \overline{BC}$ , and by the wavenumber  $k$ :

$$\begin{aligned} \vartheta_P &= -(\overline{AB} + \overline{BC})k \\ &= -\left( \overline{BC} \cos 2\vartheta + \frac{h}{\cos \vartheta} \right)k \\ &= -\frac{hk}{\cos \vartheta} (1 + \cos 2\vartheta) = -2hk \cos \vartheta \end{aligned} \quad (9.77)$$

Therefore, the condition of constructive interference assumes the form:

$$\arctan \left[ \frac{\mu_2 \left( \frac{p^2 - s_2^2}{s_1^2 - p^2} \right)^{1/2}}{\mu_1} \right] - 2hk \cos \vartheta = 2n\pi \quad (9.78)$$

**Table 9.1** Main symbols for building mantle and core phases in the IASPEI standard

Phase symbol	Description
$P$	$P$ wave travelling through the mantle
$K$	$P$ wave travelling through the outer core
$I$	$P$ wave travelling through the inner core
$S$	$S$ wave travelling through the mantle
$J$	$S$ wave travelling through the inner core
$c$	Reflection at the CMB
$i$	Reflection at the ICB

or, equivalently,

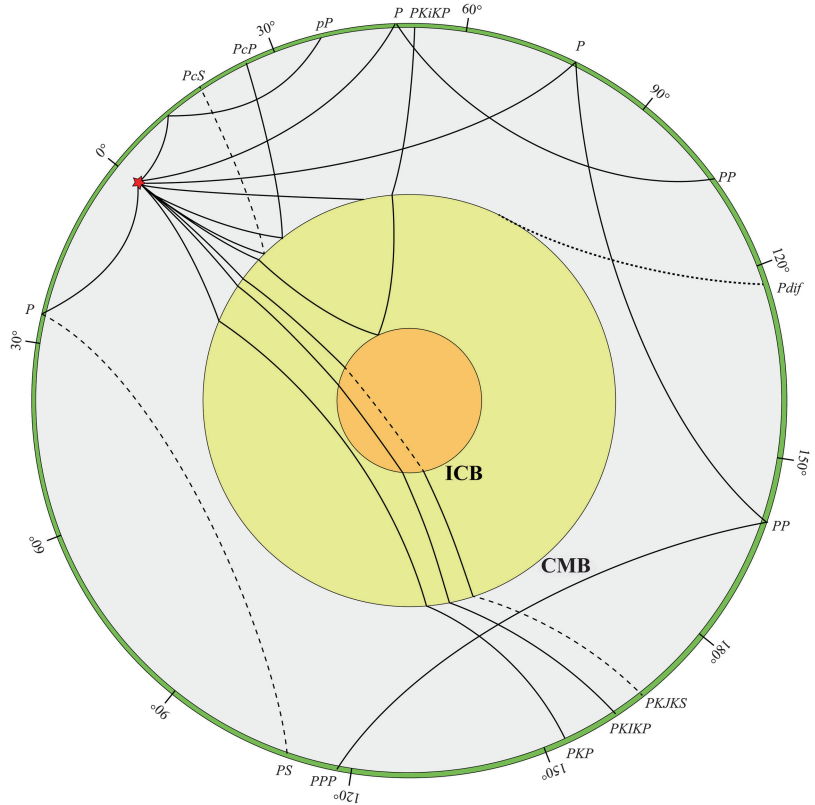
$$\frac{\mu_2}{\mu_1} \left( \frac{p^2 - s_2^2}{s_1^2 - p^2} \right)^{1/2} = \tan(hk \cos \vartheta) \quad (9.79)$$

Phases associated with Rayleigh and Love waves are indicated respectively by the symbols  $LR$  and  $LQ$ . At *teleseismic distance* ( $\Delta > 20^\circ$ ) it is possible to observe either the *mantle phases* or the *whole Earth phases*. Table 9.1 lists the basic symbols used to build mantle and core phase names, according to the *IASPEI Standard* (Storchak et al. 2003).

Phase names are built considering reflections, refractions, and conversions along raypaths. For each transformation, a new symbol is added to the phase name. For example, a raypath starting in the mantle, refracted at the CMB and bottoming in the external core will be indicated as  $PKP$  (Fig. 9.24). Similarly, the phase  $PKJKP$  indicates a  $P$  wave that penetrates the outer core, converts to an  $S$  wave at the ICB, and has a turning point in the inner core.

In the case of deep focus earthquakes, up-going rays reflected at the Earth's surface are termed *depth phases* and are indicated by  $pP$ ,

**Fig. 9.24** Main mantle and whole Earth ray paths from a deep earthquake (red star) and phase names.  $P$  and  $S$  waves are shown respectively as *solid* and *dashed* lines. Diffractions are indicated as *dotted* lines. The different colors indicate inner core (orange), outer core (yellow), mantle (grey), and lithosphere (green). Nomenclature for phases starting as  $S$  waves is similar

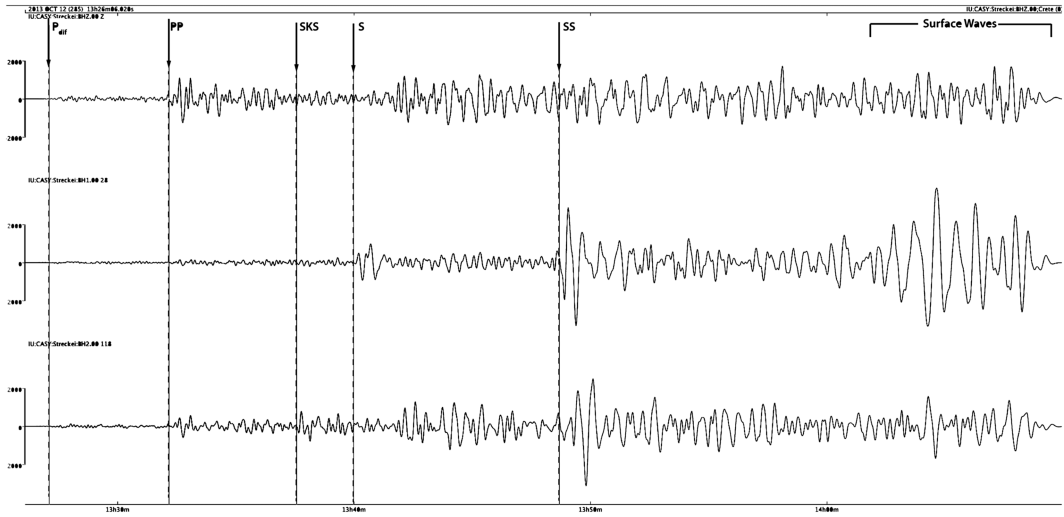


$sS$ ,  $sP$ , etc. (Fig. 9.24). The identification of phases on teleseismic seismograms requires some additional caution. First, it is often necessary to apply band-pass or low-pass filtering (with high cut-off frequency between 0.07 and 0.1 Hz) to allow a correct picking of the  $S$  phase arrival through detection of the change of frequency that is usually associated with  $S$  waves. Furthermore, at teleseismic distance the first arrival is a  $P$  wave only for  $\Delta < 110^\circ$ . In the shadow zone determined by the drop of velocity in the external core, which is the range  $110^\circ < \Delta < 145^\circ$ , the earliest arrival is a diffraction of  $P$  along the CMB, which is termed  $Pdif$  (Fig. 9.24). Beyond this range, for  $\Delta > 145^\circ$ , the first arrival is  $PKP$ . Figure 9.25 shows a teleseismic record for the same event of Fig. 9.19.

Picking of seismic phases at a number of stations for several tens of years has allowed the construction of detailed travel time curves since the 1940s and has furnished the basis for building a reliable picture of the Earth's internal

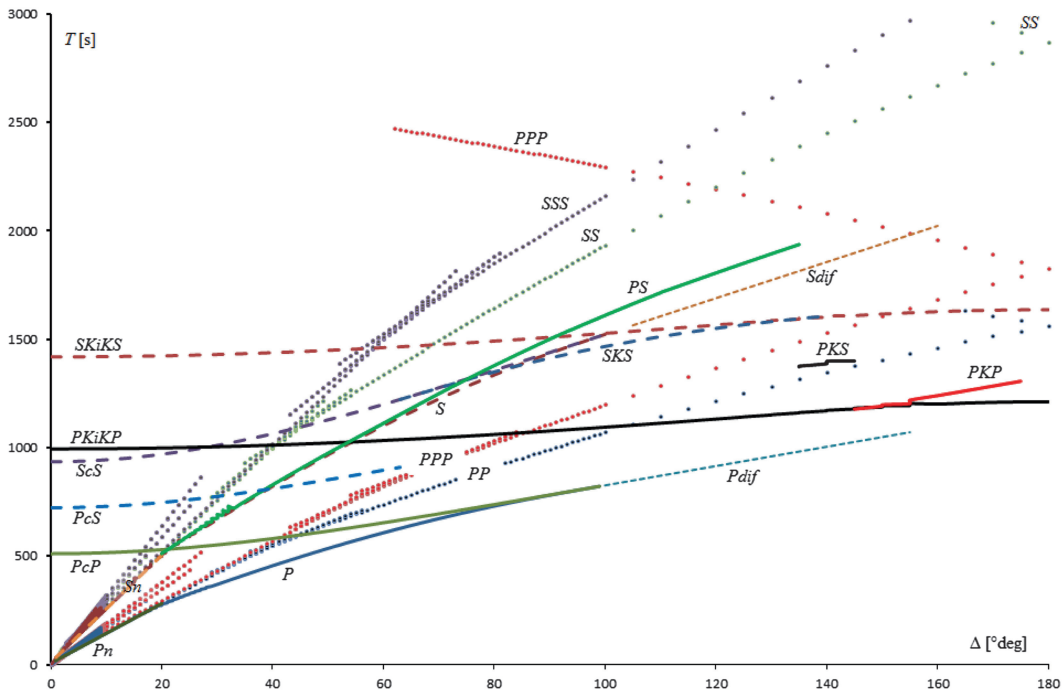
structure (e.g., the PREM model, Dziewonski and Anderson 1981). The first travel time tables were published by Jeffreys and Bullen (1940), when modern computers had not yet been invented. It is striking that the much more recent IASP91 model (Kennett and Engdahl 1991) presents only small differences with respect to the original JB tables. The most recent compilation of empirical travel times is an improvement of IASP91, which has been used to build the radial velocity model named AK135 (Kennett et al. 1995). Figure 9.26 shows travel time curves for the major phases, relative to an earthquake with focus at the Earth's surface.

Although these curves can be used to predict the approximate arrival times on a seismogram, it is important to note that lateral inhomogeneities along the raypaths, as well as local departures of the seismic velocities from the global averages, will determine a shift of the observed arrival time with respect to the theoretical value. In the future, it is likely that only minor adjustments will be



**Fig. 9.25** Teleseismic seismograms (vertical, radial, and transverse components) recorded at station IRIS/CASY (Casey, Antarctica), for the earthquake with epicenter in

Crete,  $M_w$  6.7, October 12, 2013 h13:11:52.4000. Epicentral distance is  $\Delta = 121.06^\circ$ . The original data have been band-pass filtered to the range 0.01–0.07 Hz



**Fig. 9.26** Travel time curves for the main seismic phases, determined on the basis of the AK135 model (Kennett et al. 1995)

made to the empirical global travel time curves, because they already result from the combination of a large number (almost 1,700,000) of picks (Kennett et al. 1995). Therefore, the focus will

be on the *deviations* of the observed arrival times with respect to the predicted values at regional scale, because these deviations are strictly related to the presence of lateral inhomogeneities,

especially in the upper mantle. The technique of determining the lateral variations of seismic velocity starting from the observation of arrival time differences is known as *seismic tomography* and represents a powerful tool of modern seismology. This technique has shown to be effective in providing spectacular images of the Earth's internal structure, in particular of descending slabs.

## Problems

1. What is the geometry of a seismic ray with  $\theta_0 = 0^\circ$ ;
2. Determine the travel-time curve  $T = T(X)$  for waves reflected at a horizontal discontinuity, knowing that the two-way travel time (TTWT) for vertical incidence is  $T_0$ ;
3. What is the correction that must be applied on a set of travel times of reflected arrivals to obtain proportionality between arrival time and depth to the reflector (this is known as *normal moveout*)?
4. Determine the travel-time curve  $T = T(X)$  of a  $Pn$  phase, given the crustal thickness and assuming a homogeneous crust and surficial source;
5. Assuming a homogeneous crustal layer, the first arrival at short range would be a direct wave only below the crossover distance. Determine this distance;
6. Determine the critical distance below which  $Pn$  phase arrivals are impossible, assuming a homogeneous crust;
7. Draw the delay time curve associated with the travel time curve of Fig. 9.13;
8. Download the MS Excel file ESP215.txt, containing a velocity model of oceanic crust in the Balearic Basin (Pascal et al., 1993, *Geophys. J. Int.*, **113**, 701–726.), from the supplemental material web site. This file contains a table that divides the oceanic crust into a series of flat layers with  $P$  wave velocity that increases linearly and variable thickness. Write a computer program that uses this ASCII table to build the local travel time table for  $Pg$  waves generated from a surficial source, assuming take-off angles from  $85^\circ$  to the minimum take-off angle for  $Pg$  waves, with increments of  $5^\circ$ , starting from (9.60) and (9.61).

## References

- Aki K (1969) Analysis of seismic coda of local earthquakes as scattered waves. *J Geophys Res* 74:615–631. doi:10.1029/JB074i002p00615
- Anderson DL, Spetzler H (1970) Partial melting and the low-velocity zone. *Phys Earth Planet Int* 4(1):62–64
- Dziewonski AM, Anderson DL (1981) Preliminary reference Earth model. *Phys Earth Planet Inter* 25: 297–356
- Hirschmann MM (2010) Partial melt in the oceanic low velocity zone. *Phys Earth Planet Int* 179(1):60–71
- Jeffreys H, Bullen KE (1940) *Seismological tables*, British Association for the Advancement of Science. Burlington House, London
- Karato SI, Jung H (1998) Water, partial melting and the origin of the seismic low velocity and high attenuation zone in the upper mantle. *Earth Planet Sci Lett* 157 (3–4):193–207
- Kennett BLN, Engdahl ER (1991) Travel times for global earthquake location and phase identification. *Geophys J Int* 105:429–466
- Kennett BLN, Engdahl ER, Buland R (1995) Constraints on seismic velocities in the Earth from traveltimes. *Geophys J Int* 122:108–124
- Lambert IB, Wyllie PJ (1970) Low-velocity zone of the Earth's mantle: incipient melting caused by water. *Science* 169(3947):764–766
- Sato H, Fehler MC, Maeda T (2012) *Seismic wave propagation and scattering in the heterogeneous Earth*. Springer, Berlin, p 494
- Shearer PM (2007) Seismic scattering in the deep Earth. In: Schubert G (ed) *Treatise on geophysics*, vol 1. Elsevier, The Netherlands, pp 695–730
- Shearer PM (2009) *Introduction to seismology*, 2nd edn. Cambridge University Press, Cambridge, UK, 396 pp
- Stein S, Wysession M (2003) *An Introduction to seismology, earthquakes, and earth structure*. Blackwell Publishing Ltd, Carlton, 498 pp
- Stixrude L, Lithgow-Bertelloni C (2005) Mineralogy and elasticity of the oceanic upper mantle: origin of the low-velocity zone. *J Geophys Res* 110(B3):B03204. doi:10.1029/2004JB002965
- Storchak DA, Schweitzer J, Bormann P (2003) The IASPEI standard seismic phase list. *Seismol Res Lett* 74(6):761–772
- Thybo H (2006) The heterogeneous upper mantle low velocity zone. *Tectonophysics* 416(1):53–79

## Abstract

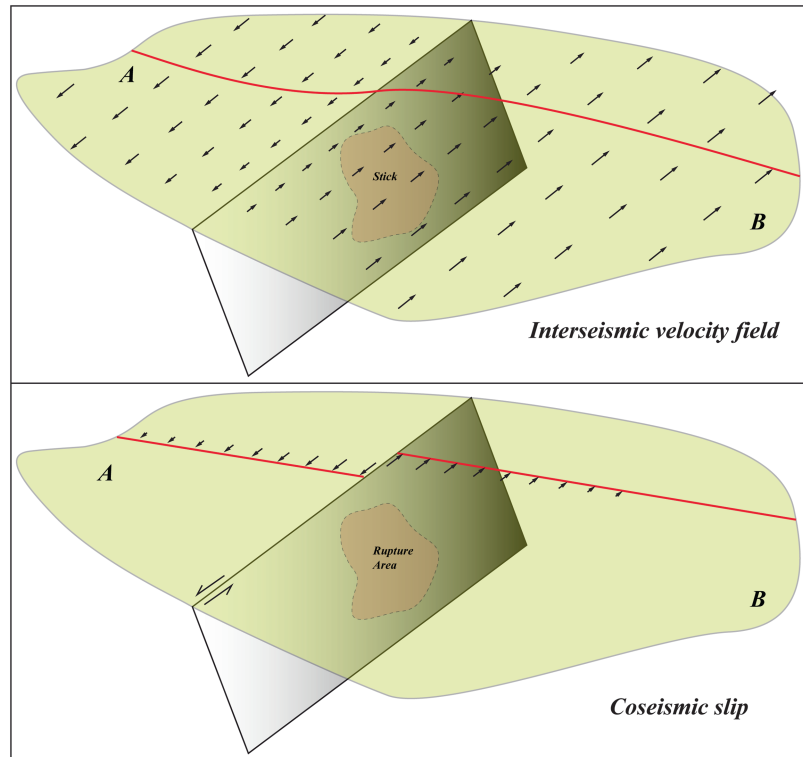
This chapter introduces earthquake seismology, in particular friction models for earthquake genesis, source mechanisms, moment tensors, and the Gutenberg-Richter law. The objective is to acquire confidence with the focal mechanisms representation (beach balls) and the recurrence time of earthquakes.

## 10.1 Reid's Model

Earthquakes represent the most evident effect of short-term plate tectonics, because of their dramatic and destructive impact on humanity. Although they are mostly concentrated along plate boundaries, strong earthquakes associated with intra-plate deformation or other phenomena are also possible and in some cases may represent an even greater hazard for human life. An earthquake can be viewed as an instantaneous release of accumulated strain energy (hence stress) in a small region along an active fault. The idea that an earthquake results from a sudden release of accumulated stress in the crust was first proposed by Reid (1910), an American engineer who studied the ground displacement field across the San Andreas Fault after the 1906 earthquake in San Francisco. His analysis led to the *elastic rebound theory* for the origin of earthquakes. In this model, slip along a fault plane may be locked for centuries, whereas the field of relative velocities is nonzero and increases progressively as we move away from a stick zone

along the fault plane. A direct consequence of the model is that the relative velocity vectors will attain the magnitude determined by global plate kinematics only at some distance from plate boundaries (Fig. 10.1). During this *interseismic phase*, which may be several thousands of years long after large earthquakes, the upper crust deforms elastically and accumulates strain energy and stress. When the accumulated stress exceeds some threshold, a “rupture” along the stick region determines sudden sliding and release of strain energy (Fig. 10.1). The (almost) instantaneous displacement during this *coseismic phase* can reach tens of meters and cause devastating earthquakes. The rupture and the subsequent evolution of the stress and strain fields leading to the next earthquake is known as the *seismic cycle*. A complete seismic cycle typically also includes a *postseismic phase*, which generally spans a few months after an earthquake but may continue for some years. In this instance, a fault that has experienced a significant rupture continues to accommodate some slip after the earthquake. Today observations of surface deformation using geodetic techniques are used to monitor the slow

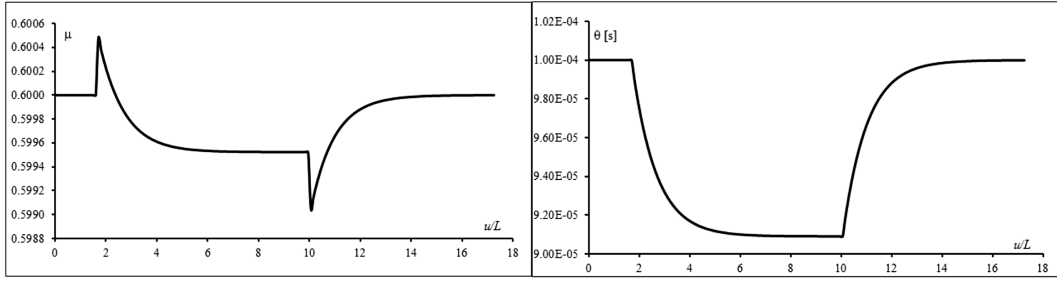
**Fig. 10.1** The seismic cycle in the elastic rebound model. The *red line* represents a reference marker on two tectonic plates, A and B, which is progressively deformed during the time interval between two earthquakes (*top, inter-seismic phase*). There is no slip along the fault plane during the interseismic phase. During an earthquake (*co-seismic phase*), the displacement field is maximum along the fault (*bottom*), so that the shape of the marker is restored on each plate



changes of strain field in seismically active regions (known as the *secular strain rate*).

A simple physical mechanism explaining the seismic cycle and the elastic rebound theory is known as the *stick-slip model* of frictional instability. This theory is based upon the observation that earthquakes do not form as a consequence of shear cracking (i.e., fracturing) of rocks, but they are ultimately frictional phenomena (Brace and Byerlee 1966). Therefore, the seismic cycle is viewed as a combination between a “stick” interseismic phase of elastic strain energy accumulation and a coseismic “slip” along an *existing* fault plane. The observation of Brace and Byerlee (1966) was followed by a number of laboratory friction experiments, with the objective to study the dynamics of sliding instability and determine a constitutive law of friction. These experiments showed that for a Saint-Venant body (see Sect. 7.4) the static friction coefficient  $\mu_s$  in Eq. (7.76) is not constant but depends on the duration of the stick interval, so that if the two surfaces are kept in static contact under load for a time interval  $\Delta t$ ,

then  $\mu_s$  increases as  $\log \Delta t$  (Dieterich 1972). The quantity  $\mu_s$  is an important parameter for understanding earthquake mechanics, because as we saw in Chap. 7 it represents the threshold ratio of shear to normal stress triggering sliding along a fault plane. During sliding, the friction coefficient decreases to a new value,  $\mu$ , which is termed the *dynamic friction coefficient*. Another significant result of the experiments was the determination of the dependence of  $\mu$  from the sliding velocity  $v$ , which was found to be:  $\mu \propto \log v$ . Finally, it was found that variations of sliding velocity determined state transitions over a characteristic distance  $L$  (for a review of the major experimental results, see Scholz 1998). At the same time, theoretical modelling of stick-slip motion revealed that the instability of frictional slip depends from a reduction of the friction force during sliding. This phenomenon was called *slip weakening*. Modelling efforts led to the formulation of an empirical constitutive law for the dynamic friction coefficient, which is known as the *rate-and-state friction law* (or *Dieterich–Ruina law*). According to this law, starting from a *steady*



**Fig. 10.2** Frictional response (*left*) and state evolution (*right*) for a positive 10 % velocity step, followed by a negative step.  $u/L$  is the normalized displacement. The plots were built assuming  $L = 10^{-5}$  m,  $a = 0.005$ , and  $b = 2a$

state with constant velocity  $v_0$  and friction coefficient  $\mu_0$ , an arbitrary velocity transition  $v_0 \rightarrow v$  will trigger a transient phase during which the friction coefficient changes continuously as a consequence of the evolution of a state variable  $\theta$ :

$$\mu(t) = \frac{\tau}{\sigma} = \mu_0 + a \ln \left[ \frac{v}{v_0} \right] + b \ln \left[ \frac{v_0 \theta(t)}{L} \right] \quad (10.1)$$

where  $a$ , and  $b$  are constants that can be determined experimentally and  $\theta(0) \equiv \theta_0 = L/v_0$ . The state variable  $\theta$  was interpreted as the age of a population of contact points supporting the load  $\sigma$  across the fault plane. The Dieterich-Ruina formula is based on the assumption that the physical state of the contact surface can be characterized at any time by a single variable  $\theta = \theta(t)$ , and that the frictional stress depends only from the normal stress  $\sigma$ , the slip rate  $v$ , and the state variable  $\theta$  (Dieterich 1979; Ruina 1983). Several evolution laws were proposed for the variable  $\theta$  (for a review, see Nakatani 2001). The most simple of them is (Dieterich and Linker 1992):

$$\dot{\theta}(t) = 1 - \frac{1}{L} \theta(t) v \quad (10.2)$$

Experiments showed that the characteristic distance  $L$  varies between 2 and 100  $\mu\text{m}$  and increases with the surface roughness and the fault gouge particle size. Equation 10.2 implies that for a stationary contact the state variable  $\theta$  increases linearly with time, while for constant  $v > 0$  we have:

$$\theta(t) = \frac{L}{v} \left[ 1 + \left( \frac{v}{v_0} - 1 \right) e^{-v t / L} \right] \quad (10.3)$$

The constitutive law (10.1) shows that the friction coefficient  $\mu$  may change either as a consequence of velocity variations or as a consequence of state transitions. Figure 10.2 illustrates the variations of dynamic friction coefficient and state after positive and negative velocity steps. After a sudden velocity increase  $v \rightarrow \alpha v$ ,  $\alpha > 1$ ,  $\mu$  has a positive transition  $\mu \rightarrow \mu + a \ln \alpha$ , which is known as the *direct velocity effect*. Such a discontinuous transition is followed by a continuous decrease in friction, having magnitude  $b \ln \alpha$ . In fact, by (10.2) we have that the state variable  $\theta$  decreases exponentially to the asymptotic value  $\theta_\infty = L/v$ , thereby the third factor at the right-hand side of (10.1) will tend asymptotically to the value  $-b \ln(v/v_0)$  (Fig. 10.2).

As a consequence, the steady state friction coefficient will be given by:

$$\mu_\infty = \frac{\tau_\infty}{\sigma} = \mu_0 + (a - b) \ln \frac{v}{v_0} \quad (10.4)$$

This solution implies that for  $a < b$  the steady state friction decreases with increasing velocity. This form of friction can be observed for a wide class of materials, and the reference friction coefficient  $\mu_0$  results to be nearly independent from the rock type and from temperature. The solution (10.4) apparently says that the steady state friction  $\mu$  at some velocity  $v$  depends from the *previous* steady state pair  $(\mu_0, v_0)$ . In this instance, a series of velocity values  $v_0, v_1, \dots, v_n$  would produce a sequence of friction coefficients

$\mu_0, \mu_1, \dots, \mu_n$ , such that the steady state coefficient for the stage  $k + 1$  would be given by:

$$\begin{aligned} \mu_{k+1} &= \mu_k + (a - b) \ln \frac{v_{k+1}}{v_k}; k \\ &= 0, 1, \dots, n - 1 \end{aligned} \quad (10.5)$$

However, this recurrence formula can be easily solved in terms of  $(\mu_0, v_0)$ , so that the steady state coefficient for the  $k$ -th stage can be calculated easily from  $(\mu_0, v_0)$ . It results:

$$\mu_k = \mu_0 + (a - b) \ln \frac{v_k}{v_0}; k = 0, 1, \dots, n \quad (10.6)$$

Therefore, in the calculation of the dynamic friction coefficient through the rate-and-state law (10.1),  $\mu_0$  and  $v_0$  can be considered as reference values, so that the resulting steady state value of  $\mu$  will be independent from the previous state of the system. Now we may wonder if the existence of a static friction coefficient  $\mu_s$  can be explained through the rate-and-state law (10.1). An interesting numerical experiment, which can also be performed as real laboratory experiment (see Nakatani 2001), will help clarifying the concept of static friction.

In a *slide-hold-slide* experiment, a sliding mass  $M$  is initially moved applying a constant shear stress  $\tau_0$ . In steady state conditions, this stress is balanced by a frictional stress  $\mu_0\sigma$ , so that  $\tau_0 = \mu_0\sigma$  and the mass moves at constant velocity  $v(t) = v_0$ . Now let us assume that at some time  $t = t_0$  the applied stress is reduced to a lower value  $\tau < \tau_0$  (possibly zero). At this point, the velocity will start decreasing according to the following equation of motion:

$$\dot{v}(t) = \frac{1}{M} [\tau - \mu(t)\sigma] \quad (10.7)$$

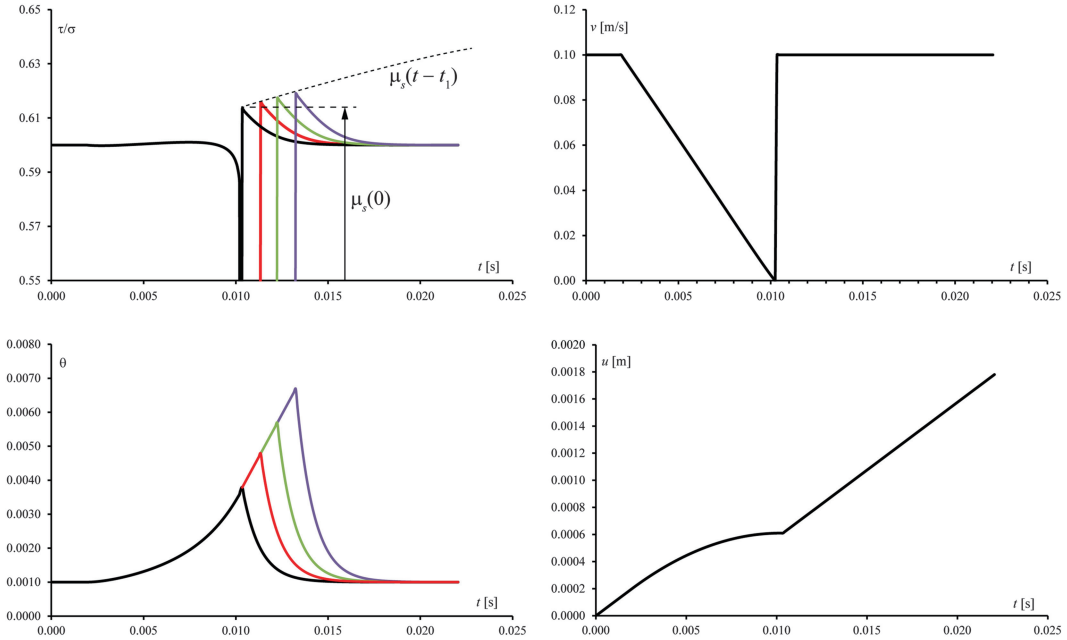
where  $M$  is mass per unit surface and  $\mu(t)$  is the dynamic friction coefficient at time  $t$ , which evolves according to the rate-and-state law (10.1) and the state evolution law (10.2). We also assume that at some later time  $t_1 = t_0$ , as soon as the velocity has decreased to zero, the slider is restarted and moved again at constant velocity  $v_0$ .

The resulting plots of frictional stress, velocity, and state are illustrated in Fig. 10.3. We note a prominent peak of the state variable  $\theta$  for  $v = 0$ , and the frictional stress peak at  $t = t_1$ , which represents the static friction stress  $\sigma\mu_s$  that must be applied to resume sliding. In this numerical experiment, the duration of the stick interval is initially zero, so that the initial value of static friction can be defined as the *dynamic* friction coefficient at the onset of slipping:

$$\mu_s(0) = \mu_0 + b \ln \left[ \frac{v_0 \theta(t_1)}{L} \right] \quad (10.8)$$

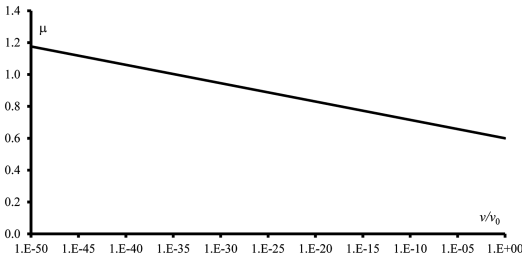
Alternatively, if we keep the slider at rest for some finite time interval  $t_{stick} = t - t_1$ , we will observe a *logarithmic* increase of the static friction  $\mu_s(t_{stick})$  as a consequence of the linear growth of the state variable  $\theta$  for  $v = 0$  (Eq. 10.2). Such an increase of the friction coefficient during stationary contact is known as *healing* (e.g., Berthoud et al. 1999). The curves in Fig. 10.3 show that static friction is *not* a separate phenomenon but can be explained in terms of rate-and-state evolution. Furthermore, they show that the initial value of  $\mu_s$  does not represent an absolute threshold determined exclusively by the physical properties of the materials. In fact, by (10.8) we see that  $\mu_s(0)$  also depends from the current value of the state variable  $\theta$  at the stop time and, apparently, from the velocity at which we resume sliding after the stick interval ( $v_0$  in this example). Clearly, this dependence from the restarting velocity is not actual, because the system does not know a priori at which velocity the slider will be moved. In reality, the appearance of a velocity dependence arises from the fact that the order of magnitude of the dynamic friction coefficient does not change for a wide range of velocities. For example, using the parameters of Fig. 10.3 and Eq. (10.4), we see that for a transition  $v_0 \rightarrow 10^{-3} v_0$ , the friction coefficient will change from  $\mu = 0.60$  to  $\mu \cong 0.63$ . Figure 10.4 shows the predicted steady state friction coefficient for a wide interval of relative velocities  $v/v_0$ .

Therefore, if we apply a small push to the slider when it is at rest, we will effectively accelerate the mass to a very small velocity  $v$ ,



**Fig. 10.3** Frictional response, state evolution, slider velocity, and displacement in a numerical slide–hold–slide experiment. The applied stress  $\tau$  is reduced by 10 % at time  $t_0 = 0.001$  s, while sliding at constant reference velocity  $v_0$  is restored as soon as  $v = v(t_1) = 0$  (black lines) or at a later time (colored lines). The assumed param-

eters are:  $L = 10^{-4}$  m,  $a = 0.005$ ,  $b = 2a$ ,  $v_0 = 0.1$  ms $^{-1}$ ,  $\mu_0 = 0.6$ ,  $M = 0.005$  kg m $^{-2}$ . The static friction coefficient  $\mu_s$  increases logarithmically with the duration of the stick interval  $t - t_1$ . During this time interval, the state variable  $\theta$  increases linearly according to (10.2) with  $v = 0$



**Fig. 10.4** Steady state dynamic friction coefficient as a function of relative velocity. Note the logarithmic scale of the horizontal axis. The assumed parameters are:  $a = 0.005$ ,  $b = 2a$ ,  $v_0 = 0.1$  ms $^{-1}$ ,  $\mu_0 = 0.6$

say  $10^{-8}$  m/s, but the resulting *dynamic* friction, which is comparable to  $\mu_0$ , will immediately stop the slider. We could not even detect the corresponding small displacement. Conversely, if the applied stress is sufficiently high to accelerate the mass to a macroscopic velocity  $v$  having the same order of magnitude of  $v_0$ , we can eventually balance the frictional stress at that velocity and

trigger the onset of sliding. As we have seen above, the stress required to put the slider in motion will increase with the stick time as a consequence of the healing phenomenon. This reflects the fact that the interface region between slider and sliding surface is effectively a multicontact interface, along which a population of contact points supporting the load  $\sigma$  increases logarithmically with the hold time.

The friction law (10.1) predicts instability of sliding under some specific conditions. Let us consider a spring-slider system pulled at constant velocity  $V_0$  (Fig. 10.5). The equation of motion for the slider simply states that the acceleration of a unit mass having unit contact area is proportional to the net force resulting from the combination of spring pull (or push) and friction:

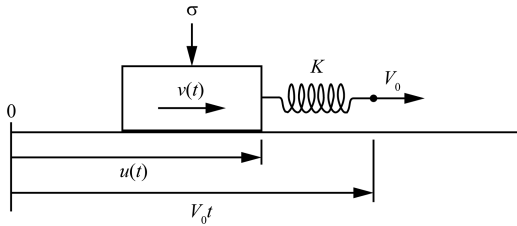
$$\ddot{u} = -K[u(t) - V_0 t] - \tau(v, \theta) \quad (10.9)$$

where  $K$  represents the spring stiffness and  $\tau$  is the frictional stress exerted on the slider.

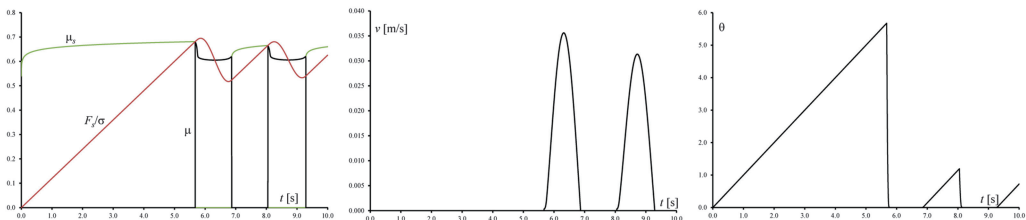
A simple numerical experiment shows that for  $a < b$  the slider motion is jerky, that is, it is characterized by phases of rapid slip separated by stick intervals, as illustrated in Fig. 10.6 Starting from a stationary contact, the spring force to normal stress ratio,  $F_s/\sigma$ , will increase linearly until the current value of the static friction coefficient  $\mu_s$  is attained. At this point, a velocity spike occurs and the spring shortly starts relaxing. When  $F_s/\sigma$  falls below the dynamic friction coefficient  $\mu$ , the slider starts decelerating and quickly stops. As shown in Fig. 10.6 during the stick time intervals, the static friction coefficient increases logarithmically, while the state variable increases linearly.

Also the spring force increases linearly during a stick interval. In fact, if at any time  $t = t_0$  the slider stops at location  $u_0 = u(t_0)$ , then during the subsequent stick interval we have:

$$\begin{aligned} F_s(t) &= -K(u_0 - V_0 t) \\ &= F_s(t_0) + K V_0 (t - t_0); t \geq t_0; \text{ stick time} \end{aligned} \quad (10.10)$$



**Fig. 10.5** The spring-slider analogy for earthquake nucleation. A spring with stiffness  $K$  is pulled at constant velocity  $V_0$ . The elastic force exerted by the spring on the attached slider is opposed by the friction stress  $\tau(t) = \sigma\mu(t)$ . This system may exhibit stick-slip instability



**Fig. 10.6** Frictional response, slider velocity, and state evolution in a numerical stick-slip motion experiment. A spring-slider system is pulled at velocity  $V_0 = 0.01 \text{ ms}^{-1}$ . It is assumed that the slider is at rest for  $t = 0$ .

At any time  $t$  in the stick interval, this force could *potentially* accelerate the slider to velocity:

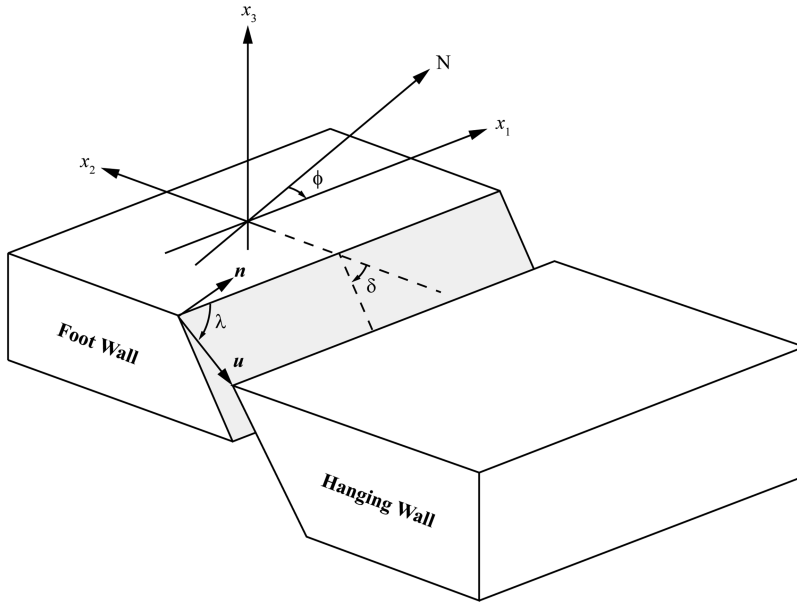
$$\begin{aligned} v(t) &= \int_{t_0}^t F_s(t') dt' \\ &= \frac{1}{2} K V_0 (t^2 - t_0^2) - K u_0 (t - t_0) \end{aligned} \quad (10.11)$$

However, the slider will effectively restart only when the dynamic friction corresponding to this potential velocity is less than the applied spring force. In conclusion, a plausible though qualitative explanation of the seismogenic mechanism can be obtained combining Reid's elastic rebound idea with the modern theory of friction and analog spring-slider modelling. However, it is important to note that this representation of earthquake nucleation cannot be used to predict the short-term occurrence of seismic events. The reason is that a fault generally includes several stick regions with different geometry and different normal loads. This complexity clearly affects the regularity of stick-slip motion. Furthermore, an earthquake associated with rupture of a stick zone usually modifies the stress field of other stick regions, determining a delay or an advance of the subsequent rupture. Therefore, earthquake prediction cannot be based on rigorous geophysical laws, but must be considered as a statistical problem.

## 10.2 Faults and Focal Mechanisms

An important step in the study of earthquake dynamics consists into the determination of the location of the rupture area along a fault plane.

The parameters are set as follows:  $K = 12 \text{ Nm}^{-1}$ ,  $\sigma = 1 \text{ Pa}$ ,  $dt = 0.0005 \text{ s}$ ,  $L = 10^{-5} \text{ m}$ ,  $a = 0.005$ ,  $b = 2a$ ,  $v_0 = 0.1 \text{ ms}^{-1}$ ,  $\mu_0 = 0.6$ .  $F_s = -K[u(t) - V_0 t]$  is the spring force



**Fig. 10.7** Kinematic parameters of a source mechanism. In earthquake studies, a fault is modeled as a planar surface with normal vector  $\mathbf{n}$ . The slip vector  $\mathbf{u}$  describes the relative displacement of the hanging wall block relative to the foot wall block. It always results  $\mathbf{u} \cdot \mathbf{n} = 0$ . The intersection of the fault plane with the Earth's surface is

a line whose orientation relative to the North is described by the *strike* angle  $\phi \in [0^\circ, 360^\circ]$ . The angle between fault plane and the Earth's surface is the *dip*  $\delta \in (0^\circ, 90^\circ]$ . The *rake* (or *slip*) angle  $\lambda \in [-180^\circ, +180^\circ]$  is between the positive  $x_1$  axis and the slip vector. It is positive when  $\mathbf{u}$  is directed upwards

This is the stick region where the “static” frictional forces were overcome by the accumulated elastic stress. Usually, this location is specified by a single representative point in the Earth's crust or mantle, which is called the earthquake *focus* or *hypocenter*. The coordinates  $(x_0, y_0, z_0)$  of this point and the corresponding source time  $t_0$  can be determined through specific inversion algorithms (e.g., Stein and Wyssession 2003) that rely on the arrival times at seismic stations. The next step is determining the *source mechanism* associated with the earthquake. This mechanism includes the orientation of the fault along which the rupture occurred and the slip vector of relative motion between the two blocks or plates.

The kinematic parameters associated with a source mechanism are the *strike*,  $\phi$ , the *dip*,  $\delta$ , and the *rake* (or *slip*),  $\lambda$ . They are illustrated in Fig. 10.7. A local reference frame can be defined aligning the  $x_1$  axis with the fault strike, in such a way the fault plane dips to the right when we look at the positive  $x_1$  direction. Then, the  $x_2$  axis is

oriented as the opposite direction with respect to the dipping trend. Therefore, in this instance the  $x_3$  axis is vertical and directed upwards. *Reverse faulting* occurs when the slip vector is directed upwards, while downward motion is referred to as *normal faulting*. They reflect the existence of compressional or extensional stress fields, respectively. Faults with  $\delta < 45^\circ$  and reverse faulting are usually termed *thrusts*. Finally, the horizontal and vertical components of motion are indicated, respectively, as *strike-slip* and *dip-slip*. In particular, in the case of pure strike-slip motion, when an observer sees the adjacent block moving rightwards the faulting mechanism is said to be *right-lateral* strike-slip, otherwise it is termed *left-lateral* strike-slip. Conventionally, the rake is set to  $\lambda = 0^\circ$  in the case of pure left-lateral strike-slip motion, whereas  $\lambda = 180^\circ$  in the case of pure right-lateral strike-slip motion.

The set of parameters  $(\phi, \delta, \lambda, \mathbf{u})$  is called the *focal mechanism* of an earthquake. It determines uniquely geometry and kinematics of the

seismogenic source. The fundamental problem of the earthquake source theory is determining the *radiation pattern* of displacements  $\mathbf{u} = \mathbf{u}(\mathbf{r}, t)$  from a source located at the origin, assuming elastic properties of the transmission medium. In this context, the building block of analog models that are representative of real seismic sources is simply a body force (per unit volume),  $\mathbf{f}$ , applied to a point  $\mathbf{r} = \mathbf{r}_0$ . In principle, the objective of determining the radiation pattern associated with this simple source can be accomplished solving the non-homogeneous version (7.50) of Cauchy's momentum equation with  $\mathbf{f}(\mathbf{r}, t) = \mathbf{g}(t)\delta(\mathbf{r}-\mathbf{r}_0)$ :

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \tau_{ij}}{\partial x_j} + g_i(t) \delta(\mathbf{r} - \mathbf{r}_0) \quad (10.12)$$

where  $\delta(\mathbf{r} - \mathbf{r}_0)$  is the Dirac delta function, which is defined by the following functional relations (e.g., Panofsky and Phillips 2005):

$$\left\{ \begin{array}{l} \delta(\mathbf{r}) = 0 \text{ for } \mathbf{r} \neq \mathbf{0} \\ \int_{\mathbf{R}} \delta(\mathbf{r} - \mathbf{r}') dx' dy' dz' = 1 \text{ for any region } \mathbf{R} \\ \text{such that } \mathbf{r} \in \mathbf{R} \\ \int_{\mathbf{R}} f(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') dx' dy' dz' = f(\mathbf{r}) \text{ for any} \\ \text{region } \mathbf{R} \text{ such that } \mathbf{r} \in \mathbf{R} \end{array} \right. \quad (10.13)$$

We also note that the second of these properties is a consequence of the third, more general, property. To determine the radiation pattern associated with  $\mathbf{f}$ , it is useful to start from the *static* displacement field generated by the application of a force  $\mathbf{f}$  at the origin of the reference frame. In equilibrium conditions, the displacement is zero far from the origin, and assuming a homogeneous medium we have:

$$0 = (\lambda + \mu) \frac{\partial \Delta}{\partial x_i} + \mu \frac{\partial^2 u_i}{\partial x_j^2} + f_i \quad (10.14)$$

where have simply rewritten (8.11) without the acceleration term and including the body force contribution. In vector notation, this equation assumes the form:

$$\mathbf{f} + (\lambda + 2\mu) \nabla (\nabla \cdot \mathbf{u}) - \mu \nabla \times \nabla \times \mathbf{u} = 0 \quad (10.15)$$

where we have used the identity:

$$\nabla \times \nabla \times \mathbf{u} = \nabla (\nabla \cdot \mathbf{u}) - \nabla^2 \mathbf{u} \quad (10.16)$$

In this problem, the body force field  $\mathbf{f} = \mathbf{f}(\mathbf{r})$  is concentrated at the origin, so that it must be expressed in terms of Dirac's delta function. By Gauss' theorem, this function can be written as the Laplacian of a scalar field:

$$\delta(\mathbf{r}) = -\frac{1}{4\pi} \nabla^2 \left( \frac{1}{r} \right) \quad (10.17)$$

Therefore following Lay and Wallace (1995) we can write:

$$\begin{aligned} \mathbf{f}(\mathbf{r}) &= \mathbf{n} \delta(\mathbf{r}) = -\nabla^2 \left( \frac{\mathbf{n}}{4\pi r} \right) \\ &= -\nabla \left( \nabla \cdot \frac{\mathbf{n}}{4\pi r} \right) + \nabla \times \nabla \times \left( \frac{\mathbf{n}}{4\pi r} \right) \end{aligned} \quad (10.18)$$

where  $\mathbf{n}$  is a unit vector representing the direction of  $\mathbf{f}$ . If we insert this force into the static equilibrium Eq. (10.15) we obtain:

$$\begin{aligned} &-\nabla \left( \nabla \cdot \frac{\mathbf{n}}{4\pi r} \right) + \nabla \times \nabla \times \left( \frac{\mathbf{n}}{4\pi r} \right) \\ &+ (\lambda + 2\mu) \nabla (\nabla \cdot \mathbf{u}) - \mu \nabla \times \nabla \times \mathbf{u} = 0 \end{aligned} \quad (10.19)$$

Therefore,

$$\begin{aligned} &\nabla \left\{ \nabla \cdot \left[ -\frac{\mathbf{n}}{4\pi r} + (\lambda + 2\mu) \mathbf{u} \right] \right\} \\ &+ \nabla \times \nabla \times \left( \frac{\mathbf{n}}{4\pi r} - \mu \mathbf{u} \right) = 0 \end{aligned} \quad (10.20)$$

We search a solution  $\mathbf{u} = \mathbf{u}(\mathbf{r})$  having the form:

$$\mathbf{u} = \nabla (\nabla \cdot \mathbf{A}_p) - \nabla \times (\nabla \times \mathbf{A}_s) \quad (10.21)$$

To this purpose, we note that for an arbitrary vector field  $\mathbf{u} = \mathbf{u}(\mathbf{r})$  we can always determine a vector field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$  such that:

$$\nabla^2 \mathbf{A}(\mathbf{r}) = \mathbf{u}(\mathbf{r}) \quad (10.22)$$

In fact, (10.22) is a classic *Poisson's equation*, which has a unique solution when  $\mathbf{u}$  decreases with sufficient rapidity (at least as  $1/r$ ) for  $\mathbf{r} \rightarrow \infty$ :

$$\mathbf{A}(\mathbf{r}) = -\frac{1}{4\pi} \int_{R^3} \frac{\mathbf{u}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' \quad (10.23)$$

Therefore, using the vector identity (10.16) for the field  $\mathbf{A}$  gives:

$$\mathbf{u} = \nabla^2 \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla \times (\nabla \times \mathbf{A}) \quad (10.24)$$

Now, setting:

$$\phi = \nabla \cdot \mathbf{A}; \quad \Psi = -\nabla \times \mathbf{A} \quad (10.25)$$

we obtain:

$$\mathbf{u} = \nabla \phi + \nabla \times \Psi \quad (10.26)$$

This equation implies that a vector field  $\mathbf{u} = \mathbf{u}(\mathbf{r})$  always has an irrotational component ( $\nabla \phi$ ) and a solenoidal component ( $\nabla \times \Psi$ ). The fields  $\phi$  and  $\Psi$  are termed respectively the *scalar potential* and the *vector potential* of the displacement field, and (10.26) is known as *Helmholtz's decomposition theorem*. Now, let us consider two vector fields,  $\mathbf{A}_p$  and  $\mathbf{A}_s$ , such that  $\mathbf{A} = \mathbf{A}_p + \mathbf{A}_s$ . In order to satisfy (10.21), we must have that:  $\nabla \times \mathbf{A}_p = \mathbf{0}$  and  $\nabla \cdot \mathbf{A}_s = 0$ . Therefore, by (10.16) we have:  $\nabla^2 \mathbf{A}_p = \nabla(\nabla \cdot \mathbf{A}_p)$  and  $\nabla^2 \mathbf{A}_s = \nabla \times (\nabla \times \mathbf{A}_s)$ . Substituting (10.21) into (10.20) and taking into account that  $\nabla^2 \mathbf{A} = \nabla^2 \mathbf{A}_p + \nabla^2 \mathbf{A}_s$  leads to:

$$\begin{aligned} \nabla \left\{ \nabla \cdot \left[ -\frac{\mathbf{n}}{4\pi r} + (\lambda + 2\mu) \nabla^2 \mathbf{A}_p \right] \right\} \\ + \nabla \times \nabla \times \left( \frac{\mathbf{n}}{4\pi r} - \mu \nabla^2 \mathbf{A}_s \right) = 0 \end{aligned} \quad (10.27)$$

This equation is clearly satisfied when:

$$\begin{cases} \nabla^2 \mathbf{A}_p = \frac{\mathbf{n}}{4\pi r(\lambda + 2\mu)} \\ \nabla^2 \mathbf{A}_s = \frac{\mathbf{n}}{4\pi r\mu} \end{cases} \quad (10.28)$$

Now let us assume that the potentials  $\mathbf{A}_p$  and  $\mathbf{A}_s$  have the same direction of  $\mathbf{n}$ :  $\mathbf{A}_p = A_p \mathbf{n}$ ,  $\mathbf{A}_s = A_s \mathbf{n}$ . In this instance, the equations (10.28) reduce to just two standard Poisson's equations:

$$\begin{cases} \nabla^2 A_p = \frac{1}{4\pi r(\lambda + 2\mu)} \\ \nabla^2 A_s = \frac{1}{4\pi r\mu} \end{cases} \quad (10.29)$$

These equations are easily integrated taking into account that  $\nabla^2 r = 2/r$ . Therefore, the solutions have the form:

$$\begin{cases} A_p = \frac{r}{8\pi(\lambda + 2\mu)} \\ A_s = \frac{r}{8\pi\mu} \end{cases} \quad (10.30)$$

To determine the displacement, we must multiply (10.30) by  $\mathbf{n}$  and insert the vector potentials  $\mathbf{A}_p$  and  $\mathbf{A}_s$  into (10.21). It is convenient to set  $\mathbf{n} = \mathbf{e}_j$  and determine the  $i$ -th component of the displacement field associated with a force at the origin in the direction  $\mathbf{e}_j$ . We shall indicate this quantity by  $u_i^j$ :

$$u_i^j = \frac{1}{8\pi\mu} \left[ \delta_{ij} (2 - \Gamma) \frac{1}{r} + \Gamma \frac{x_i x_j}{r^3} \right] \quad (10.31)$$

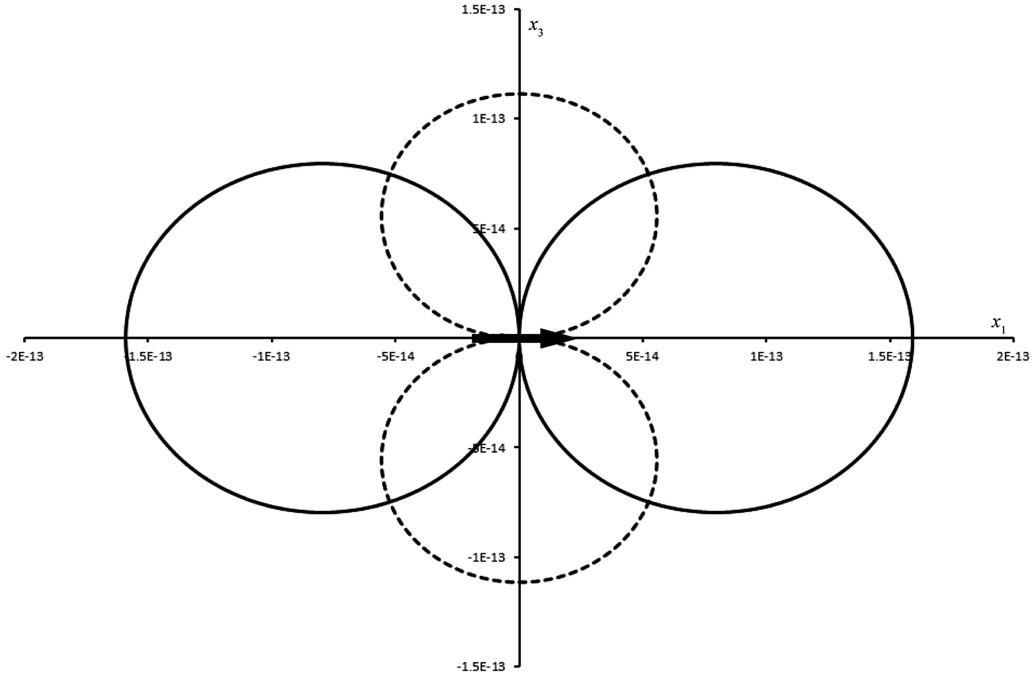
where:

$$\Gamma \equiv \frac{\lambda + \mu}{\lambda + 2\mu} \quad (10.32)$$

The quantities  $u_i^j$  form a symmetric tensor that is known as the *Somigliana tensor*. To visualize the pattern of deformation, it is usually convenient to express the components of displacement in spherical coordinates applying the following transformation (see Appendix I):

$$\begin{bmatrix} u_r^j \\ u_\theta^j \\ u_\phi^j \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\sin \phi & \cos \phi & 0 \end{bmatrix} \begin{bmatrix} u_1^j \\ u_2^j \\ u_3^j \end{bmatrix} \quad (10.33)$$

For example, for a force applied in the  $x_1$  direction, it is easy to calculate the components of the displacement field in the plane  $x_1 x_3$  ( $\phi = 0$ ). Using the transformation (2.27) to represent the



**Fig. 10.8** Radial (solid lines) and tangential (dashed lines) components of displacement after the application of a unit force at the origin in the  $x_1$  direction (black arrow). The plot has been drawn assuming  $\mu = 50$  GPa and  $\lambda = 25$  GPa

components  $x_i$  of the position vector  $\mathbf{r}$  we obtain after some steps:

$$\begin{aligned} u_r^1 &= u_1^1 \sin \theta + u_3^1 \cos \theta = \frac{1}{4\pi\mu r} \sin \theta \\ u_\theta^1 &= u_1^1 \cos \theta - u_3^1 \sin \theta = \frac{1}{4\pi\mu r} \left(1 - \frac{\Gamma}{2}\right) \cos \theta \end{aligned} \quad (10.34)$$

Therefore, the radial component of displacement is zero along the vertical axis and attains its maximum in the horizontal plane. Conversely, the tangential displacement (which is associated with shear strain) is maximum along the  $z$  axis and zero in the horizontal plane. Both components have a two-lobe pattern as illustrated in Fig. 10.8. If a force is applied at position  $\xi$  instead that at the origin, the previous formulae must be corrected by the transformation:  $x_i \rightarrow x_i - \xi_i$ . In particular, we are going to consider the case of the application of a *force couple* in the  $x_1$  direction at locations  $(\xi_1, \xi_2 + 1/28\xi_2, \xi_3)$  and  $(\xi_1, \xi_2 - 1/28\xi_2, \xi_3)$ . By the superposition principle, we have that in this instance the displacement at a location  $\mathbf{r}$

will be given by the sum of the displacements associated with the single forces.

Let us indicate the displacement at  $\mathbf{r}$  generated by the application of a force at  $\xi$  by  $\mathbf{u}(\xi, \mathbf{r})$ . Using this notation, the displacement associated with a force couple in the  $x_1$  direction and with a small arm in the  $x_2$  direction (Fig. 10.9) can be written as:

$$\begin{aligned} u_i^1(\xi, \mathbf{r}) &= u_i^1(\xi_1, \xi_2 + \delta\xi/2, \xi_3, x_1, x_2, x_3) \\ &\quad - u_i^1(\xi_1, \xi_2 - \delta\xi/2, \xi_3, x_1, x_2, x_3) \\ &= \frac{\partial u_i^1}{\partial \xi_2} \delta\xi_2 + O(\delta\xi_2^2) \end{aligned} \quad (10.35)$$

Considering that  $r^2 = (x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 + (x_3 - \xi_3)^2$ , we have:  $\partial r / \partial \xi_i = -\partial r / \partial x_i$ . Therefore,

$$\frac{\partial u_i^j}{\partial \xi_k} = -\frac{\partial u_i^j}{\partial x_k} \quad (10.36)$$

Consequently, the total displacement can be written as follows:

$$u_i^1(\xi, r) = -\frac{\partial u_i^1}{\partial x_2} \delta \xi_2 + O(\delta \xi_2^2) \quad (10.37)$$

To obtain the static field of displacement associated with a force couple, we take the limit as  $\delta \xi_2 \rightarrow 0$  and  $f \rightarrow \infty$ , so that we can always consider a finite moment  $f \delta \xi_2 \rightarrow M$ . Therefore, combining (10.31) and (10.37) we see that the displacement will be given by:

$$u_j^1(r) = -\frac{M}{8\pi\mu} \left[ -\delta_{i1}(2-\Gamma) \frac{x_2}{r^3} + \delta_{i2}\Gamma \frac{x_1}{r^3} - 3\Gamma \frac{x_1 x_2 x_i}{r^5} \right] \quad (10.38)$$

A similar calculation can be performed to determine the displacement field generated by a

force couple in the  $x_2$  direction with an arm in the  $x_1$  direction. It results:

$$u_i^2(r) = -\frac{M}{8\pi\mu} \left[ -\delta_{i2}(2-\Gamma) \frac{x_1}{r^3} + \delta_{i1}\Gamma \frac{x_2}{r^3} - 3\Gamma \frac{x_1 x_2 x_i}{r^5} \right] \quad (10.39)$$

Therefore, a double couple like that in Fig. 10.10 would generate the following displacement field:

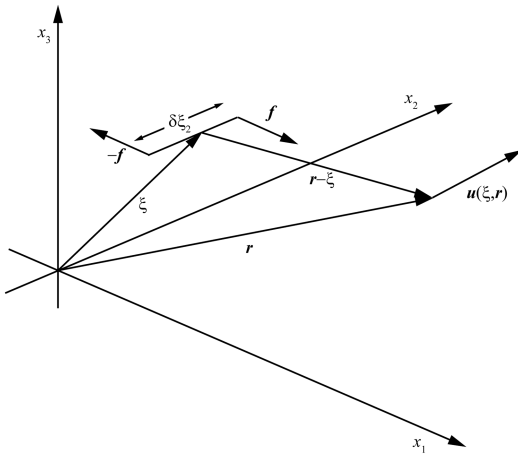
$$u_i(r) = -\frac{M}{4\pi\mu} \left[ (\Gamma-1) \left( \delta_{i1} \frac{x_2}{r^3} + \delta_{i2} \frac{x_1}{r^3} \right) - 3\Gamma \frac{x_1 x_2 x_i}{r^5} \right] \quad (10.40)$$

Finally, in spherical coordinates the double-couple displacement field assumes the form:

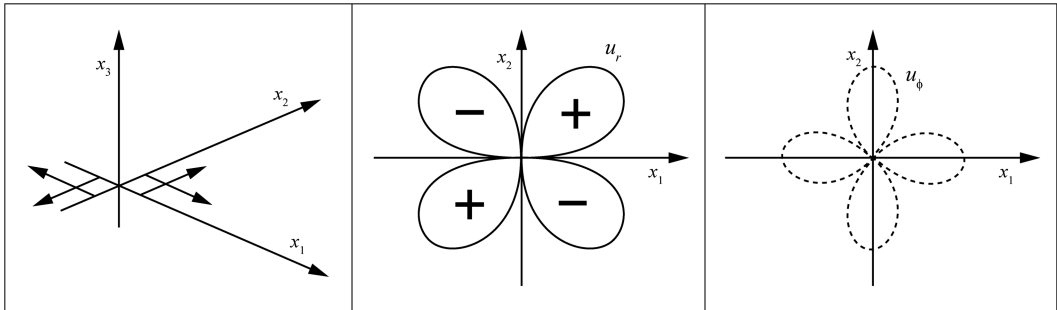
$$\begin{cases} u_r = \frac{M}{4\pi\mu r^2} (1 + \frac{\Gamma}{2}) \sin^2 \theta \sin 2\phi \\ u_\theta = \frac{M}{8\pi\mu r^2} (1 - \Gamma) \sin 2\theta \sin 2\phi \\ u_\phi = \frac{M}{4\pi\mu r^2} (1 - \Gamma) \sin \theta \cos 2\phi \end{cases} \quad (10.41)$$

A comparison of this solution with (10.34) shows that the displacement field associated with a double couple decreases with distance more rapidly (as  $1/r^2$ ) than that generated by a point force. In the  $x_1 x_2$  plane we have that  $\theta = \pi/2$ , thereby  $u_\theta = 0$  and:

$$\begin{cases} u_r = \frac{M}{4\pi\mu r^2} (1 + \frac{\Gamma}{2}) \sin 2\phi \\ u_\phi = \frac{M}{4\pi\mu r^2} (1 - \Gamma) \cos 2\phi \end{cases} \quad (10.42)$$



**Fig. 10.9** A force couple in the  $x_1$  direction is applied at location  $\xi = (\xi_1, \xi_2, \xi_3)$  with an arm  $\delta \xi_2$  in the  $x_1 x_2$  plane



**Fig. 10.10** A double force couple in the  $x_1 x_2$  plane and the corresponding pattern of displacement for  $u_r$  and  $u_\phi$

The corresponding pattern of displacement for these components is illustrated in Fig. 10.10. The importance of these solutions in seismology arises from the idea that *a distribution of equivalent double couples produces a displacement field that is indistinguishable from the displacements around a fault plane after an earthquake*. This is one of the most fundamental principles of seismology, which underlies the construction of equivalent systems of forces in the modeling of real seismic events, both in the elasto-static and in the elasto-dynamic cases.

Now we are ready to consider the elasto-dynamic solutions of the momentum equation. In the general non-static case this equation can be written, in vector notation, as follows:

$$\mathbf{f} + (\lambda + 2\mu) \nabla (\nabla \cdot \mathbf{u}) - \mu \nabla \times \nabla \times \mathbf{u} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad (10.43)$$

This time we shall assume a time-dependent body force of the form:  $\mathbf{f}(\mathbf{r}, t) = g(t)\delta(\mathbf{r})\mathbf{n}$ , where  $g(t)$  represents the time history. For example,  $g(t)$  could be a delta function  $\delta(t)$  or a step function  $H(t)$ . Therefore, (10.18) can be rewritten as follows:

$$\begin{aligned} \mathbf{f}(\mathbf{r}) &= g(t)\delta(\mathbf{r})\mathbf{n} = -g(t)\nabla^2 \left( \frac{\mathbf{n}}{4\pi r} \right) \\ &= -g(t)\nabla \left( \nabla \cdot \frac{\mathbf{n}}{4\pi r} \right) + \nabla \times \nabla \times \left( \frac{\mathbf{n}}{4\pi r} \right) \end{aligned} \quad (10.44)$$

Again, we assume that two vector potentials exist,  $\mathbf{A}_p$  and  $\mathbf{A}_s$ , such that the displacement  $\mathbf{u}$  has the form (10.21). In this instance, the wave equation splits into two distinct equations, one for each potential.

We obtain:

$$\begin{cases} \nabla^2 \mathbf{A}_p - \frac{1}{\alpha^2} \frac{\partial^2 \mathbf{A}_p}{\partial t^2} = \frac{g(t)\mathbf{n}}{4\pi r(\lambda + 2\mu)} \\ \nabla^2 \mathbf{A}_s - \frac{1}{\beta^2} \frac{\partial^2 \mathbf{A}_s}{\partial t^2} = \frac{g(t)\mathbf{n}}{4\pi r\mu} \end{cases} \quad (10.45)$$

where  $\alpha$  and  $\beta$  are the  $P$  and  $S$  wave velocities respectively (Eqs. 8.13 and 8.28). Assuming as before that the potentials  $\mathbf{A}_p$  and  $\mathbf{A}_s$  have the same direction of  $\mathbf{n}$ , so that  $\mathbf{A}_p = A_p\mathbf{n}$  and  $\mathbf{A}_s$

$= A_s\mathbf{n}$ , gives the following inhomogeneous wave equations:

$$\begin{cases} \nabla^2 A_p - \frac{1}{\alpha^2} \frac{\partial^2 A_p}{\partial t^2} = \frac{g(t)}{4\pi r(\lambda + 2\mu)} \\ \nabla^2 A_s - \frac{1}{\beta^2} \frac{\partial^2 A_s}{\partial t^2} = \frac{g(t)}{4\pi r\mu} \end{cases} \quad (10.46)$$

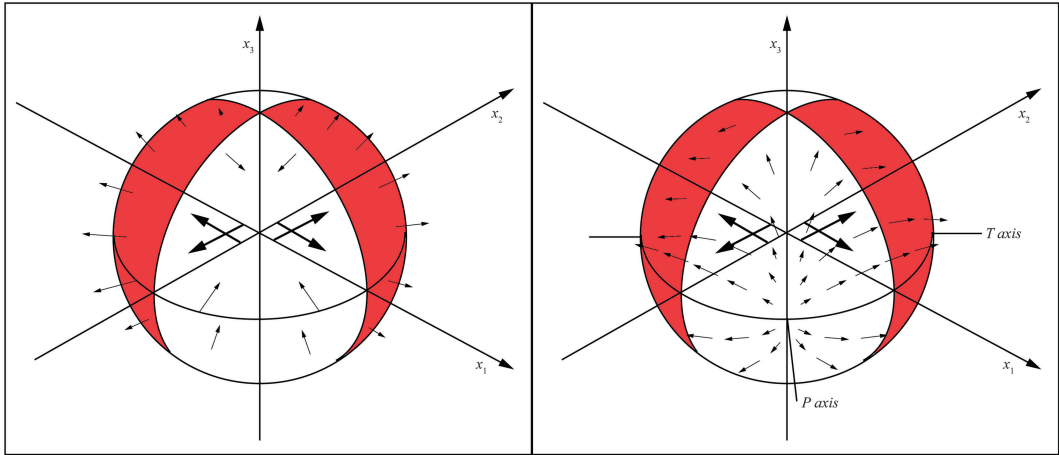
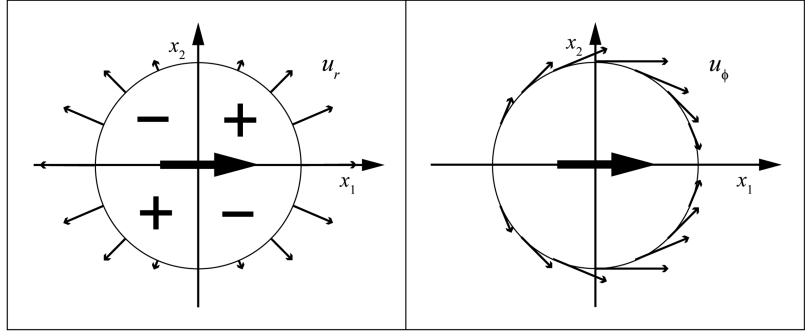
Finding a solution to these equations is rather complicate and involves Fourier transform techniques. The interested reader is referred to the book of Lay and Wallace (1995) for a detailed treatment of this subject. Therefore, we will give without proof the classic *Stokes solution* for the displacement associated with a point force at the origin in the direction  $x_j$ :

$$\begin{aligned} u_i^j(\mathbf{r}, t) &= \frac{1}{4\pi\rho r^3} \left( 3 \frac{x_i x_j}{r^2} - \delta_{ij} \right) \int_{r/\alpha}^{r/\beta} g(t-t') t' dt' \\ &\quad + \frac{1}{4\pi\rho\alpha^2 r} \frac{x_i x_j}{r^2} g\left(t - \frac{r}{\alpha}\right) \\ &\quad - \frac{1}{4\pi\rho\beta^2 r} \left( \frac{x_i x_j}{r^2} - \delta_{ij} \right) g\left(t - \frac{r}{\beta}\right) \end{aligned} \quad (10.47)$$

We note that the first term of this solution behaves like  $1/r^2$ , thereby it is usually referred to as the *near-field* term, while the other terms behave like  $1/r$ . Consequently, they are called the *far-field* terms. The first of them corresponds to a  $P$  wave that propagates with velocity  $\alpha$ . Its contribution to the displacement is a radial component. The second far-field term is associated with an  $S$  wave propagating with velocity  $\beta$ . It is easy to prove that the scalar product with  $\mathbf{r}$  is zero, so that its contribution consists of a tangential displacement. Both far-field terms are proportional to the magnitude of the applied force. Figure 10.11 illustrates the radiation pattern for a single force applied in the  $x_1$  direction. Finally, to obtain the displacement field associated with force couples and double couples we can apply the same procedure described in the elasto-static context.

For a double couple in the  $x_j x_k$  plane, oriented as the coordinate axes, the far-field  $P$  and  $S$  wave radiation patterns are given by:

**Fig. 10.11** Far-field components of displacement in the  $x_1x_2$  plane for a single force in the  $x_1$  direction. *Left:*  $P$  wave field; *Right:*  $S$  wave field



**Fig. 10.12** Far-field components of displacement for a double couple in the  $x_1x_2$  plane. *Left:*  $P$  wave field; *Right:*  $S$  wave field. The orientation of the arrows shows the direction of first motion and the length is proportional to the displacement. In the red quadrants,  $P$  wave first motion is outward, thereby they are called the *compressional*

*quadrants*. Conversely, in the *dilatational quadrants* (white regions)  $P$  wave first motion is inward. *Nodal lines* of zero motion separate the  $P$ -wave polarities into four quadrants. The *tension axis* (or  $T$  axis) is in the middle of the compressional quadrant; the *pressure axis* ( $P$  axis) is in the middle of the dilatational quadrant

$$u_i^P(\mathbf{r}, t) = \frac{1}{2\pi\rho\alpha^3} \frac{x_i x_j x_k}{r^3} \frac{\dot{M}(t - r/\alpha)}{r} \quad (10.48)$$

$$u_i^S(\mathbf{r}, t) = \frac{1}{4\pi\rho\beta^3} \frac{x_k}{r} \left( \delta_{ij} - \frac{x_i x_j}{r^2} \right) \frac{\dot{M}(t - r/\beta)}{r} \quad (10.49)$$

where  $\dot{M}$  is termed the *moment rate* function. A three-dimensional view of the first-motion far-field radiation pattern corresponding to a double couple in the  $x_1x_2$  plane is illustrated in Fig. 10.12. As in the elasto-static context, we note

that the displacement field is four-lobed, but now the lobes represent *first-motion* displacements, not static dislocations.

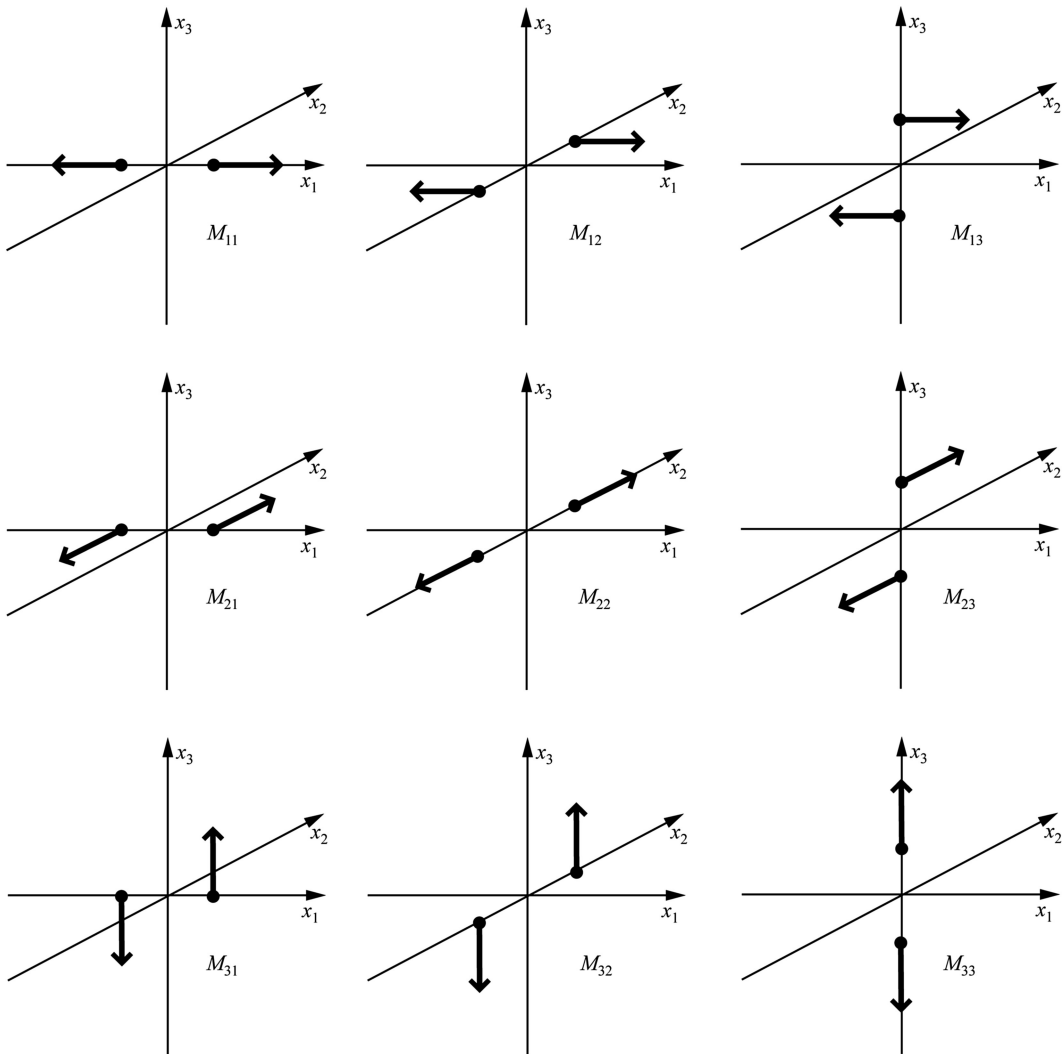
So far, we have not discussed the meaning of the near-field term in (10.47). This term combines radial and tangential contributions and a complete solution for double couple sources can be found in Aki and Richards (2002). However, the pattern of near-field displacement is important only at short distances from an earthquake focus, where it represents permanent static deformation. In the next section, we shall use the double couple description of seismic sources and the

corresponding radiation pattern to set up a quantitative characterization of the kinematic and dynamic parameters associated with an earthquake.

### 10.3 Moment Tensor

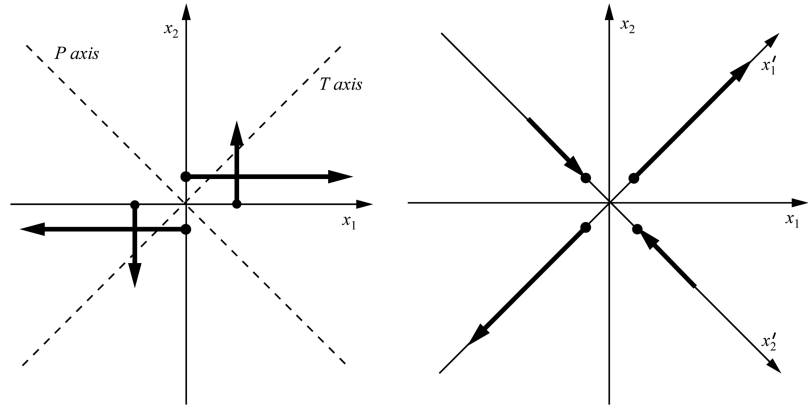
Let us consider the set of all force couples in a Cartesian coordinate system, such that the opposing forces are directed as the  $x_i$  axis and separated in the  $x_j$  direction. The nine possibilities are illustrated in Fig. 10.13. If assign a magnitude  $M_{ij}$  to each of these

couples, we obtain a rank-two tensor that is termed the *moment tensor*. Clearly, if the force couples must be used to build double couples that are representative of real seismic sources, we need to ensure that the total torque be zero. This requirement then translates into a specific condition of symmetry for the moment tensor components:  $M_{ij} = M_{ji}$ . Therefore, the moment tensor represents three different double couples and three force pairs with zero moment, corresponding to the diagonal elements. The latter system of forces can be used in the modelling of explosive sources but not



**Fig. 10.13** The nine force couples composing the seismic moment tensor

**Fig. 10.14** Principal axes and diagonalization of the moment tensor



for the representation of earthquake sources. The moment tensor allows to build a simple representation of arbitrarily oriented double couples by a combination of double couples that have force vectors oriented as the coordinate axes. For example, for a double couple in the  $x_1x_2$  plane, we would have that  $\mathbf{M}$  assumes the form:  $M_{ii} = 0, M_{12} = M_{21} = M_0, M_{13} = M_{31} = M_{23} = M_{32} = 0$ , where  $M_0$  is the magnitude of the single couples.

Using the moment tensor formalism and on the basis of the superposition principle, we have that for an arbitrary double couple the far-field solutions (10.48) and (10.49) assume the following general form:

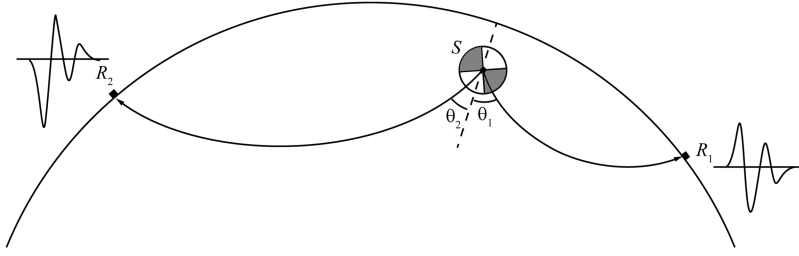
$$u_i^P(\mathbf{r}, t) = \frac{1}{2\pi\rho\alpha^3} \frac{x_i x_j x_k}{r^3} \frac{\dot{M}_{jk}(t - r/\alpha)}{r} \quad (10.50)$$

$$u_i^S(\mathbf{r}, t) = \frac{1}{4\pi\rho\beta^3} \frac{x_k}{r} \left( \delta_{ij} - \frac{x_i x_j}{r^2} \right) \frac{\dot{M}_{jk}(t - r/\beta)}{r} \quad (10.51)$$

Now we recall from the previous section that double couples are systems of equivalent body forces that produce the same radiation pattern of real shear dislocations. Consequently, their orientation is related to slip directions and fault plane orientations associated with earthquakes. For example, we expect that the pattern of displacements for an E–W oriented right-lateral strike-slip fault can be described through an equivalent model that includes an E–W oriented couple of forces and a conjugate couple having N–S ori-

entation. Therefore, the E–W directed coseismic slip along the fault plane would have the same direction as the main force couple. However, it is not possible to infer a unique focal mechanism for an earthquake, in terms of strike, dip, and rake, starting from the observed radiation pattern. For example, the pattern illustrated in Fig. 10.12 could have been generated either by an E–W oriented right-lateral strike-slip fault or by an N–S oriented left-lateral fault. This ambiguity in the source mechanism is intrinsic in the moment tensor representation and cannot be avoided. There are always two complementary focal mechanisms that are consistent with far-field seismic observations. The true fault plane is termed the *primary fault plane*, while the other solution is called the *auxiliary fault plane*. While the determination of the correct primary fault plane is easy when direct geological observation of the structures is available, in the case of deep faults (e.g., associated with intra-slab deformation) it is necessary to consider the distribution of *aftershock locations*. In fact, all large earthquakes are followed by a sequence of smaller earthquakes, the aftershocks, which are distributed along the fault plane and are associated with readjustments of the stress field after the mainshock.

The symmetry of the moment tensor implies that it can be diagonalized by solving an eigenvalue problem and applying the corresponding similarity transformation. For a double couple in the  $x_1x_2$  plane, the result of this operation is illustrated in Fig. 10.14 and the components of the diagonalized tensor will be:  $M_{ij} = 0$  for  $i \neq j$ ,



**Fig. 10.15** Focal sphere,  $S$ , about an earthquake focus. The first-motion polarity recorded at two seismic stations  $R_1$  and  $R_2$  depends from their azimuth, from the take-off

angle ( $\theta_1$  or  $\theta_2$ ), and from the orientation of the compressional and dilatational quadrants on the focal sphere

$M_{11} = M_0$ ,  $M_{22} = -M_0$ ,  $M_{33} = 0$ . Consequently, the new  $x_1$  and  $x_2$  axes are termed, respectively, the *tension axis*,  $T$ , and the *pressure axis*,  $P$ . These axes indicate the directions of minimum and maximum compressional stress, respectively, and their orientation with respect to the far-field radiation pattern is illustrated in Fig. 10.12.

The polarity (up or down) of first motion associated with the arrival of  $P$  waves, measured on several vertical-component seismograms distributed around a seismic source, can be used to infer the radiation pattern and the focal mechanism of an earthquake. A convenient approach is to consider a small sphere around the source, which is called the *focal sphere*. The polarity of a  $P$  phase arrival determines whether the corresponding seismic ray left the focal sphere from a compressional (upward first motion) or dilatational quadrant (downward first motion). The next step is then to determine the points on the focal sphere that are crossed by the seismic rays linking the earthquake focus to each receiver  $R$  (Fig. 10.15). The location of one of these points on the focal sphere can be specified assigning a take-off angle,  $\theta_0$ , and a ray azimuth  $\phi_0$ . The former quantity can be easily determined using Snell's law (9.42), while the latter parameter depends from the location of the seismic station relative to the epicenter. The results from many observations are plotted using stereographic or equal-area projections and specific software is run to determine the best-fitting orientation of the compressional and dilatational quadrants on the focal sphere. The distribution of first-motion polarities on the focal sphere constrains the orientation of the primary and auxiliary planes as-

sociated with the focal mechanism. Usually, this is displayed plotting the lower hemisphere of the focal sphere, but in some cases it could be convenient to show a lateral view. These plots are referred to as *beach ball* plots and represent the standard way to describe the pattern of seismic deformation in a region. Some basic beach ball plots associated with common focal mechanisms are illustrated in Fig. 10.16.

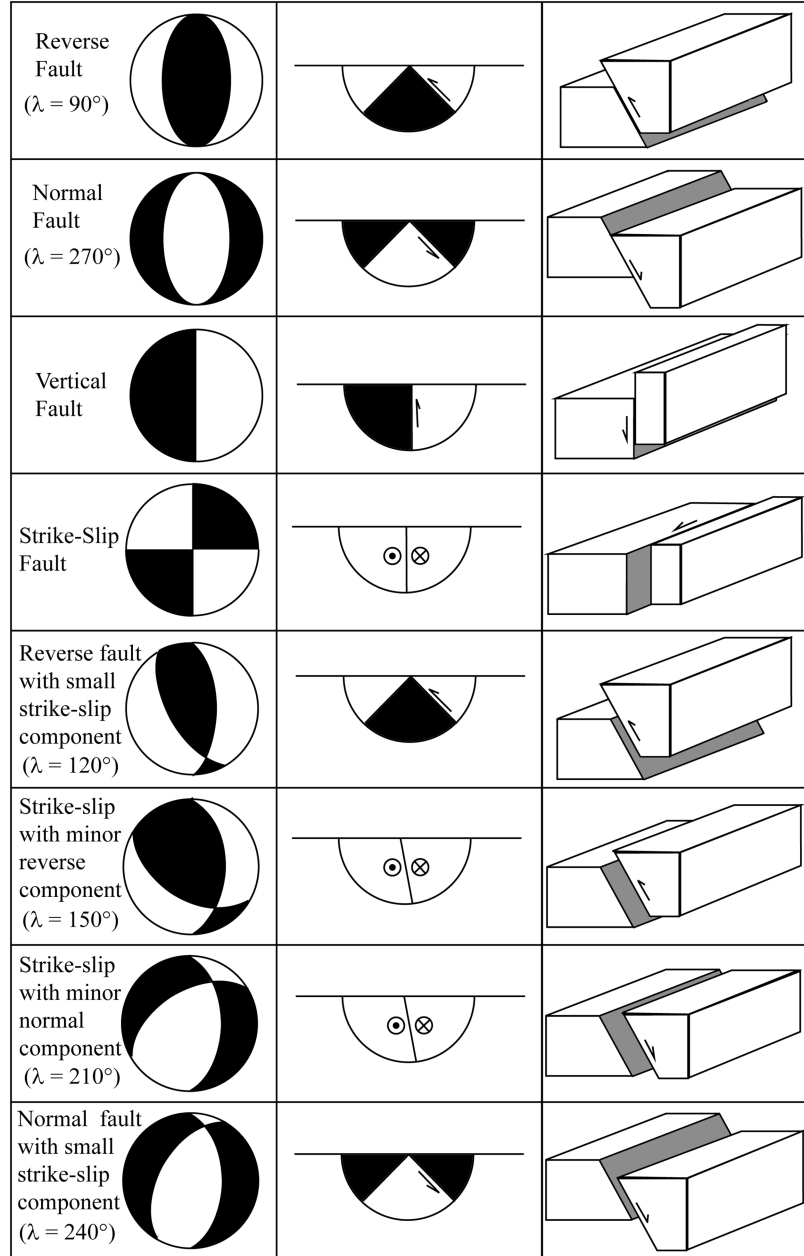
So far, we have focused on the *kinematic* parameters associated with an earthquake, which are grouped in the focal mechanism variables ( $\phi, \delta, \lambda$ ), and on a description of the procedure used for determining them. Therefore, it is time to consider the parameters that describe the *strength* of earthquakes, which are representative of the seismic energy released in the coseismic phase.

An earthquake results from sudden slip along a fault plane, with *finite* average magnitude  $\bar{u}$  and direction represented by the rake parameter  $\lambda$ . Such dislocation, which is unrecoverable and cannot be considered as an elastic displacement, determines a *stress drop* along the fault plane, as illustrated in Fig. 10.17. Burridge and Knopoff (1964) showed that if  $S$  is a fault plane and  $u(x, y)$  is the coseismic displacement at a point  $(x, y)$  on  $S$ , then a measure of the earthquake size is given by:

$$M_0 = \mu \int_S u(x, y) dS = \mu \bar{u} S \quad (10.52)$$

where  $S$  is the fault area and  $\mu$  is the rigidity modulus. The quantity  $M_0$  is called the *scalar seismic moment* and has dimension [Nm]. It represents the most important parameter in seismology for the specification of the strength of earthquakes

**Fig. 10.16** Lower-half (left) and lateral (center) views of some basic focal spheres and corresponding fault geometry and kinematics (right). Only one of the two possible focal mechanisms is displayed. The compressional quadrants are shown in *black*

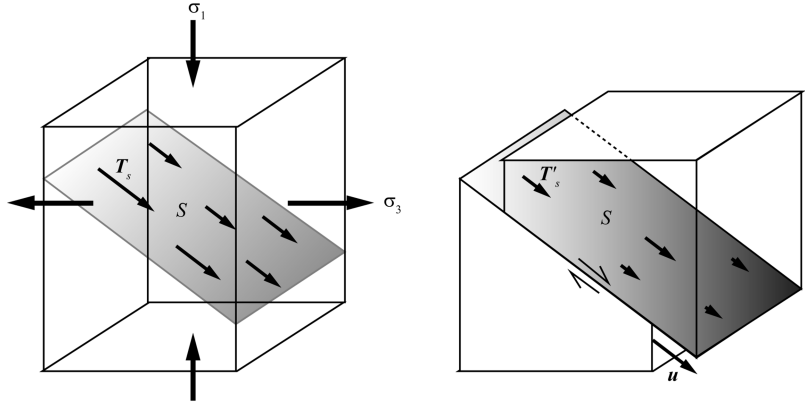


caused by fault slip. Observed value of the seismic moment range between  $\sim 4 \times 10^{22}$  Nm (the largest earthquake included in the Harvard CMT catalog) and  $\sim 10^5$  Nm for microearthquakes. If  $T_s(x, y)$  and  $T'_s(x, y)$  are the shear stresses exerted on the fault plane before and after an earthquake, then the stress drop can be defined as the average change of stress along the fault plane:

$$\delta T_s = \frac{1}{S} \int_S [T'_s(x, y) - T_s(x, y)] dS \quad (10.53)$$

The stress drop represents the fraction of stress that is employed to generate slip along the fault. In a similar way, we can define the average shear stress during the coseismic phase.

**Fig. 10.17** Stress drop across a fault plane after an earthquake.  $T_s$  and  $T'_s$  are the local fields of shear stress along  $S$  before and after slip  $u$



This quantity is clearly:

$$M_{ij} = M_0 (n_i \hat{u}_j + n_j \hat{u}_i) \quad (10.56)$$

$$\bar{T}_s = \frac{1}{S} \int_S [T'_s(x, y) + T_s(x, y)] dS \quad (10.54)$$

Now we note that  $\bar{T}_s S$  represents the total surface force exerted across the fault plane. Therefore, the total energy release during an earthquake will be given by:

$$E = \bar{T}_s S \bar{u} = \frac{\bar{T}_s}{\mu} M_0 \quad (10.55)$$

The stress drop associated with large earthquakes is generally estimated from the seismic moment. In the case of shallow events, it varies between 1 and 10 MPa. Often earthquakes along plate boundaries have lower stress drops than intraplate events (Shearer 2009 and references therein). It is estimated that the average  $\delta T_s$  is  $\sim 3$  MPa in the case of plate boundary earthquakes and  $\sim 6$  MPa for intraplate events. This is possibly a consequence of a lower area of intraplate faults with respect to the typical dimensions of plate boundary faults.

The scalar seismic moment measures the strength of an earthquake. Therefore, it is quite intuitive that it must be related to the moment tensor components  $M_{ij}$ . In fact, it is possible to show (e.g., Udías 1999) that if  $\mathbf{n}$  is the versor normal to a fault plane and  $\hat{\mathbf{u}}$  is the seismic slip versor, so that  $\mathbf{n} \cdot \hat{\mathbf{u}} = 0$ , then an analytic expression for the moment tensor is:

As we expected, this expression gives a symmetric tensor. Furthermore, it results:

$$Tr(\mathbf{M}) = M_{kk} = 2M_0 n_k \hat{u}_k = 2M_0 (\mathbf{n} \cdot \hat{\mathbf{u}}) = 0 \quad (10.57)$$

Therefore, the trace of the moment tensor is always zero. The expression (10.56) shows that the moment tensor can be separated into a scalar strength, corresponding to the seismic moment  $M_0$ , and a kinematic component associated with the tensor quantity  $m_{ij} \equiv n_i \hat{u}_j + n_j \hat{u}_i$ . The latter can be rewritten to make explicit the relation between the components of  $\mathbf{M}$  and the focal mechanism. In fact, the components of the versor  $\mathbf{n}$  depend from both the fault strike,  $\phi$ , and the dip,  $\delta$ , while  $\hat{\mathbf{u}}$  depends from the rake  $\lambda$ . Therefore, it is easy to prove that the components of  $\mathbf{M}$  in a local reference frame oriented as  $x \equiv \text{N}$ ,  $y \equiv \text{E}$ ,  $z \equiv \text{Down}$  are given by:

$$\begin{cases} M_{11} = -M_0 (\sin \delta \cos \lambda \sin 2\phi \\ \quad + \sin 2\delta \sin \lambda \sin^2 \phi) \\ M_{12} = +M_0 (\sin \delta \cos \lambda \cos 2\phi \\ \quad + 0.5 \sin 2\delta \sin \lambda \sin 2\phi) \\ M_{13} = -M_0 (\cos \delta \cos \lambda \cos \phi \\ \quad + \cos 2\delta \sin \lambda \sin \phi) \\ M_{22} = +M_0 (\sin \delta \cos \lambda \sin 2\phi \\ \quad - \sin 2\delta \sin \lambda \cos^2 \phi) \\ M_{23} = -M_0 (\cos \delta \cos \lambda \sin \phi \\ \quad - \cos 2\delta \sin \lambda \cos \phi) \\ M_{33} = +M_0 \sin 2\delta \sin \lambda \end{cases} \quad (10.58)$$

## 10.4 Earthquake Magnitude

The size of earthquakes is usually measured through their *magnitude*. There are several empirical magnitude scales, which are based on simple measurements of the largest seismic wave amplitude as recorded on seismograms. In the 1930s, Charles Richter proposed what is now known as the *local magnitude*,  $M_L$  (Richter 1935). While cataloging data from the southern California seismic network, Richter observed that plots of the logarithm of the largest amplitude,  $A$ , recorded on a standard instrument, the *Wood-Anderson seismograph*, versus epicentral distances exhibited the same decay rate (Fig. 10.18).

We recall that the amplitude of seismic waves decays as a consequence of geometrical spreading and attenuation. Therefore, the displacement recorded on seismograms may reflect the earthquake size only after a correction that accounts for the epicentral distance. Furthermore, the displacement results to be proportional to the moment *rate* (see Eqs. 10.50 and 10.51), not to the scalar moment (which is the correct quantity describing earthquake sizes). However, the necessity to determine promptly earthquake magnitudes through simple routine measurements has led to the widespread usage of scales based on waveform amplitudes. Richter, in particular, observed that the *distance* between the decay curves of two different earthquakes appeared to be approximately independent from the epicentral distance. Therefore, it was possible to define a distance-independent measure of the

earthquake size through the offset of  $\log_{10} A$  from a reference value:

$$M_L = \log_{10} A(\Delta) - \log_{10} A_0(\Delta) \quad (10.59)$$

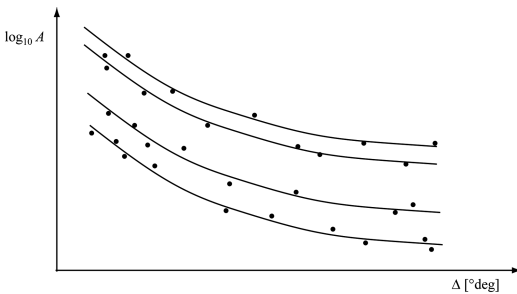
Richter built his scale using a reference epicentral distance  $X = 100$  km and a reference amplitude  $A_0 = 10^{-3}$  m. Using these values, (10.59) can be rewritten as follows:

$$M_L = \log_{10} A(\Delta) + 2.76 \log_{10} \Delta - 2.48 \quad (10.60)$$

The largest amplitude on Wood-Anderson torsion seismometers is generally associated with the  $S$  wave arrival on horizontal sensors. Earthquakes with  $M_L \leq 2.5$  are not generally felt by people, so that they are called *microearthquakes*. Since dominant period of the Wood-Anderson instrument (0.8 s) is similar to that of most buildings, the Richter (or local) scale is especially useful in engineering seismology. This scale is also important because all the more recent magnitude scales have been tied to it. A much more useful scale in plate tectonics is the *moment magnitude* scale proposed by Kanamori (1977), which is defined as follows:

$$M_w = \frac{2}{3} \log_{10} M_0 - 6.03 \quad (10.61)$$

where  $M_0$  is the scalar moment measured in N m. This expression shows that the moment magnitude scale provides a simple way to write the large numbers usually associated with the scalar moment. We also note that a unit increase in  $M_w$  does not coincide with a tenfold increase in earthquake size. Rather, if  $M_0$  changes from  $10^n$  to  $10^{n+1}$ , then  $M_w$  will increase by  $2/3$ .

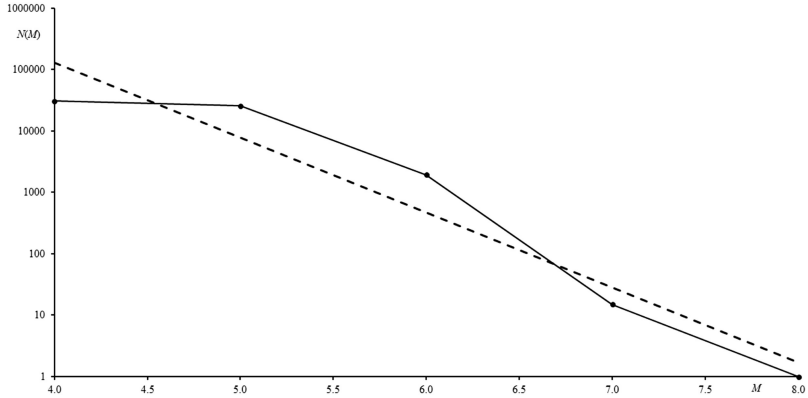


**Fig. 10.18** Amplitude decay curves for four earthquakes

## 10.5 Gutenberg-Richter Law

Tens of years of seismological observation have shown that there exists a precise statistical relationship between the magnitude of earthquakes and their recurrence time. Gutenberg and Richter (1956) were among the first seismologists to

**Fig. 10.19** Fit of the Gutenberg-Richter distribution (*dashed line*) to the population of earthquakes in the global CMT catalog 1976–2009 (*solid line*). The best-fitting value of  $b$  is  $\sim 1.22$



recognize that such relation could be described by a power-law:

$$\log_{10} N(M) = a - bM \quad (10.62)$$

where  $N(M)$  is the number of earthquakes with magnitude greater than or equal to an assigned value  $M$  in some time interval and region. An example of modelling of the global seismicity by this law is illustrated in Fig. 10.19.

In expression (10.62), the parameter  $a$  depends from the total number of events with magnitude greater than or equal to zero, while  $b$  measures the relative number of large earthquakes versus small earthquakes. This parameter is referred to as the  $b$ -value and varies between 0.8 and 1.2 for a number of different regions. In seismically active regions, it is close to the unity. A  $b$ -value of 1 implies that for each unitary magnitude increase, there is a tenfold reduction in the number of earthquakes.

An empirical relation similar to the Gutenberg-Richter law holds for the seismic moment  $M_0$ . In this instance, the annual number of earthquakes with seismic moment greater than or equal to an assigned value  $M_0$  in some tectonic region is given by (Molnar 1979):

$$N(M_0) = \alpha M_0^{-\beta} \quad (10.63)$$

where  $\alpha$  and  $\beta$  are constants. This distribution does not take into account the important observation that for any given tectonic region there is a threshold seismic moment above which there

are no earthquakes. Molnar (1979) has proved that if  $M_{0,\max}$  is the maximum observed seismic moment, then the annual rate of scalar moment release is given by:

$$\dot{M}_0 = \frac{\alpha}{1-\beta} M_{0,\max}^{1-\beta} \quad (10.64)$$

As pointed out by Kagan (1991), there is evidence that the  $\beta$ -value represents a universal constant. At global scale, a statistical analysis of the CMT catalog gives  $\beta = 2/3$ . Although this is also the value suggested by Molnar (1979), theoretical arguments lead to the value  $\beta = 0.5$  for any specific tectonic region (Kagan 1991, 1993, 2010). We shall see in the next chapter that (10.63) and (10.64) play a key role in the study of the seismic deformation of the lithosphere.

## Problems

1. An active fault has been mapped by geologists. After the last earthquake, they have found that its length is 10 km, the strike is  $\phi = 30^\circ\text{N}$ , the dip is  $\delta = 45^\circ$ , an average 10 cm downward dip-slip displacement of the hanging wall occurred, while seismic reflection studies suggest that the fault reaches a depth of 5 km. Estimate the moment tensor components for this earthquake;
2. Prove the validity of expression (10.58) for the moment tensor components;

3. Expression (10.58) for  $\mathbf{M}$  can be rewritten as follows:

$$\begin{aligned}\mathbf{M} = & \cos \delta \cos \lambda \mathbf{M}^{(1)} + \sin \delta \cos \lambda \mathbf{M}^{(2)} \\ & - \cos 2\delta \sin \lambda \mathbf{M}^{(3)} + \sin 2\delta \sin \lambda \mathbf{M}^{(4)}\end{aligned}$$

where  $\mathbf{M}^{(k)}$  are four elementary moment tensors. Find expressions for  $\mathbf{M}^{(k)}$  and discuss the result.

## References

- Aki K, Richards PG (2002) Quantitative seismology, 2nd edn. University Science, Sausalito, 700 pp
- Berthoud P, Baumberger T, G'Sell C, Hiver J-M (1999) Physical analysis of the state-and rate-dependent friction law: static friction. *Phys Rev B* 59(22): 14313–14327
- Brace WF, Byerlee JD (1966) Stick slip as a mechanism for earthquakes. *Science* 153:990–992
- Burridge R, Knopoff L (1964) Body force equivalents for seismic dislocations. *Bull Seismol Soc Am* 54(6A):1875–1888
- Dieterich JH (1972) Time-dependent friction in rocks. *J Geophys Res* 77(20):3690–3697. doi:10.1029/JB077i020p03690
- Dieterich JH (1979) Modeling of rock friction: 1. Experimental results and constitutive equations. *J Geophys Res* 84(B5):2161–2168
- Dieterich JH, Linker MF (1992) Fault stability under conditions of variable normal stress. *Geophys Res Lett* 19(16):1691–1694
- Gutenberg B, Richter CF (1956) Magnitude and energy of earthquakes. *Ann Geofis* 9:1–15
- Kagan YY (1991) Seismic moment distribution. *Geophys J Int* 106:123–134
- Kagan YY (1993) Statistics of characteristic earthquakes. *Bull Seismol Soc Am* 83(1):7–24
- Kagan YY (2010) Earthquake size distribution: power-law with exponent  $\beta \equiv 1/2$ ? *Tectonophysics* 490:103–114
- Kanamori H (1977) The energy release in great earthquakes. *J Geophys Res* 82(20):2981–2987
- Lay T, Wallace TC (1995) Modern global seismology, vol 58, International Geophysics Series. Academic, San Diego, 521 pp
- Molnar P (1979) Earthquake recurrence intervals and plate tectonics. *Bull Seismol Soc Am* 69(1):115–133
- Nakatani M (2001) Conceptual and physical clarification of rate and state friction: frictional sliding as a thermally activated rheology. *J Geophys Res* 106(B7):13347–13380
- Panofsky WKH, Phillips M (2005) Classical electricity and magnetism. Courier Dover Publications.
- Reid HF (1910) The mechanics of the earthquake, the California earthquake of April 18, 1906. Report of the State Investigation Commission, 2. Carnegie Institution of Washington, Washington, DC
- Richter CF (1935) An instrumental earthquake magnitude scale. *Bull Seismol Soc Am* 25:1–32
- Ruina A (1983) Slip instability and state variable friction laws. *J Geophys Res* 88(B12):10359–10370
- Scholz CH (1998) Earthquakes and friction laws. *Nature* 391(6662):37–42
- Shearer PM (2009) Introduction to seismology, 2nd edn. Cambridge University Press, Cambridge, UK, 396
- Stein S, Wysession M (2003) An introduction to seismology, earthquakes, and earth structure. Blackwell Publishing Ltd., Carlton, 498
- Udías A (1999) Principles of seismology. Cambridge University Press, Cambridge, UK, 475



## Abstract

Here the principles of seismic deformation are illustrated, which represent an essential tool to unravel the kinematics of subduction. In particular, the classic Kostrov and Brune formulae are presented, as well as the more advanced representation of Molnar, based on an asymmetric strain tensor. Finally, this chapter introduces the application of these techniques to the study of slab deformation in subduction zones.

## 11.1 Kostrov's Formula

In the previous chapters, we have learnt that lithosphere deforms both by viscous creep, at the time scale of millions of years, and elastically in the short range (up to thousands of years). Therefore, at first glance seismicity does not provide a relevant source of data for studying neither the long-term pattern of deformation of tectonic plates nor plate motions. In this chapter, we are going to prove that such an appearance is misleading, because earthquakes supply both valuable qualitative information about the *style* of deformation (e.g., the modes of deformation of slabs), and a quantitative measure of the long-term strain rate under external loads.

The first attempt to link earthquakes to long-term tectonics was done by Brune (1968) in a seminal paper. This author considered the seismicity along a large shear zone with total surface  $S$ , consisting into a series of earthquakes with scalar moment  $M_0^{(k)}$  ( $k = 1, 2, \dots, n$ ), rupture surface  $S_k$ , and average displacement  $\bar{u}^{(k)}$  (Fig. 11.1).

By (10.52), we have that the average slip along the  $k$ -th rupture area is:

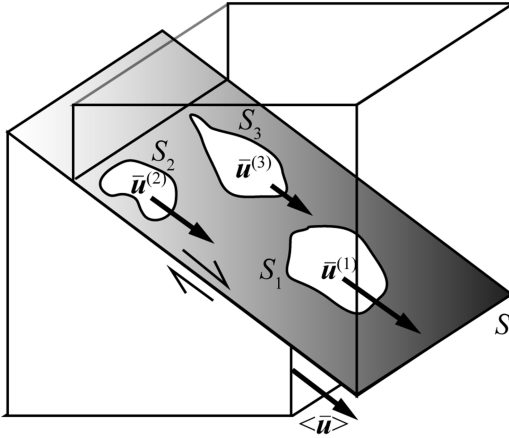
$$\bar{u}^{(k)} = \frac{M_0^{(k)}}{\mu S_k} \quad (11.1)$$

Therefore, if the overall displacement,  $\bar{u}$ , along the entire shear zone is calculated taking the weighted average of these quantities, with weights given by the size of the individual rupture areas, we have:

$$\bar{u} = \frac{1}{S} \sum_k \bar{u}^{(k)} S_k = \frac{1}{\mu S} \sum_k M_0^{(k)} \quad (11.2)$$

This is called *Brune's formula*. Assuming that the observed seismicity is relative to a sufficiently long time interval  $T$ , the formula can be used to estimate the average *rate of seismic slip* along a shear zone:

$$\frac{\partial \bar{u}}{\partial t} = \frac{1}{\mu S T} \sum_k M_0^{(k)} \quad (11.3)$$



**Fig. 11.1** Average long-term seismic slip in the Brune's model

Differently from Brune (1968), Kostrov (1974) considered the problem of describing the seismic deformation within a *finite volume*  $V$ , rather than dislocations along a planar plate boundary. In this instance, earthquakes are not confined to a planar surface, individual faults within the deforming region have different orientations, and slip along these faults occurs along distinct directions. Therefore, the contribution of each event to the overall deformation must include the source focal mechanism parameters or, alternatively, the components of the moment tensor. Kostrov (1974) proved that the components of the average *strain rate tensor* due to seismic slip are related to the moment tensor components of the individual earthquakes,  $M_{ij}^{(k)}$ , by the following expression:

$$\dot{\epsilon}_{ij} = \frac{1}{2\mu VT} \sum_k M_{ij}^{(k)} = \frac{\overline{M}_0}{2\mu VT} \sum_k m_{ij}^{(k)} \quad (11.4)$$

where  $\overline{M}_0$  is the average scalar seismic moment and  $m_{ij}$  is the geometrical part of the moment tensor, which can be expressed in terms of strike, dip, and rake (see 10.58). This formula has been widely used to determine at regional scale the seismic deformation of continental blocks

(e.g., Kiratzi and Papazachos 1996) or subducting slabs (e.g., Kiratzi and Papazachos 1995), and at large scale to build global strain rate models for the lithosphere (Kreemer et al. 2000). Although (11.4) appears as a generalization of (11.3) (and Kostrov himself considered his formula this way), this interpretation is not completely correct. As pointed out by Molnar (1983), Kostrov's formula leads to results that differ from those obtained using Brune's formula when the former is applied to a block that deforms along a unique shear zone (which is the situation considered by Brune). To understand the difference between the two formulations, let us consider the deformation field associated with the coseismic slip along a fault plane. The sudden non-elastic displacement during an earthquake can be viewed as a rotation of one block relative to the opposing block about an axis perpendicular to the fault plane. Clearly, conservation of the angular momentum requires that the moment associated with slip be balanced by an opposite moment with equal magnitude. Such a complementary moment will determine elastic deformation in the region surrounding the rupture area border. Brune's formulation allows to calculate the net rotation associated with seismic slip independently from the elastic deformation occurring near the fault ends. In this instance, the strain tensor will be identically zero, while the displacement field can be described uniquely by an antisymmetric rotation tensor  $\omega_{ij}$  (see Sect. 7.2). Consequently, this model cannot completely describe the deformation within a 3-D volume. Conversely, in his formulation Kostrov assumed that the deformation was pure shear (i.e., without any rotational component). In fact, the symmetric strain rate tensor that results from summation of symmetric moment tensors excludes any rotational displacement. As a consequence, Kostrov's model cannot describe situations where any of the faults in the deforming region intersects the region boundary. In particular, it cannot describe the simple shear of a block. In the next section, we shall consider a more general formulation that combines both approaches.

## 11.2 The Asymmetric Strain Tensor

Molnar (1983) proposed the following solution to the problem of estimating the average strain rate of a deforming region from its seismicity. Let us consider a rectangular region of thickness  $h$  and dimensions  $w$  and  $a$ , crossed from side to side by a vertical fault of length  $L$  and strike  $\theta$  (Fig. 11.2). Let us assume that a small (compared to  $w$ ,  $a$ , and  $L$ ) strike-slip displacement  $\delta u$  occurs along the fault during an earthquake. In this instance, the average strain in the  $y$  direction will be given by:

$$\varepsilon_{yy} = \frac{\delta u \cos \theta}{w} \quad (11.5)$$

According to (10.52), the coseismic displacement  $\delta u$  determines a scalar moment  $M_0$  given by:

$$M_0 = \mu L h \delta u \quad (11.6)$$

Therefore, (11.5) can be rewritten as follows:

$$\varepsilon_{yy} = M_0 \frac{\sin \theta \cos \theta}{\mu V} \quad (11.7)$$

where  $V = awh$  is the volume of the region. Similarly, the shear strain will be given by:

$$\varepsilon_{xy} = \frac{\delta u \sin \theta}{w} = M_0 \frac{\sin^2 \theta}{\mu V} \quad (11.8)$$

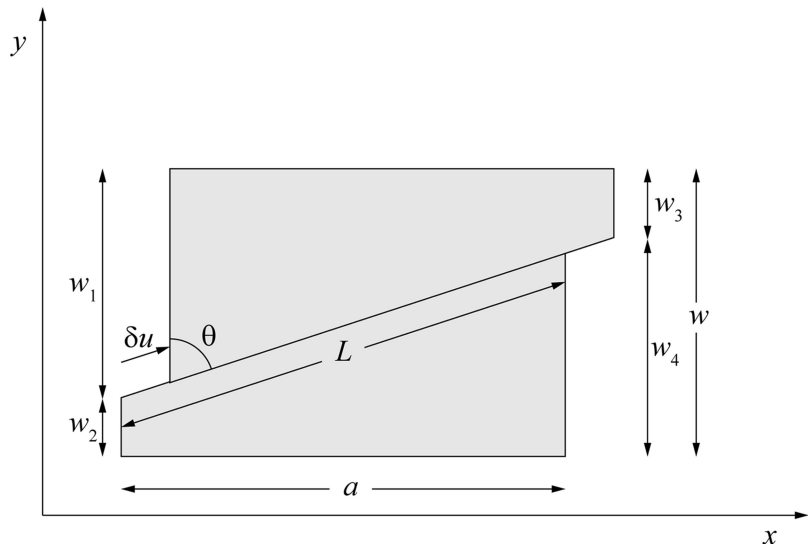
Let us consider now the components  $\varepsilon_{xx}$  and  $\varepsilon_{yx}$ . The points along the left side of the region have displacement zero in the  $x$  direction for a length  $w_2$ , and  $\delta u \sin \theta$  for a length  $w_1$ . Therefore, the average displacement of the left side in the  $x$  direction is  $(w_1/w)\delta u \sin \theta$ . Similarly, the mean displacement of the right side in the  $x$  direction is  $(w_3/w)\delta u \sin \theta$ . Therefore, the average change in length is  $\delta a = (w_3 - w_1)\delta u \sin \theta / w = -L \delta u \sin \theta \cos \theta / w$ .

Then, the average strain in the  $x$  direction will be given by:

$$\varepsilon_{xx} = \frac{\delta a}{a} = -\frac{L \delta u \sin \theta \cos \theta}{aw} = -M_0 \frac{\sin \theta \cos \theta}{\mu V} \quad (11.9)$$

A similar calculation leads to the following expression for  $\varepsilon_{yx}$ :

$$\varepsilon_{yx} = -\frac{L \delta u \cos^2 \theta}{aw} = -M_0 \frac{\cos^2 \theta}{\mu V} \quad (11.10)$$



**Fig. 11.2** A rectangular region cut by a vertical fault with strike  $\theta$

Expressions (11.7), (11.8), (11.9), and (11.10) define an *asymmetric strain tensor* that is proportional to the scalar moment  $M_0$ . Now let us introduce the slip versor  $\hat{\mathbf{u}}$  and the versor  $\mathbf{n}$  perpendicular to the fault plane. It results:

$$\begin{cases} \hat{\mathbf{u}} = \sin \theta \mathbf{i} + \cos \theta \mathbf{j} \\ \mathbf{n} = -\cos \theta \mathbf{i} + \sin \theta \mathbf{j} \end{cases} \quad (11.11)$$

Now let us define an *asymmetric moment tensor* as follows:

$$\Gamma_{ij} \equiv M_0 \hat{u}_i n_j \quad (11.12)$$

Then, from (11.7), (11.8), (11.9), and (11.10) we obtain:

$$\varepsilon_{ij} = \frac{1}{\mu V} \Gamma_{ij} \quad (11.13)$$

In the general case of many faults, the average strain will be calculated by summation of the asymmetric moment tensors associated with the individual seismic events:

$$\varepsilon_{ij} = \frac{1}{\mu V} \sum_k \Gamma_{ij}^{(k)} = \frac{\overline{M}_0}{\mu V} \sum_k \gamma_{ij}^{(k)} \quad (11.14)$$

where  $\overline{M}_0$  is the average scalar seismic moment and  $\gamma_{ij}$  is the geometrical part of the asymmetric moment tensor, which can be expressed in terms of strike, dip, and rake just like the corresponding symmetric tensor appearing in (11.4). From the definition (11.12), it is easy to prove that in the North–East–Down reference frame it results:

$$\begin{cases} \gamma_{11} = -0.5 (\sin \delta \cos \lambda \sin 2\phi + \sin 2\delta \sin \lambda \sin^2 \phi) \\ \gamma_{12} = \sin \delta \cos \lambda \cos^2 \phi + 0.25 \sin 2\delta \sin \lambda \sin 2\phi \\ \gamma_{13} = -\cos \delta \cos \lambda \cos \phi - \cos^2 \delta \sin \lambda \sin \phi \\ \gamma_{21} = -\sin \delta \cos \lambda \sin^2 \phi + 0.25 \sin 2\delta \sin \lambda \sin 2\phi \\ \gamma_{22} = +0.5 (\sin \delta \cos \lambda \sin 2\phi + \sin 2\delta \sin \lambda \cos^2 \phi) \\ \gamma_{23} = -\cos \delta \cos \lambda \sin \phi + \cos^2 \delta \sin \lambda \cos \phi \\ \gamma_{31} = \sin^2 \delta \sin \lambda \sin \phi \\ \gamma_{32} = -\sin^2 \delta \sin \lambda \cos \phi \\ \gamma_{33} = 0.5 \sin 2\delta \sin \lambda \end{cases} \quad (11.15)$$

Finally, assuming that the observed seismicity is relative to a sufficiently long time interval  $T$ , we can have an estimate of the average strain rate during that time interval:

$$\dot{\varepsilon}_{ij} = \frac{1}{\mu VT} \sum_k \Gamma_{ij}^{(k)} = \frac{\overline{M}_0}{\mu VT} \sum_k \gamma_{ij}^{(k)} \quad (11.16)$$

This is the asymmetric analogous of Kostrov's formula (11.4). The correct application of this formula requires that the time interval  $T$  be large enough to include events having large magnitude. These earthquakes have long recurrence intervals, as we have seen in Sect. 10.5. Therefore, we cannot assign to the variable  $T$  in (11.16) a value corresponding to the number of years represented in the local data base of seismic events (for example, a few tens of years in the case of the Harvard CMT catalog), because large events with recurrence intervals of the order of  $\sim 10^3$  years (Molnar 1979) could not be represented in this data base. This problem can be solved rewriting (11.16) as follows:

$$\dot{\varepsilon}_{ij} = \frac{\dot{M}_0}{\mu V} \sum_k \gamma_{ij}^{(k)} \quad (11.17)$$

where  $\dot{M}_0$  is the annual rate of seismic moment release.

The quantity  $\dot{M}_0$  can be estimated statistically assuming a distribution density,  $\phi$ , of seismic moment release. The simplest distribution is a power-law or Pareto distribution (Molnar 1979; Kagan 1991, 1993):

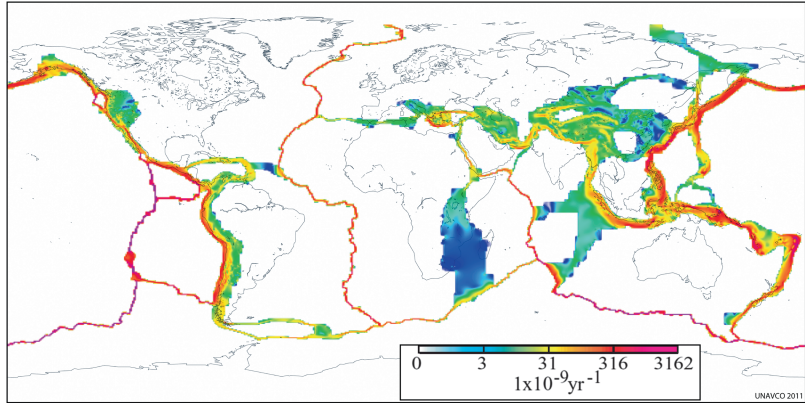
$$\phi(M) = \beta M_c^\beta M^{-1-\beta} \quad (M_c \leq M < \infty) \quad (11.18)$$

where  $M_c$  is a lower threshold and  $\beta \cong 0.5$  (Kagan 1991, 2010). In this instance, the annual number of events having moment  $M \geq M_0$  will be given by (10.63), with  $\alpha \equiv N_T M_c^\beta$ ,  $N_T$  being the annual number of events.

### 11.3 Global Pattern of Lithospheric Deformation

Kostrov's (1974) formula or one of its variants have been widely used to estimate geological strain rates at regional or global scale. For example, England and Molnar (1997) studied the active deformation of Asia covering the region

**Fig. 11.3** Second invariant of the strain rate field (After Kreemer et al. 2003)



by a mesh of triangles. For each mesh element, they estimated the strain rate using the seismicity of Quaternary faults and Kostrov's formula. Similarly, Corredor (2003) determined the seismic strain rate of the northern Andes region, while Papazachos and Kiratzi (1992) used the same technique to study the pattern of active deformation in central Greece. At global scale, Kreemer et al. (2000, 2003) used geodetic velocities, seismic moment tensors from the Harvard CMT catalog, and Quaternary fault slip rate data to build a model strain rate field along the major plate boundaries, in particular along the wide Alpine-Hymalaian collision zone. These authors assumed a constant shear modulus  $\mu = 35$  GPa, and seismogenic depths of 20 km for the continental areas, 30 km along subduction zones, 20 km for zones of diffuse oceanic deformation, and 7.5 km for oceanic ridges and transforms. Their global map, which is illustrated in Fig. 11.3, shows the distribution of the second invariant of the strain rate, which is defined by:

$$\dot{\epsilon}^{II} = \frac{1}{2} \left[ \dot{\epsilon}_{ij} \dot{\epsilon}_{ij} - (\dot{\epsilon}_{kk})^2 \right] = \dot{\epsilon}_{12}^2 + \dot{\epsilon}_{23}^2 + \dot{\epsilon}_{31}^2 - (\dot{\epsilon}_{11} \dot{\epsilon}_{22} + \dot{\epsilon}_{11} \dot{\epsilon}_{33} + \dot{\epsilon}_{22} \dot{\epsilon}_{33}) \quad (11.19)$$

The total annual release of seismic moment results to be  $7.7 \times 10^{21}$  Nm yr<sup>-1</sup> within the shallow seismogenic thickness of the lithosphere, but  $\sim 17\%$  of this total moment rate is concentrated in areas of diffuse deformation on continents.

In their approach, Kreemer et al. (2000, 2003) used Brune's formula (11.3) to estimate

the geodetic average rate of release of seismic moment starting from slip rates. For a system of  $n$  faults having length  $L_k$  and dip  $\delta_k$ , we can rewrite (11.3) as follows:

$$\langle \dot{M}_0^{(k)} \rangle = \mu S_k \langle \dot{u}_k \rangle = \frac{\mu L_k h}{\sin \delta_k} \langle \dot{u}_k \rangle \quad (11.20)$$

where  $h$  is the seismogenic thickness and the average slip rate is estimated from a combination of geodetic data (GPS, VLBI, etc.), regional Quaternary fault slip rate data, and seismic moment tensor information from shallow earthquakes. Inserting this expression into Kostrov's formula (11.4), Kreemer et al. (2000, 2003) obtained an estimate of the *geodetic-seismic* strain rate tensor:

$$\dot{\epsilon}_{ij} = \frac{1}{2A} \sum_k \frac{L_k \langle \dot{u}_k \rangle}{\sin \delta_k} m_{ij}^{(k)} \quad (11.21)$$

where  $A$  is the area of a grid cell for which the strain rate is estimated. In their study, Kreemer and colleagues divided the Earth's deforming regions into 24,500 grid cells having dimension  $0.6 \times 0.5^\circ$ . The possibility of using geologic and geodetic data to estimate strain rates had already been exploited by Ward (1998) in a study on the differences between geodetic, seismic, and geologic strain rates for the US region. This author determined the field of maximum geodetic strain rate through the largest eigenvalues of the geodetic strain rate tensor. The geodetic strain rates were used in turn to estimate the average geodetic moment rates. An interesting result

obtained by Ward (1998) was that the average release of seismic moment was 73–86 % of the average geodetic moment rate in California and ~70 % outside this state. Given the unlikelihood of a high rate of aseismic creep, he concluded that the earthquake catalogues failed to be representative of the long-term deformation. However, in the next section we shall see that the correct application of Kostrov's formula to the determination of the seismic deformation requires some caution, because the Gutenberg-Richter law (10.62) (and the equivalent expression in terms of seismic moment) demand a rigorous statistical analysis in the estimation of annual seismic rates. Similarly, using *observed* geologic slip rates for estimating the annual rate of seismic moment release through (11.20) should require an in-depth analysis of the statistical properties of the seismic slip data set. Therefore, the correct application of Kostrov's and Brune's formulae to the analysis of lithospheric deformation is still an open problem that demands further studies.

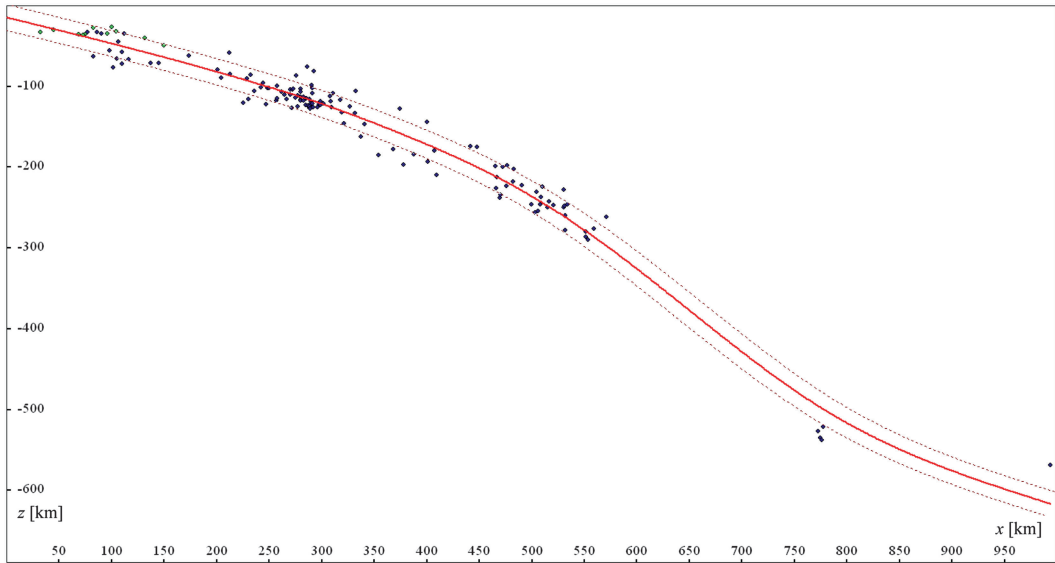
## 11.4 Bending of Slabs

Seismicity supplies an important source of data for the study of the geometry and deformation of subducting slabs. As pointed out by Isacks et al. (1968) in a seminal paper, almost all the global seismicity in the deep and intermediate range, and most of the world's shallow earthquakes are associated with island arcs or arc-like structures. Oliver and Isacks (1967), in a study of the Fiji-Tonga region, were among the first to recognize the existence in the upper mantle of an anomalous zone whose thickness was of the order of 100 km and whose upper surface was characterized by elevated seismicity up to depths of about 700 km. The zone was anomalous in that attenuation of seismic waves was low and seismic velocities were high relative to those of the surrounding asthenosphere. Subsequent studies confirmed the existence of such anomalous regions beneath all the world's island arcs, and they are today known as *Wadati-Benioff* zones. These zones delineate the geometry of slabs sinking into the mantle as a consequence of their negative buoyancy.

Figure 11.4 shows an example of subducting slab cross-section as outlined by its seismicity.

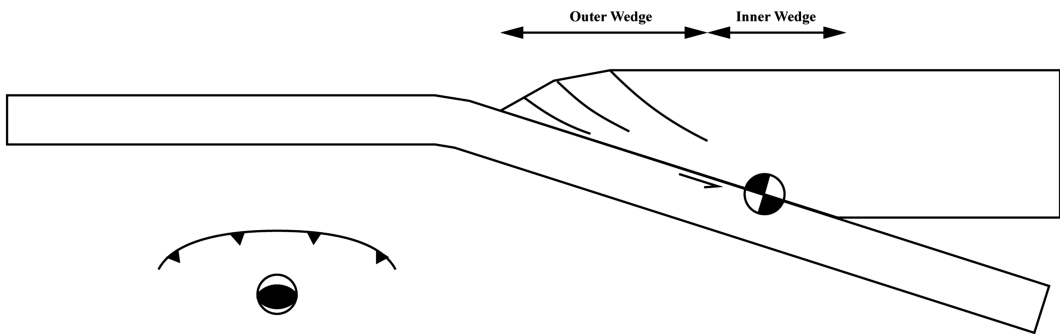
Seismicity of Wadati-Benioff zones can be divided in three categories. At shallow depths, up to ~80 km, slip of the bending lithosphere beneath the accretionary wedge produces very strong events that may have devastating effects on the people living in the forearc region. An example is the large Sumatra-Andaman earthquake of 26 December 2004, which had magnitude  $M_w = 9.1$  and was accompanied by a tsunami that caused more than 283,000 deaths (for a general description of this event, see Lay et al. 2005). Figure 11.5 illustrates the typical geological setting and focal mechanism associated with this class of events.

The other two sources of seismicity are associated with intra-plate earthquakes that reflect two modes of slab deformation. The first of them is referred to as *downdip* deformation and can be either *downdip shortening* or *downdip extension*. In both cases, the *P* or *T* axes are parallel to the local slab dip and the fault strike is approximately normal to the local downdip direction. Isacks and Molnar (1971) were the first to perform a systematic study of this kind of seismicity, with the objective of studying the distribution of stresses within the sinking lithosphere. They observed that there were remarkable systematic differences between intermediate-depth (between 70 and 300 km) and deep solutions. The deep earthquakes were dominated by down-dip compression, whereas the intermediate-depth solutions showed a variety of focal mechanisms, including downdip shortening, downdip extension, and many solutions that could not be classified as downdip deformation. Isacks and Molnar (1971) suggested that this distribution of seismicity was a consequence of the gravitational instability of the lithosphere, which is more dense than the surrounding mantle and sinks into the asthenosphere under its own weight. Therefore, slabs are subject to downdip extension when they penetrate the rheologically weak asthenosphere, and downdip compression when they reach regions of the upper mantle characterized by higher strength (hence resistance to penetration), in particular the transition zone.



**Fig. 11.4** Vertical cross-section across the Nazca slab, which is subducting beneath the Chile Trench. *Black dots* represent foci of intra-plate earthquakes, while underthrusting events are shown as *green circles*. All the earthquake foci in a vertical stripe 200 km wide have been projected onto this vertical cross-section. The Wadati-

Benioff zone is modelled by a 30 km thick seismogenic slab, built around a cubic spline regression curve of the events locations (*red line*). Most world's slabs bend downwards (increasing dip) in the asthenosphere and upwards in the transition zone (decreasing subduction angles)



**Fig. 11.5** Vertical cross-section illustrating the main geometric features of accretionary prisms and the typical focal mechanisms associated with slab underthrusting

The second source of intra-plate seismicity requires a more complex explanation. In Chap. 2, we have seen that the Earth's subduction zones, in particular the island arcs, are represented geometrically by small circle arcs. Frank (1968) was the first to prove that the curvature of the subduction zones is a function of the slab dip  $\alpha$ . He proposed a suggestive explanation that is known as the "ping-pong ball" model. In this model, the lithosphere is considered as a flexible-inextensible

spherical shell that bends downwards, determining a deformation pattern that can be assimilated to the dent of a ruptured ping-pong ball. In this process, there is a quantity that is conserved: the Gaussian curvature of the Earth's surface. To understand the concept of Gaussian curvature, let us consider a surface  $S$  having arbitrary shape. The curvature of  $S$  at a point  $P$  clearly varies if we consider different directions along the tangent plane at  $P$ . Euler in 1760 showed that the

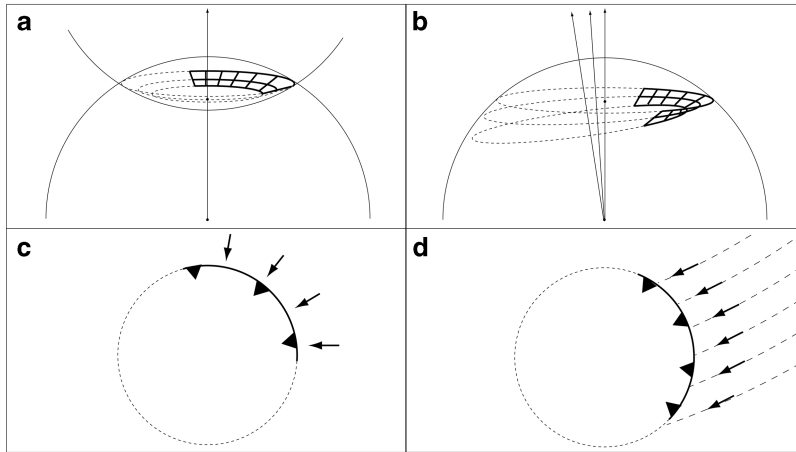
curvature has a maximum  $\kappa_1$  and a minimum  $\kappa_2$  along two orthogonal directions. They are known as the *principal curvatures*. The quantity  $K \equiv \kappa_1 \kappa_2$  is called *Gaussian curvature* and plays a key role in differential geometry. It can be used to classify points on  $S$  according to the value and sign of  $K$ . At the Earth's surface, which is biconvex, we have always that  $K$  is positive, because  $\kappa_1 \cong \kappa_2 \cong 1/R = 1/6373 \text{ km}^{-1}$ ,  $R$  being the Earth's radius, so that  $K = 1/R^2 = 2.462 \times 10^{-8} \text{ km}^{-2}$ . In Frank's theory, this quantity is conserved when the lithosphere bends to form the Wadati-Benioff zone, because we have that for a slab  $\kappa_1 \cong \kappa_2 \cong -1/R$ . Therefore, the slab acquires a *biconcave* shape, similar to the dent on a ping-pong ball (Bevis 1986). This is in agreement with an important theorem proved by Gauss in 1827, which states that  $K$  remains unchanged under continuous deformation of a flexible and inextensible surface. Finally, Frank (1968) proved that the simple bending of the lithosphere through an angle  $\alpha$  can only occur along a small circle having angular radius of curvature  $\beta = \alpha/2$ .

Despite its appeal and the capacity to explain the geometry of subduction zones, Frank's model has proved to be incorrect. During the 1970s, several authors recognized that Frank's formula, relating dip angle and radius of curvature, did not seem to be satisfied by the observed values of dip and trench curvature. The conclusion was that subducting slabs were subject to significant lateral stress as a consequence of the observed deviations from Frank's equilibrium formula (Strobach 1973; De Fazio 1974; Laravie 1975; Tovish and Schubert 1978). Another discrepancy was the prediction of Wadati-Benioff zones having concave downdip curvature  $-1/R$ , whereas most modern subduction zones show convex radial curvatures, associated with the downward bending of the lithosphere in the upper mantle and a general increase of the subduction angle at shallow depths. As pointed out by Bevis (1986), modern subducting slabs do not seem to conserve the Gaussian curvature of the unsubducted lithosphere, thereby the Earth's lithosphere cannot be considered as a flexible-inextensible shell and Frank's theorem is not applicable.

In a more recent study, Schettino and Tassi (2012) have argued that the classic "ping-pong ball" model of Frank (1968) is not compatible with the kinematics of subduction. First, the model is in contrast with Euler's theorem, according to which the relative motion between any two plates is a rotation about an axis (see Sect. 2.2). Figure 11.6a shows the concavo-concave geometry of a subducting slab in Frank's model. To obtain this configuration, subduction must have started at a point with an infinitely small island arc. Then, the process has continued by increasing both radius of curvature and width of the trench zone, in such a way that the lithosphere has been subducted radially from the initial point. In this instance, the Wadati-Benioff zone isodepth lines would have the shape of small circle arcs having a unique pole and length decreasing with the depth, as shown in Fig. 11.6a. Therefore, in a reference frame fixed to the overriding plate the relative velocity field of the subducting plate would be oriented radially with respect to the trench (Fig. 11.6c). Clearly, such a kinematics is not compatible with the plate tectonics paradigm, not only because it violates Euler's theorem, but also because it requires the presence of a non-rigid overriding plate, which extends radially in so far as subduction proceeds. In any case, the plate velocities predicted by this model are not compatible with those effectively observed along convergent plate boundaries.

The last major problem of Frank's model is represented by the very strong lateral deformation (extension or shortening) predicted when dip angle and radius of curvature assume values that deviate from an equilibrium combination. It can be easily proved, using the formulation of Strobach (1973), that the theoretical lateral strain associated with non-equilibrium values of dip and trench curvature is very high and attains a singularity at downdip distance  $\zeta = \rho / \sqrt{1 - \rho^2/R^2}$ ,  $\rho$  being the radius of curvature of the trench (in km).

An alternative model of subduction has been proposed recently by Schettino and Tassi (2012). As illustrated in Fig. 11.6b, the set of points placed along a subducting plate margin, which are entering the trench at an initial time  $t_0$  are



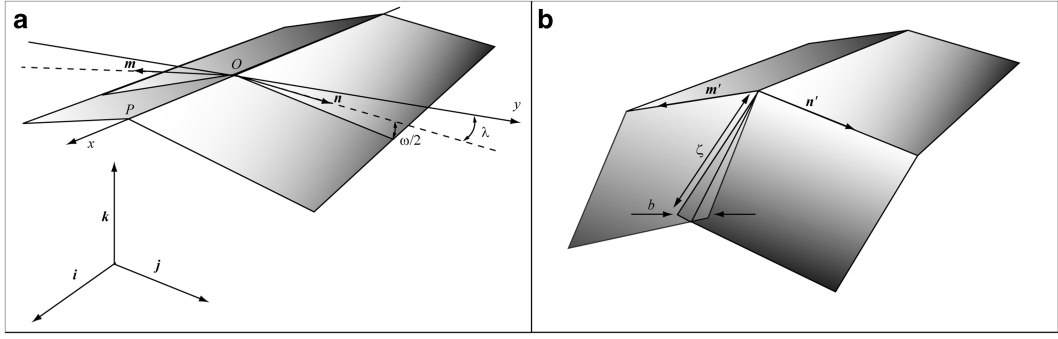
**Fig. 11.6** Alternative kinematic models of lithospheric bending at a subduction zone. (a) Classic “ping-pong ball” model (Frank 1968). During the subduction the trench curvature is progressively reduced and the trench length increases. Isodepth lines of the Wadati-Benioff zone are small circle arcs having a common pole, the slab geometry is concavo-concave, and the Gaussian curvature is conserved. (b) In the model proposed by Schettino and Tassi (2012), the radius of curvature of the trench does not

change through time. In this instance, the slab geometry is concavo-convex, thereby the Gaussian curvature is no longer invariant. The subducted trench lines are isodepth lines (*small circle arcs*) having constant curvature but different poles. (c) Predicted relative velocity field in the “ping-pong ball” model. (d) Typical velocity field along a subduction zone. *Dashed lines* represent flow lines about the Euler pole of relative motion (From Schettino and Tassi 2012)

aligned along a small circle arc and form an isodepth line having depth  $z = 0$  and length  $\ell$  at time  $t_0$ . In Frank’s model, such a small circle arc would sink vertically while conserving its length, thereby no lateral deformation would occur during the subduction in so far as the equilibrium condition is satisfied. In the model of Schettino and Tassi (2012), the points which entered the trench at time  $t_0$  also form isodepth lines at any time  $t > t_0$ , thereby the Wadati-Benioff zone can be still represented by isodepth lines that are small circle arcs. However, in this instance a condition of lateral mechanical equilibrium cannot exist at *any* time  $t$ , so that subduction *requires* some amount of lateral shortening and extension during bending. Furthermore, in conditions of lateral mechanical equilibrium, each isodepth line conserves at some depth its initial length  $\ell$  and curvature, but differently from Frank’s model has a continuously changing pole. Therefore, in the model of Schettino and Tassi (2012) the trench geometry can be conserved, at least in principle, for million years and the relative velocity field, which is shown in

Fig. 11.6d, is compatible with plate kinematics. An interesting feature of the Wadati-Benioff zone geometry predicted by Schettino and Tassi (2012) is that the *radial* curvature of the unsubducted lithosphere,  $\kappa_1 \cong 1/R$ , can be potentially conserved, when the subduction angle does not change substantially, although the Gaussian curvature of the subducting plate would be no longer conserved, because it would assume a negative value. Conversely, in Frank’s model the slab dip must decrease with depth and  $K$  is always conserved. Therefore, in the model of Schettino & Tassi a subducting plate cannot be considered as a flexible-inextensible spherical cap.

To understand the theoretical grounds of this model, let us start from the simple example of a planar lamina that has been bent along a hinge line coinciding with the  $x$  axis (Fig. 11.7a) to simulate the effect of the Earth’s curvature. If we bend the lamina again by rotating progressively the margin about the hinge versors  $\mathbf{n}'$  and  $\mathbf{m}'$  (Fig. 11.7b), it can be easily proved, for example through a sheet of paper, that such a new bending will cause lateral shortening, until the dip angle



**Fig. 11.7** Lateral deformation associated with the bending of a non-planar lamina along a segmented hinge line. (a) A planar lamina has been bent along the  $x$ -axis to simulate the effect of the Earth's curvature. Then, further bending is applied along a segmented hinge line. In the tangent plane, the segments of the hinge line forms an angle  $\lambda$  with the  $y$ -axis. (b) Slab bending is performed

through progressive rotation about the axes  $n'$  and  $m'$ . For small values of the bending angle  $\alpha$ , lateral shortening occurs along the segment  $OP$  after subduction, as the internal margin of the left and right sectors of the slab overlap by a quantity  $b$ , which increases with the distance  $\zeta$  from the hinge line

reaches an equilibrium value. It is easy to prove that such an equilibrium value of the bending angle is determined by Frank's formula.

The state of lateral shortening that occurs for small bending angles is represented in Fig. 11.7b by the overlap of the internal margins of the two slab components by a quantity  $b$ . If  $n$  and  $m$  are the versors describing the projection of the hinge line onto the tangent plane and  $R(q, \varphi)$  is the rotation matrix about an axis having versor  $q$  through an angle  $\varphi$  (positive counter-clockwise), then the axes of the hinge line are described by the versors:

$$n' = R(i, -\omega/2)n \quad ; \quad m' = R(i, +\omega/2)m \quad (11.22)$$

where  $i$  is the  $x$ -axis versor and  $\omega/2$  is the angle of bending of each side of the original lamina about the  $x$  axis. In order to determine the overlap  $b$ , we rotate the point  $P$ , having coordinates  $p \equiv (\zeta, 0, 0)$ , respectively by an angle  $\alpha$  about the axis  $n'$ , and by an angle  $-\alpha$  about the axis  $m'$ . The amount of overlap  $b$  will be given by:

$$\begin{aligned} b &= \{[R(n', \alpha) - R(m', -\alpha)]p\}_y \\ &= 2\zeta \left[ \sin \lambda \cos \lambda \cos \frac{\omega}{2} (1 - \cos \alpha) \right. \\ &\quad \left. - \cos \lambda \sin \frac{\omega}{2} \sin \alpha \right] \end{aligned} \quad (11.23)$$

where we have used expression (2.18) for the components of  $R$ .

Overlap occurs for  $b < 0$ , whereas for  $b = 0$  we get the equilibrium condition of Frank (1968). Finally, for  $b > 0$  we obtain lateral extension along the internal margin. This result implies that Frank's formula correctly describes the states of lateral mechanical equilibrium during the bending. The angle  $\lambda$  in (11.23) must be taken in the tangent plane (i.e., the plane of  $n$  and  $m$ ). The relationship of this quantity with the corresponding angle,  $\lambda'$ , in the plane of  $n'$  and  $m'$  is:

$$\sin^2 \lambda' = \sin^2 \lambda \cos^2 \frac{\omega}{2} + \sin^2 \frac{\omega}{2} \quad (11.24)$$

The application of these expressions to the case of the Earth's subduction zones can be made taking the limit as  $\lambda \rightarrow 0$  and  $\omega \rightarrow 0$  and considering that the variation  $d\omega$  per unit arc length  $\ell$  must be equal to the Earth's curvature:  $\kappa_e = 1/R = d\omega/d\ell$ . Let  $\kappa_T = d(2\lambda)/d\ell = 2d\lambda/d\ell$  and  $\kappa = 1/\rho = 2d\lambda'/d\ell$  be respectively the trench curvatures in the tangent plane and in the plane of  $n'$  and  $m'$ . From (11.24) we easily obtain the relation between these quantities:

$$\kappa_T = \left( \kappa - \frac{1}{R^2} \right)^{1/2} = \left( \frac{1}{\rho^2} - \frac{1}{R^2} \right)^{1/2} \quad (11.25)$$

Therefore, if we take the limit as  $\lambda \rightarrow 0$  and  $\omega \rightarrow 0$  in the expression (11.23) for  $b$ , we obtain an expression for the *lateral strain*  $\varepsilon = db/d\ell$  at downdip distance  $\zeta$  from the hinge line:

$$\varepsilon = \zeta \left[ (1 - \cos \alpha) \left( \frac{1}{\rho^2} - \frac{1}{R^2} \right)^{1/2} - \frac{1}{R} \sin \alpha \right] \quad (11.26)$$

For  $\rho = R \sin \alpha / 2$  (Frank's formula), the lateral strain vanishes. Therefore, although the model of Frank (1968) does not correctly describe the subduction process, the corresponding relation between subduction angle and trench curvature holds as a condition of lateral mechanical equilibrium at any depth. Expression (11.26) holds when the lithosphere bends along a single hinge line and differs substantially from the formula proposed by Strobach (1973). It shows that the lateral strain state increases linearly with the downdip distance  $\zeta$ . From (11.26) we easily obtain the strain rate,  $\dot{\varepsilon}$ , which is independent from the downdip distance.

If  $v$  is the subduction velocity, then:

$$\dot{\varepsilon} = v \left[ (1 - \cos \alpha) \left( \frac{1}{\rho^2} - \frac{1}{R^2} \right)^{1/2} - \frac{1}{R} \sin \alpha \right] \quad (11.27)$$

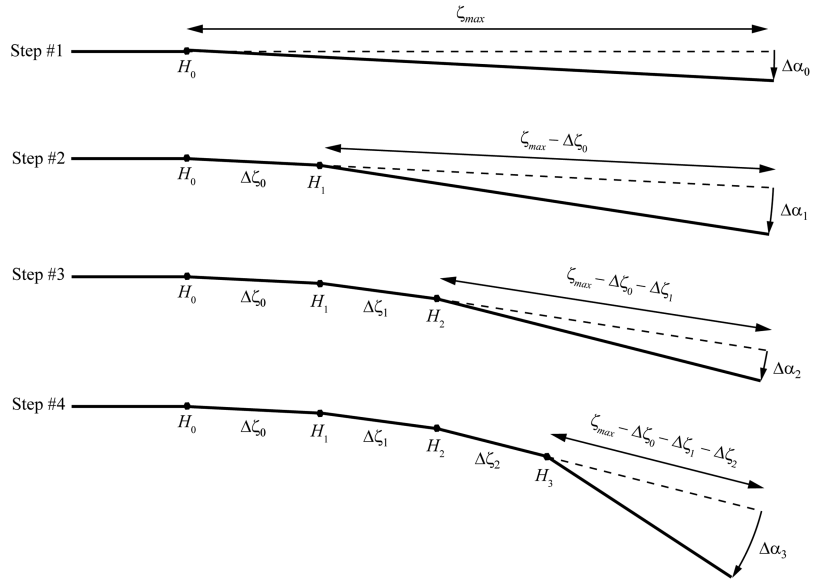
The previous formulation started from the simple bending of a lamina whose initial geometry simulates the effect of the Earth's curvature along one direction only, because the radial curvature is assumed to be zero in the starting state (the lamina is flat in the  $x$  direction). In order to estimate the effect of the radial curvature  $\kappa_e = 1/R$  of a tectonic plate, we note that this is equivalent to assume an initial radial bending of the lamina by a small angle  $\delta\alpha$ , which is not accompanied by lateral strain. However, for a small patch of lithosphere having width  $\zeta = 1$  km we would have:  $\delta\alpha = 180/(\pi R) = 0.008993^\circ$ , which is effectively a negligible quantity. Therefore, the

approximation used to obtain (11.26) and (11.27) is acceptable.

More problematic is the unrealistic hypothesis that the rate of lithospheric bending along the trench zone is compensated at shallow depths by an equal rate of lithospheric unbending, so that the subduction angle remains approximately constant along the Wadati-Benioff zone. Real subduction zones exhibit a more complex behaviour, characterized by a general increase of the dip angle  $\alpha$  and eventually by zones of upward bending, especially when an old slab reaches the 670 km discontinuity (e.g., Goes et al. 2008). Considering a vertical cross-section, we can ideally reproduce the real slab geometry through a sequence of small rotations about horizontal hinge lines. The method is illustrated in Fig. 11.8. For a slab having total downdip length  $\zeta_{max}$ , we start with a small rotation of a straight segment having length  $\zeta_{max}$  about the surface hinge line by an angle  $\Delta\alpha_0$ . At the next step, we rotate a straight segment having length  $\zeta_{max} - \Delta\zeta_0$  about a hinge line at downdip distance  $\Delta\zeta_0$  by an angle  $\Delta\alpha_1$ . Then, the procedure is repeated iteratively. If the slab geometry has been correctly reproduced as far as downdip distance  $\zeta$ , then the next step is to rotate by a small angle  $\Delta\alpha(\zeta)$  a segment having length  $\zeta_{max} - \zeta$  about a hinge line at downdip distance  $\zeta$ . At each step, the incremental strain can be calculated by differentiating (11.26). Therefore, we can generalize (11.26) and (11.27) to an arbitrary downdip geometry of the Wadati-Benioff zone. Let  $x$  be the horizontal offset from the surface hinge line and  $\alpha = \alpha(x)$  a dip function. Then, generalized formulae for the finite lateral strain and the lateral strain rate at offset  $x$  are:

$$\begin{aligned} \varepsilon(x) &= \int_0^x [\zeta(x) - \zeta(x')] \\ &\times \left[ \left( \frac{1}{\rho^2} - \frac{1}{R^2} \right)^{1/2} \sin \alpha(x') - \frac{1}{R} \cos \alpha(x') \right] d\alpha(x') \end{aligned} \quad (11.28)$$

**Fig. 11.8** Conceptual method of construction of down-dip slab geometries. At step #1, a line having length  $\zeta_{max}$  is rotated by angle  $\Delta\alpha_0$  about a hinge  $H_0$  at depth zero. Then, at the next step, a line segment having length  $\zeta_{max} - \Delta\zeta_0$  is rotated by angle  $\Delta\alpha_1$  about  $H_1$ . At step # $n$ , a line segment having length  $\zeta_{max} - \sum_{k=0}^{n-1} \Delta\zeta_k$  is rotated by an angle  $\Delta\alpha_n$

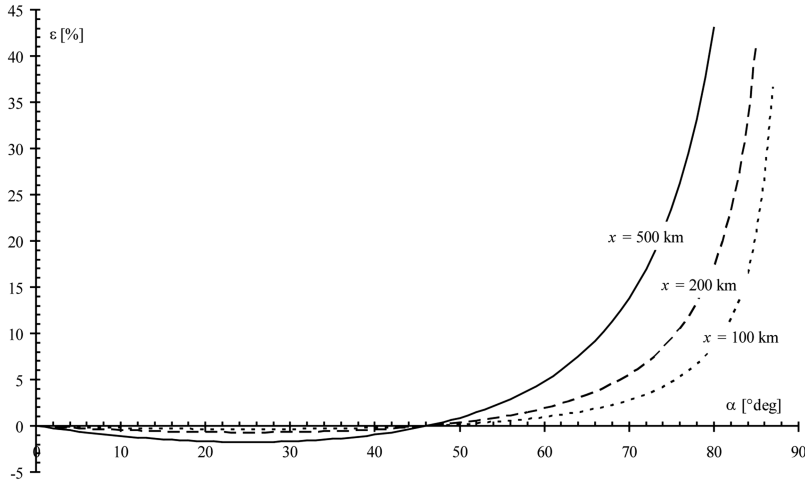


$$\begin{aligned}
 \dot{\epsilon}(x) &= v \int_0^x \left[ \left( \frac{1}{\rho^2} - \frac{1}{R^2} \right)^{1/2} \sin \alpha(x') \right. \\
 &\quad \left. - \frac{1}{R} \cos \alpha(x') \right] d\alpha(x') \\
 &= v \left[ (1 - \cos \alpha(x)) \left( \frac{1}{\rho^2} - \frac{1}{R^2} \right)^{1/2} \right. \\
 &\quad \left. - \frac{1}{R} \sin \alpha(x) \right]
 \end{aligned}
 \quad (11.29)$$

Expression (11.29) is particularly interesting, because it shows that the solution (11.27) is valid also when the slab dip changes with depth. In general, the lateral strain rate  $\dot{\epsilon}$  at some depth  $z$ , or offset  $x$ , will depend only from the slab dip  $\alpha$  at the same depth (or at the corresponding horizontal offset). Plots of the lateral strain as a function of the bending angle,  $\alpha$ , are illustrated in Fig. 11.9 for different horizontal offsets  $x$  from the hinge line. These curves have been traced in the simple case of a constant bending angle ( $\alpha(x) = \text{const}$ ) through (11.26). They show that a small component of compressional strain at shallow depths is always present during the bending of a lithospheric slab. However, the typical geometry of many Wadati-Benioff zones presents a dip angle increasing progressively from the bulge to

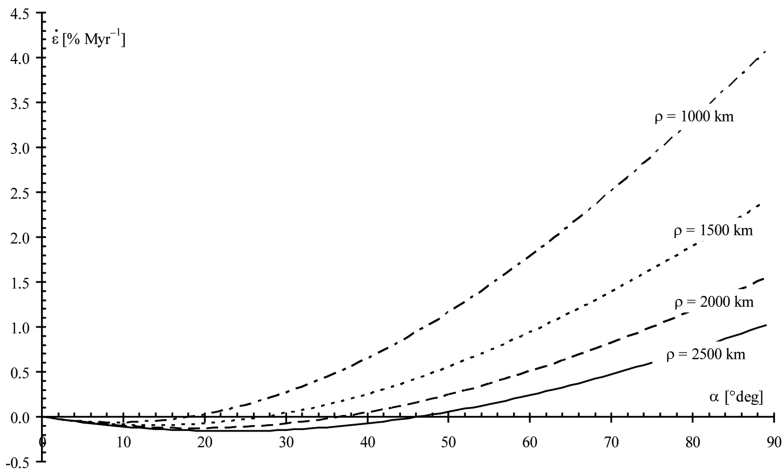
intermediate depths (300–400 km), where values close to Frank's equilibrium condition are often reached.

Expression (11.29) can be used to plot the theoretical lateral strain rate as a function of the slab dip. Such curves are more readily comparable with the observed data along vertical cross sections. Figure 11.10 illustrates the dependence of the lateral strain rate from the dip angle for several trench curvatures. We note that the range of angles for which lateral shortening is predicted is reduced when the radius of curvature decreases. These results imply that bending of slabs is always accompanied by lateral deformation that superimposes on down-dip strain. In particular, slabs are always in a state of lateral shortening at shallow depths, and they eventually go in a state of lateral extension at intermediate depth, especially when the radius of curvature  $\rho$  is small and the dip angle  $\alpha$  is large. Therefore, we expect that subducting lithospheres like that which is sinking beneath the Mariana Trench ( $\rho = 738$  km) in the western Pacific, or that which is subducting beneath the South Sandwich Trench ( $\rho = 346$  km) in the South Atlantic, are subject to large lateral extension at intermediate depths. Schettino and Tassi (2012) showed that this is exactly what is observed on the basis of the Wadati-Benioff seismicity.



**Fig. 11.9** Lateral strain as a function of the bending angle at different distances  $x$  from the trench. The three plots show that the lateral strain is shortening at any offset  $x$  for small bending angles, whereas large extensional strain is predicted when the bending angle exceeds the equilibrium

value. The curves have been built assuming a curvature radius  $\rho = 2,500$  km. Note that the range of subduction angles for which lateral shortening (or extension) occurs is independent from the horizontal offset from the trench (After Schettino and Tassi 2012)



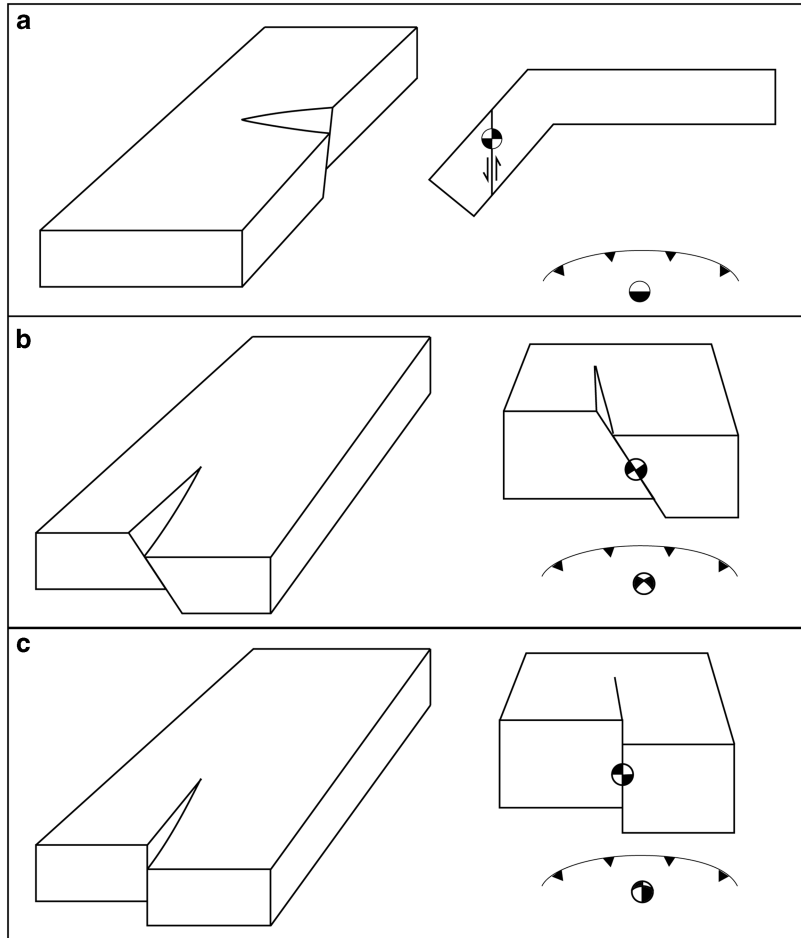
**Fig. 11.10** Lateral strain rate as a function of the bending angle for different trench curvatures. These plots show that the lateral strain rate has a small component of shortening for small bending angles, whereas large extensional strain

rates (up to  $\sim 4$  %Myr $^{-1}$ ) are predicted when the bending angle exceeds the equilibrium value. The curves have been built assuming a convergence velocity  $v = 50$  mm/year (After Schettino and Tassi 2012)

Intraslab seismicity has been widely used to study geometry, pattern of deformation, and state of stress of slabs. After the classic papers of Isacks et al. (1968) and Isacks and Molnar (1971) mentioned above, Bevis and Isacks (1984) performed an analysis of the geometry of Wadati-Benioff zones through a determination of trend surfaces fitting earthquake hypocenters. The existence of a large number of events that could

not be associated with downdip deformation was also observed by Apperson and Frohlich (1987), while Bevis (1988) estimated an average downdip strain rate of  $\sim 10^{-15}$  s $^{-1}$  ( $\sim 0.032$  %/Myr) using Brune's formula (11.2). Nothard et al. (1996), in a study of the deformation of the Tonga slab based on Kostrov's formula (11.4), suggested that the downdip and lateral strain rates were of the order of  $\sim 10^{-16} - 10^{-15}$  s $^{-1}$ , one-two

**Fig. 11.11** Examples of seismic deformation along Wadati-Benioff zones. (a) Downdip extension; (b) Lateral extension; (c) Vertical tearing



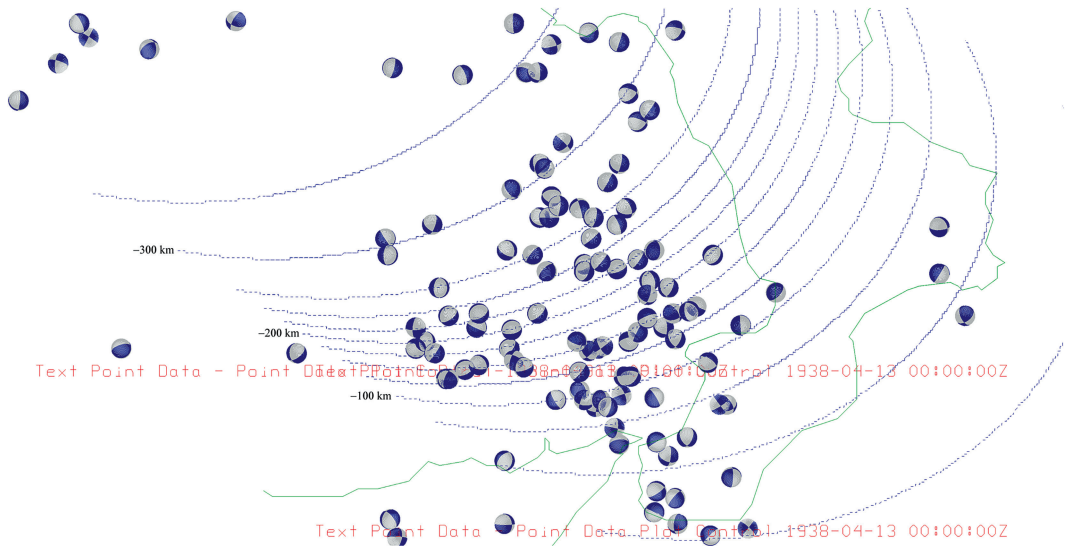
orders of magnitude smaller than the Bevis (1988) estimate. More recently, Chen et al. (2004), in a global study of the state of stress associated with intermediate-depth earthquakes, found significant lateral deformation everywhere in the Mariana and South Sandwich slabs, and in part of the Aleutian and Tonga subducting lithosphere. Finally, in their study about the relation between trench curvature and lateral deformation of the slabs, Schettino and Tassi applied Molnar's formula (11.17) to the events included in the Harvard CMT catalogue (1976 through 2009) of eight slabs. They found very large strain rates associated with lateral extension for the South Sandwich and Mariana slabs, and moderate lateral extension in the case of the Tonga and Japan slabs.

Conversely, Costa Rica, N. Kurile, Sumatra, and Kermadec were found to be in a state of

small lateral shortening above 400 km. On average, Schettino and Tassi (2012) calculated downdip and lateral strain rates of the order of  $\sim 10^{-14} \text{ s}^{-1}$ , one order of magnitude larger than Bevis' estimate. Three common kinds of intraslab deformation at intermediate depth, and the corresponding focal mechanisms, are illustrated in Fig. 11.11. Conversely, lateral deformation is generally shortening at shallow depth and in the transition zone, while downdip deformation is almost always shortening below  $\sim 300 \text{ km}$ .

## Problems

1. The figure below shows focal mechanisms along the Wadati-Benioff zone of the Ionian slab in central Mediterranean. Curved lines represent isodepth lines of the slab. Classify



- each focal mechanism and discuss the style of deformation of this slab;
- Download the MS Excel file CMT7609.xlsx, containing the global CMT catalog 1976–2009. Select underthrusting earthquakes (depth  $\leq 40$  km) from the Hellenic Trench between ( $38^{\circ}\text{N}$ ,  $20^{\circ}\text{E}$ ) and ( $34.6^{\circ}\text{N}$ ,  $24^{\circ}\text{N}$ ). Determine the average velocity of convergence between Africa and Greece using Brune's formula. Compare with the velocity of convergence Africa – Eurasia in the same area estimated by MORVEL (see Chap. 2);
  - Write a computer program to calculate the strain rate tensor from a set of moment tensors using Kostrov's formula (11.4);
  - Write a computer program to calculate the strain rate tensor from a set of asymmetric moment tensors using formula (11.14);
  - Perform a study of seismic deformation of Turkey using Kostrov's formula and the asymmetric moment tensor approach. Discuss the differences between the two approaches.

## References

- Apperson KD, Frohlich C (1987) The relationship between Wadati-Benioff Zone geometry and  $P$ ,  $T$  and  $B$  axes of intermediate and deep focus earthquakes. *J Geophys Res* 92(B13):13821–13831. doi:10.1029/JB092iB13p13821
- Bevis M (1986) The curvature of Wadati-Benioff zones and the torsional rigidity of subducting plates. *Nature* 323:52–53
- Bevis M (1988) Seismic slip and down-dip strain rates in Wadati-Benioff zones. *Science* 240:1317–1319
- Bevis M, Isacks BL (1984) Hypocentral trend surface analysis: probing the geometry of Benioff zones. *J Geophys Res* 89(B7):6153–6170
- Brune JN (1968) Seismic moment, seismicity, and rate of slip along major fault zones. *J Geophys Res* 73(2): 777–784. doi:10.1029/JB073i002p00777
- Chen P-F, Bina CR, Okal EA (2004) A global survey of stress orientations in subducting slabs as revealed by intermediate-depth earthquakes. *Geophys J Int* 159:721–733
- Corredor F (2003) Seismic strain rates and distributed continental deformation in the northern Andes and three-dimensional seismotectonics of northwestern South America. *Tectonophysics* 372(3):147–166
- De Fazio TL (1974) Island-arc and underthrust-plate geometry. *Tectonophysics* 23:149–154
- England P, Molnar P (1997) Active deformation of Asia: from kinematics to dynamics. *Science* 278(5338):647–650
- Frank FC (1968) Curvature of island arcs. *Nature* 220:363
- Goes S, Capitanio FA, Morra G (2008) Evidence of lower-mantle slab penetration phases in plate motions. *Nature* 451:981–984
- Isacks B, Molnar P (1971) Distribution of stresses in the descending lithosphere from a global survey of focal-mechanism solutions of mantle earthquakes. *Rev Geophys* 9(1):103–174
- Isacks B, Oliver J, Sykes LR (1968) Seismology and the new global tectonics. *J Geophys Res* 73(18):5855–5899. doi:10.1029/JB073i018p05855
- Kagan YY (1991) Seismic moment distribution. *Geophys J Int* 106:123–134

- Kagan YY (1993) Statistics of characteristic earthquakes. *Bull Seism Soc Am* 83(1):7–24
- Kagan YY (2010) Earthquake size distribution: power-law with exponent  $\beta \equiv 1/2$ ? *Tectonophysics* 490:103–114
- Kiratzí AA, Papazachos CB (1995) Active deformation on the shallow part of the subducting lithospheric slab in the southern Aegean. *J Geodynamics* 19(1):65–78
- Kiratzí AA, Papazachos CB (1996) Moment-tensor summation to derive the active crustal deformation in Japan. *Bull Seismol Soc Am* 86(3):821–831
- Kostrov VV (1974) Seismic moment and energy of earthquakes, and seismic flow of rock. *Izv Acad Sci USSR Phys Solid Earth* 1:23–44
- Kreemer C, Haines J, Holt WE, Blewitt G, Lavalée D (2000) On the determination of a global strain rate model. *Earth Planets Space* 52:765–770
- Kreemer C, Holt WE, Haines AJ (2003) An integrated global model of present-day plate motions and plate boundary deformation. *Geophys J Int* 154(1):8–34. doi:10.1046/j.1365-246X.2003.01917.x
- Laravie JA (1975) Geometry and lateral strain of subducted plates in island arcs. *Geology* 3(9):484–486
- Lay T, Kanamori H, Ammon CJ, Nettles M, Ward SN, Aster RC, Beck SL, Bilek SL, Brudzinski MR, Butler R, DeShon HR, Ekström G, Satake K, Sipkin S (2005) The great Sumatra-Andaman earthquake of 26 December 2004. *Science* 308(5725):1127–1133
- Molnar P (1979) Earthquake recurrence intervals and plate tectonics. *Bull Seismol Soc Am* 69(1):115–133
- Molnar P (1983) Average regional strain due to slip on numerous faults of different orientations. *J Geophys Res* 88(B8):6430–6432
- Nothard S, Haines J, Jackson J, Holt B (1996) Distributed deformation in the subducting lithosphere at Tonga. *Geophys J Int* 127(2):328–338
- Oliver J, Isacks B (1967) Deep earthquake zones, anomalous structures in the upper mantle, and the lithosphere. *J Geophys Res* 72(16):4259–4275
- Papazachos CB, Kiratzí AA (1992) A formulation for reliable estimation of active crustal deformation and its application to central Greece. *Geophys J Int* 111(3):424–432
- Schettino A, Tassi L (2012) Trench curvature and deformation of the subducting lithosphere. *Geophys J Int* 188(1):18–34. doi:10.1111/j.1365-246X.2011.05262.x
- Strobach K (1973) Curvature of island arcs and plate tectonics. *Z Geophys* 39:819–831
- Tovish A, Schubert G (1978) Island arc curvature, velocity of convergence and angle of subduction. *Geophys Res Lett* 5(5):329–332
- Ward SN (1998) On the consistency of earthquake moment rates, geological fault data, and space geodetic strain: the United States. *Geophys J Int* 134(1):172–186

---

## Abstract

In this chapter, I present the important theme of heat conduction across and within the continental and oceanic lithospheres. In particular, the non-steady state heat conduction equation is solved and applied to the cooling of the oceanic lithosphere. The chapter also describes the main forces driving plate tectonics: ridge push and slab pull.

---

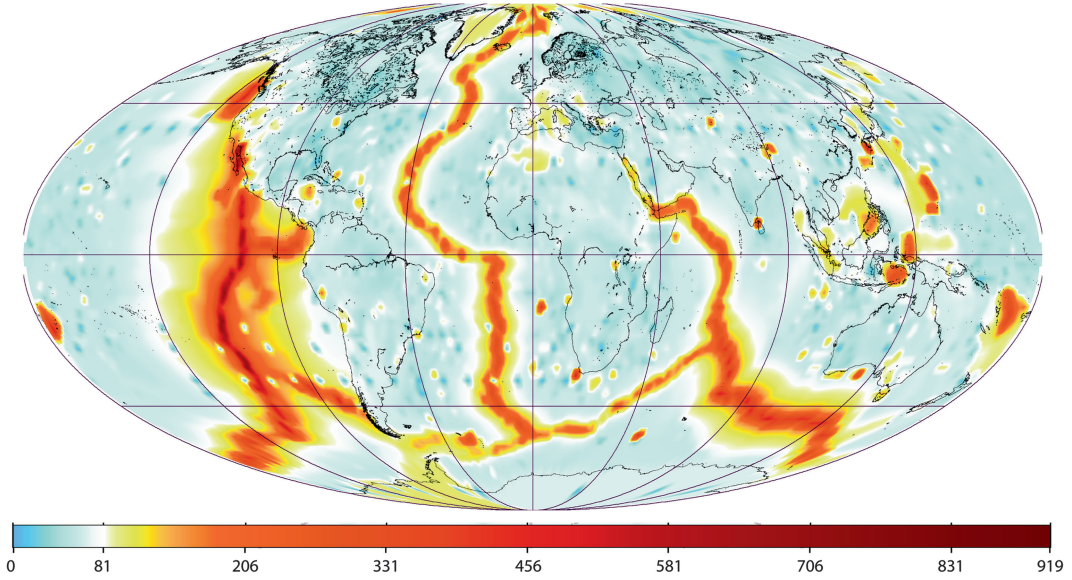
## 12.1 Fourier's Law

In this chapter, we are going to study the *thermal structure* of the lithosphere, which is strongly related to its rheology and dynamic behavior. By “thermal structure”, we mean the scalar field of temperatures,  $T = T(\mathbf{r}, t)$ , and its temporal evolution. This field depends in turn by both the rate of heat transfer through the Earth's interior and the heat loss at the Earth's surface. It is known that heat can be transferred by conduction, convection, or radiation. *Conduction* essentially results from micro-scale interaction between energy carriers within a material. The nature of these carriers depends from the state of the matter and from the material structure, so that they can be individual molecules in fluids, electrons or phonons in solids. In any case, conduction consists into a direct energy transfer from more energetic carriers to low-energy carriers through molecular or particle collisions, thereby is a diffusive process that requires spatial variability of the temperature field. *Convection* is a process that changes the temperature field by the large-scale

motion of energy carriers. In this instance, heat is conveyed from one region to another together with the carriers, rather than being transferred between energy carriers. Clearly, convection and conduction may operate at the same time within a fluid. Finally, heat can be transported by electromagnetic *radiation*. This is a process that occurs at the top of the Earth's atmosphere, thereby it will not be investigated in this book.

The distribution of temperatures in the continental and oceanic lithosphere is largely controlled by the conductive loss of heat at the Earth's surface, although convective heat transport by water circulating through the oceanic basalts or intrusive igneous bodies may be locally an important mechanism of cooling for these rocks. This heat originates both by the secular cooling of the Earth's hot interior and by the radioactive decay of some elements that are present in crustal and mantle rocks.

The fundamental equation describing the conductive heat transport is known as *Fourier's law*. If we define the *heat flux*  $q(\mathbf{n})$  as the quantity of heat per unit area and per unit time that flows by conductive transport through a small



**Fig. 12.1** Global map of Earth surface heat flow in  $\text{mWm}^{-2}$ , based on the data of Davies (2013)

surface element having normal vector  $\mathbf{n}$ , then the isotropic Fourier's law reads:

$$\mathbf{q}(\mathbf{n}) = -k\nabla T \cdot \mathbf{n} \quad (12.1)$$

where  $k$ , the *coefficient of thermal conductivity*, has units  $\text{Wm}^{-1} \text{K}^{-1}$ . Therefore, the heat flux through a surface element  $d\mathbf{S} = \mathbf{n}dS$  is proportional to the directional derivative of the temperature field along the direction  $\mathbf{n}$ . Clearly, changing the orientation of  $d\mathbf{S}$  will also change the heat flux through  $d\mathbf{S}$ . Consequently, Fourier's law agrees with the quite intuitive concept that the *maximum* heat flow must occur in the direction of maximum decrease of the temperature field, which is the direction of  $-\nabla T$ . To this purpose, it is useful to define a vector field  $\mathbf{q} = \mathbf{q}(\mathbf{r}, t)$  that is at any point  $\mathbf{r}$  orthogonal to the isotherm surface passing through  $\mathbf{r}$  at time  $t$ . By the properties of the gradient operator (see Appendix 1), we can define the vector field of maximum heat flux as follows:

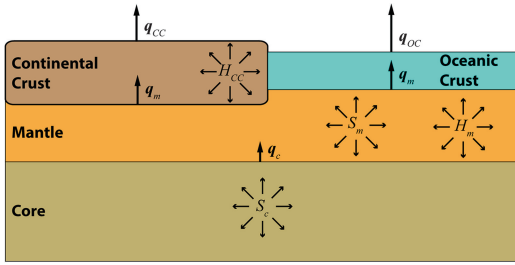
$$\mathbf{q} = -k\nabla T \quad (12.2)$$

Although this equation in many textbooks is indicated as "Fourier's law", it represents only a

corollary of the complete law of heat conduction (12.1). Combining the two equations gives:

$$\mathbf{q}(\mathbf{n}) = \mathbf{q} \cdot \mathbf{n} \quad (12.3)$$

Measurements of the vertical heat flow through the Earth's surface are performed lowering *thermistors*, which are thermally sensitive semiconductor resistors, down drill holes. The average heat flux provides essential information about the quantity of heat that was generated in the Earth's interior and the temperature field within the lithosphere. Figure 12.1 illustrates a recent global compilation of heat flux data, which is based on  $\sim 38,000$  measurements (Davies 2013). The average value of  $q$  on the continents is relatively low in N. America, Fennoscandia, and East Europe ( $33\text{--}40 \text{ mWm}^{-2}$ ), intermediate in Africa ( $\sim 52 \text{ mWm}^{-2}$ ), and quite large in Brazil and Australia ( $65\text{--}68 \text{ mWm}^{-2}$ ). Such a variability is caused by regional changes in radioactive heat production. In the oceanic regions, the average heat flux is greater than  $\sim 100 \text{ mWm}^{-2}$  when the lithosphere has an age less than  $\sim 10 \text{ Ma}$  and rapidly decreases in so far as the age attains  $30 \text{ Ma}$  (Stein and Stein 1992).



**Fig. 12.2** Sketch illustrating the sources of heat in the Earth and the components of the global heat flow.  $q_{CC}$  and  $q_{OC}$  are respectively the surface heat flux from continental and oceanic regions,  $q_m$  is the heat flux through the Moho, and  $q_c$  is the heat flux through the CMB. Four important heat sources exist in the Earth: radiogenic decay in the continental crust ( $H_{CC}$ ), radiogenic decay in the mantle ( $H_m$ ), secular cooling of the mantle ( $S_m$ ), and secular cooling of the core ( $S_c$ )

Then,  $q \sim 60 \text{ mWm}^{-2}$  for older lithosphere and decreases smoothly to about  $50 \text{ mWm}^{-2}$  in the oldest lithosphere.

The typical thermal conductivity of near-surface rocks depends from rock type, composition, grain size, grain orientation, density, porosity, composition of pore fluid, and temperature. In the case of sedimentary and volcanic rocks, the main controlling factor is the porosity, while the thermal conductivity of plutonic and metamorphic rocks depends from the dominant mineral phase (Clauser and Huenges 1995). In general, most sedimentary rocks have values of  $k$  between  $0.5$  and  $2.5 \text{ Wm}^{-1} \text{ K}^{-1}$ , while for the majority of volcanic and plutonic rocks  $k$  ranges from  $1.5$  to  $3.5 \text{ Wm}^{-1} \text{ K}^{-1}$ . Finally, in the case of metamorphic rocks, the thermal conductivity is generally between  $2$  and  $4 \text{ Wm}^{-1} \text{ K}^{-1}$  when the quartz content is low, and between  $5$  and  $6 \text{ Wm}^{-1} \text{ K}^{-1}$  in the case of quartzite. Assuming  $k = 2.5 \text{ Wm}^{-1} \text{ K}^{-1}$ , we have an average vertical temperature gradient at the Earth's surface of  $\sim 16 \text{ Kkm}^{-1}$  when  $q = 40 \text{ mWm}^{-2}$  and  $\sim 24 \text{ Kkm}^{-1}$  when  $q = 60 \text{ mWm}^{-2}$ .

A consistent part of the heat flow at the Earth's surface undoubtedly originates by the radioactive decay of  $^{40}\text{K}$ ,  $^{235}\text{U}$ ,  $^{238}\text{U}$ , and  $^{232}\text{Th}$  in the mantle and, to a lesser extent, in the continental crust (Fig. 12.2). The remaining part of the surface heat

flow must come from cooling of the Earth over the geological time. It is important to note that the average rate of production of radiogenic heat in the crust (including the oceanic crust), which is  $H_C = 2.9 \times 10^{-10} \text{ Wkg}^{-1}$ , is much greater than that of the mantle, which is estimated to be  $H_m = 5.1 \times 10^{-12} \text{ Wkg}^{-1}$ . This larger productivity is overcome by a mantle to crust mass ratio that is  $\sim 143$ , thereby the total crustal radioactivity results to be  $8.2 \times 10^{12} \text{ W}$ , with respect to a total mantle production of  $20.0 \times 10^{12} \text{ W}$  (Stacey and Davis 2008). As mentioned above, radiogenic decay is not the unique source of heat in the Earth's interior. Another important source is represented by the basal heating of the lithosphere along the LAB, associated with mantle convection, which also contributes to the heat flux  $q_m$  through the Moho (Fig. 12.2). Turcotte and Schubert (2002) estimated that  $\sim 75\text{--}80\%$  of the present-day surface heat flow should be attributed to decay of radioactive isotopes, while about  $20\text{--}25\%$  would originate from the *secular cooling* of the Earth. This is the continuous loss of primordial heat stored in the Earth's mantle and core studied by Kelvin (1864), which drives mantle convection. The relative contribution of secular cooling and radioactive decay to the total heat budget estimated by Turcotte and Schubert (2002) is controversial and other authors give very different values of the relative importance of these sources of heat (e.g., Korenaga 2003, 2008). The *Urey ratio* is a quantity that is commonly used to measure the relative importance of the radiogenic heat generated in the Earth's crust and mantle. It is defined as the ratio of internal heat production to surface heat flux. This is a key parameter characterizing the global thermal budget and a strong constraint for both thermal history and mantle convection models. Calling  $q(H_{CC})$ ,  $q(H_m)$ ,  $q(S_m)$ , and  $q(S_c)$  the components of surface heat flux associated with each of the sources illustrated in Fig. 12.2, we have the following budget relations:

$$\begin{cases} q_{CC} = q_m + q(H_{CC}) \\ q_{OC} = q_m \\ q_m = q(H_m) + q(S_m) + q_c \\ q_c = q(S_c) \end{cases} \quad (12.4)$$

where the production of radiogenic heat in the oceanic crust and in the core has been neglected. The *bulk Earth Urey ratio* measures the relative contribution of internal heat generation with respect to the total surface heat flux:

$$U_b = \frac{q(H_{CC}) + q(H_m)}{q_{CC} + q_{OC}} \quad (12.5)$$

A quantity that is more commonly used by geophysicists in studies about the thermal history of Earth is the *convective Urey ratio*, which Korenaga (2008) defines as the ratio of heat generation in the mantle over the mantle heat flux:

$$U_c = \frac{q(H_m)}{q_m} \quad (12.6)$$

Finally, it is possible to introduce an *internal heating ratio*,  $I$ , as the ratio of surface heat flow associated with mantle sources to the mantle heat flux:

$$I = \frac{q_m - q_c}{q_m} \quad (12.7)$$

Korenaga (2008) pointed out that while the  $U_b$  is probably  $\sim 0.35$ , the convective Urey ratio  $U_c$  is estimated to be  $\sim 0.2$ . Therefore, according to this author only  $\sim 20\%$  of the basal lithospheric heating associated with mantle convection would originate from radioactive decay, while  $\sim 80\%$  of  $q_m$  would come from secular cooling. In the next section, we shall determine the temperature field in the continental crust starting from the surface heat flow data and from an estimate of the radiogenic heat produced in the Earth's interior.

## 12.2 Continental Geotherms

Let us consider the temperature distribution in a homogeneous region  $\mathbf{R}$ , bounded by the surface  $S(\mathbf{R})$ . According to the form (12.2) of Fourier's law, the maximum heat flux  $\mathbf{q}$  is a potential field and the temperature itself is the associated potential. Assuming that there are no sources of heat in  $\mathbf{R}$ , the total free flux of heat through  $S(\mathbf{R})$  will be given by:

$$\begin{aligned} \Phi(S) &= \oint_{S(\mathbf{R})} \mathbf{q}(\mathbf{n}) \cdot d\mathbf{S} = \oint_{S(\mathbf{R})} \mathbf{q} \cdot d\mathbf{S} \\ &= -k \oint_{S(\mathbf{R})} \nabla T \cdot d\mathbf{S} \end{aligned} \quad (12.8)$$

At any time, the total heat,  $Q$ , in  $\mathbf{R}$  is given by:

$$Q(t) = c\rho \int_{\mathbf{R}} T(\mathbf{r}, t) dV \quad (12.9)$$

where  $c$  is the specific heat and  $\rho$  is the density in  $\mathbf{R}$ . Clearly, the rate of variation of  $Q$  must coincide with the net *incoming* flow through  $S(\mathbf{R})$ , so that applying the divergence theorem (see Appendix 1) and reversing the sign we obtain:

$$\dot{Q}(t) = k \oint_{S(\mathbf{R})} \nabla T \cdot d\mathbf{S} = k \int_{\mathbf{R}} \nabla^2 T dV \quad (12.10)$$

By (12.9), the rate of variation of  $Q$  can be also written as follows:

$$\dot{Q}(t) = c\rho \int_{\mathbf{R}} \frac{\partial T}{\partial t} dV \quad (12.11)$$

Therefore, combining this expression with (12.10) provides:

$$\int_{\mathbf{R}} \left( c\rho \frac{\partial T}{\partial t} - k \nabla^2 T \right) dV = 0 \quad (12.12)$$

The fact that this integral vanishes for any choice of the region  $\mathbf{R}$  implies that the integrand itself must be zero throughout  $\mathbf{R}$ . Therefore, introducing the *thermal diffusivity*  $\kappa \equiv k/c\rho$  [ $\text{m}^2\text{s}^{-1}$ ], we obtain the following *equation of conductive heat transfer* (or *heat diffusion equation*):

$$\frac{\partial T}{\partial t} - \kappa \nabla^2 T = 0 \quad (12.13)$$

This equation holds, in the present form, in the hypothesis that the thermal conductivity  $k$  and

the diffusivity  $\kappa$  are independent from  $T$ , which is only approximately true (e.g., Hofmeister 1999). Therefore, more general equations of conductive heat transfer can be written. In steady-state conditions, the temperature does not change with time, so that (12.13) assumes the form:

$$\nabla^2 T = 0 \quad (12.14)$$

Consequently, in steady-state conditions and in absence of sources the temperature field satisfies Laplace's equation in  $\mathbf{R}$ , so that it is harmonic. Accordingly, it will not have neither maxima nor minima in  $\mathbf{R} - S(\mathbf{R})$ . For example, in a problem where the temperature depends only from the depth  $z$ , we would have  $dT/dz = \text{const}$ , so that  $T$  would change linearly with depth. In presence of heat sources, the diffusion Eq. (12.13) must be generalized to take into account of the local rate of heat production. Let  $H = H(\mathbf{r}, t)$  be the rate of heat generation per unit mass at location  $\mathbf{r}$  and time  $t$ . In this instance, the heat diffusion equation assumes the following more general form:

$$\frac{\partial T}{\partial t} - \kappa \nabla^2 T - \frac{H}{c} = 0 \quad (12.15)$$

Therefore, even in steady-state conditions, the presence of heat sources implies that a temperature field depending only from  $z$  cannot increase linearly with the depth. We can easily apply these results to the simple case of a vertical heat flow through a thin horizontal slab having thickness  $dz$ . In this instance, the net flux through the slab is simply:

$$dq = q(z + dz) - q(z) = \frac{dq}{dz} dz = -k \frac{d^2 T}{dz^2} dz \quad (12.16)$$

This expression implies that  $dq \neq 0$  iff  $T$  is not a linear function of depth, so that its one-dimensional Laplacian  $d^2 T/dz^2$  is not zero. In this instance, by the conservation of energy, the net flux through the lamina must be supplied by internal sources of heat. Let  $H$  be the rate of heat produced per unit mass within the slab. The heat

generated in a volume of slab  $dV$  having unit surface area, thickness  $dz$ , and mass  $dm$  is given by:

$$dQ = H dm = \rho H dz \quad (12.17)$$

Therefore, the one-dimensional version of the stationary equation of conductive heat transfer with sources now reads:

$$k \frac{d^2 T}{dz^2} + \rho H = 0 \quad (12.18)$$

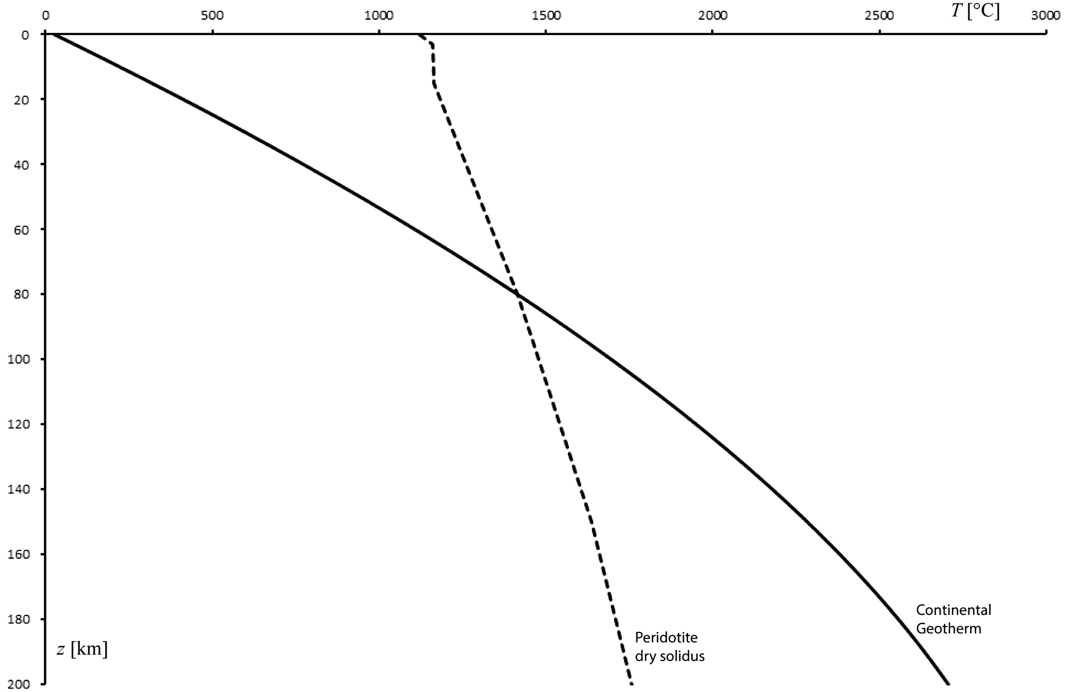
A solution to this equation can be easily found assuming a half-space with top boundary at  $z = 0$ . Let us assign the boundary conditions as follows:  $T(0) = T_0$ ,  $q(0) = -q_0$ . In this case, the solution to (12.18) is the following parabolic function:

$$T(z) = T_0 + \frac{q_0}{k} z - \frac{\rho H}{2k} z^2 \quad (12.19)$$

The function  $T = T(z)$  is called a *geotherm*. In principle, it could be used to predict the variations of temperature within the continental lithosphere, granted that the rate of radiogenic heat production  $H$  can be considered constant in the crust and in the lithospheric mantle. Assuming  $T_0 = 300$  K,  $q_0 = 60$  mWm<sup>-2</sup>,  $\rho = 3,300$  kg m<sup>-3</sup>,  $k = 3$  Wm<sup>-1</sup> K<sup>-1</sup>, and  $H = 6 \times 10^{-11}$  W kg<sup>-1</sup> gives the continental geotherm shown in Fig. 12.3. This plot would fail to describe the distribution of temperature in the asthenosphere, because it predicts partial melting of the mantle peridotite starting from  $\sim 80$  km depth. Conversely, it represents an acceptable approximation of the effective geotherm in continental areas.

An improved conductive geotherm model can be obtained taking into account that the production of radiogenic heat is not constant but decreases with depth. A good method is to consider different crustal layers, each with constant radiogenic heat production rate, and decrease  $H$  stepwise with depth. Alternatively, Turcotte and Schubert (2002) argued that a good choice for the function  $H = H(z)$  is:

$$H(z) = H_0 e^{-z/h} \quad (12.20)$$



**Fig. 12.3** Continental geotherm (solid line) for constant rate of heat production  $H$ . The dashed line shows the dry peridotite solidus, determined using expression (1.12) and PREM distribution of hydrostatic pressure

where  $H_0$  is the radiogenic heat production rate per unit mass at the Earth's surface and  $h$  is a length scale parameter for the decrease of  $H$  with depth. In this instance, (12.18) assumes the form:

$$k \frac{d^2 T}{dz^2} + \rho H_0 e^{-z/h} = 0 \quad (12.21)$$

To solve this equation, we assume that at great depth, at the lower boundary of the radiogenic heat production layer (e.g., at the Moho), we have an upward flux  $q = -q_m$  (note that for  $dT/dz > 0$ ,  $T$  increases downwards in the positive  $z$  direction, so that heat flows in the *negative*  $z$  direction). Such a basal flow can be approximated through the boundary condition:  $q(z) \rightarrow -q_m$  as  $z \rightarrow \infty$ . A first integration of (12.21) gives:

$$q(z) = -\rho H_0 h e^{-z/h} - q_m \quad (12.22)$$

Therefore, the surface heat flux will be given by:

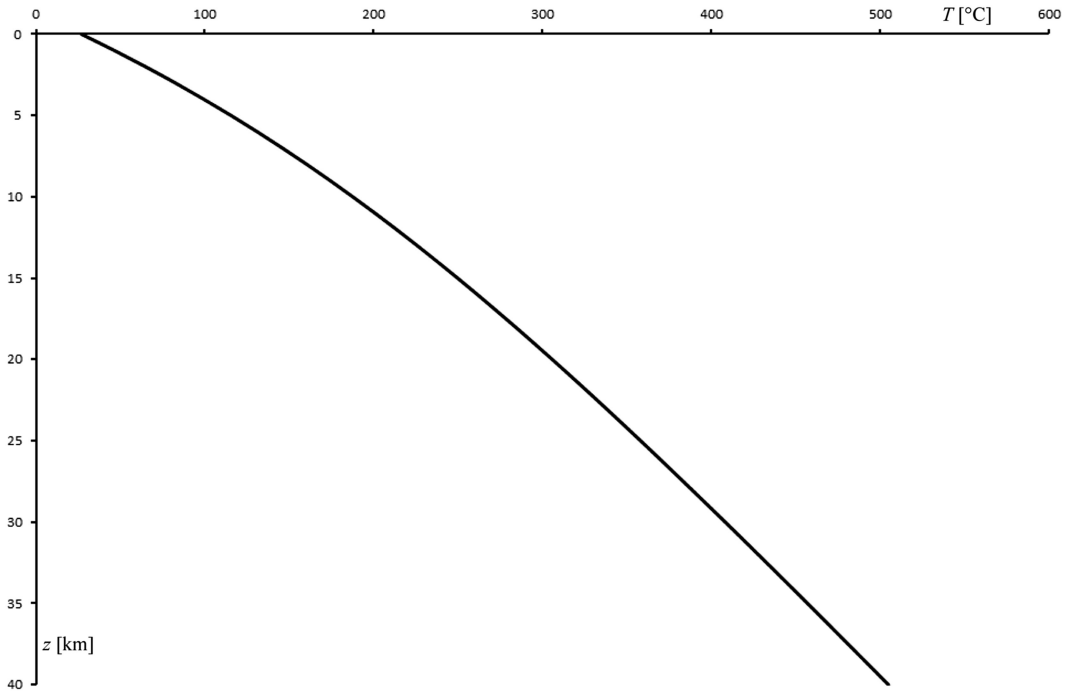
$$q_0 = -q(0) = \rho H_0 h + q_m \quad (12.23)$$

This result, which has been confirmed by direct observations, shows that the surface heat flow depends linearly from the radiogenic heat production rate at the Earth's surface. The length scale parameter  $h$  and the basal flux  $q_m$  can be estimated fitting a linear regression curve through pairs of observed values of  $q_0$  and  $H_0$ . It results that  $h \cong 10$  km and  $q_m \sim 28 \text{ mWm}^{-2}$  (Turcotte and Schubert 2002). Further integration of (12.20) gives:

$$T(z) = T_0 + \frac{q_m}{k} z + \frac{\rho H_0 h^2}{k} (1 - e^{-z/h}) \quad (12.24)$$

A possible distribution of temperature in the continental crust, based on (12.24), is shown in Fig. 12.4. We note a substantial difference with respect to the trend illustrated in Fig. 12.3, which shows a more linear growth in the crust.

This improved geotherm predicts a temperature of  $\sim 500^\circ \text{C}$  at 40 km depth, which is much less than the  $\sim 770^\circ \text{C}$  of the previous model. However, a downward continuation of



**Fig. 12.4** A more realistic continental geotherm, based on exponential decrease of heat production  $H$  with depth

the geotherm (12.24) to the lithospheric mantle would give anyway unrealistic temperatures on the basis of the estimated values from kimberlites. The mineral compositions from garnet peridotite nodules can be used to estimate both depth and temperature at which a nodule reached the equilibrium. These data can be used to constrain the geotherms of several continental regions. McKenzie et al. (2005) built steady state continental geotherms solving the equations separately for the crust, an underlying mechanical boundary layer (MBL), and a lower thermal boundary layer (TBL). In fact, as we mentioned in Sect. 1.4, the lithospheric mantle can be considered as formed by an elastic-ductile upper layer and a lower more viscous layer that are separated by the 650 °C isotherm and a sharp cut-off of seismicity. In their study, McKenzie et al. (2005) assumed a constant thermal conductivity  $k = 2.5 \text{ W m}^{-1} \text{ K}^{-1}$  for the crust, whereas the radiogenic heat generation rate  $H$  was expressed in terms of energy density and set to  $1.12 \mu\text{W m}^{-3}$  for the upper crust and  $0.4 \mu\text{W m}^{-3}$  in the lower crust. The heat generation within the thermal

and mechanical boundary layers was set to zero, thereby the steady state heat flux throughout these layers was constant and coincident with the heat flow through the Moho.

Figure 12.5 illustrates the model geotherm proposed by McKenzie et al. (2005), which is in agreement with heat flow measurements and fits  $(P, T)$  estimates from kimberlite nodules. Differently from previous models, in this geotherm the crustal contribution to the heat flow is increased, while the mantle heat flow is decreased. The surface heat flux determined from this geotherm results to be  $\sim 52 \text{ mW m}^{-2}$ .

## 12.3 Non-steady State Heat Conduction

The steady-state equation of heat conduction (12.18) can be used to model the temperature distribution in old continental crust and in the MBL. However, it cannot describe time-dependent processes such as the cooling of the oceanic lithosphere. Usually, radiogenic

heat production plays a minor role in these phenomena, so that we can assume  $H=0$ . However, in this instance, the diffusion Eq. (12.13) must be used to determine both the distribution of temperature and its time evolution. In this equation, the diffusivity  $\kappa \equiv k/c\rho$  has units [ $\text{m}^2\text{s}^{-1}$ ]. Therefore, if the temperature changes over a characteristic time interval  $\tau$ , then such a variation will propagate over a distance of the order  $L \sim \sqrt{\kappa\tau}$ , while a time  $\tau \sim L^2/\kappa$  is required for a temperature change to propagate over a distance  $L$ . Now we are going to face the case of instantaneous heating or cooling of a half-space. As shown by Turcotte and Schubert (2002), the corresponding solution can be used

in the modelling of several important geological problems. Let us assume that the temperature is uniform along horizontal planes, so that heat transfer occurs along the  $z$  direction only. In this instance, heat conduction is described by the one-dimensional version of the diffusion Eq. (12.13):

$$\frac{\partial T}{\partial t} - \kappa \frac{\partial^2 T}{\partial z^2} = 0 \quad (12.25)$$

This equation can be easily solved in a half-space having uniform temperature  $T(z,t) = T_i$  for  $t \leq 0$  and whose surface is instantaneously heated (or cooled) and maintained at a different constant temperature  $T_0$ , so that  $T(0,t) = T_0$  for  $t > 0$ . In this instance, for  $T_0 > T_i$  heat is transferred into the half-space and the internal temperature increases, whereas for  $T_0 < T_i$  the half-space cools and its temperature decreases. An example of the former situation is illustrated in Fig. 12.6. As shown in this figure, at any time  $t > 0$  we have the boundary condition:

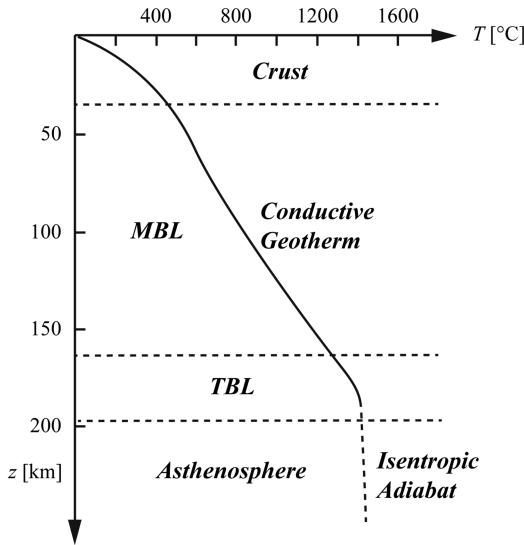
$$\lim_{z \rightarrow \infty} T(z,t) = T_i \text{ for any } t > 0 \quad (12.26)$$

The diffusion Eq. (12.25) can be solved by similarity introducing a dimensionless temperature ratio:

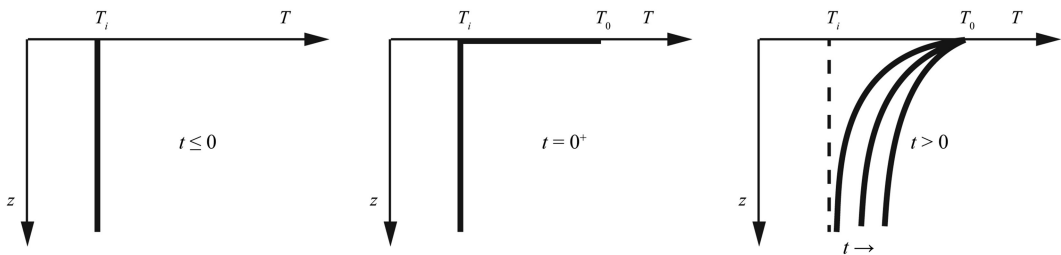
$$\theta \equiv \frac{T - T_i}{T_0 - T_i} \quad (12.27)$$

In terms of  $\theta$ , the diffusion Eq. (12.25) assumes the form:

$$\frac{\partial \theta}{\partial t} - \kappa \frac{\partial^2 \theta}{\partial z^2} = 0 \quad (12.28)$$



**Fig. 12.5** A continental lithosphere geotherm (solid line) that fits pressure and temperature estimates of garnet peridotite nodules from kimberlites (Redrawn from McKenzie et al. 2005)



**Fig. 12.6** Heating of a half-space by a sudden increase of the surface temperature

The boundary conditions for  $\theta$  arise immediately from those for  $T$ . We have that  $\theta(z, t) = 0$  for  $t \leq 0$  and  $\theta(0, t) = 1$  for  $t > 0$ . Furthermore:

$$\lim_{z \rightarrow \infty} \theta(z, t) = 0 \text{ for any } t > 0 \quad (12.29)$$

A solution by similarity is based on the idea that two solutions of (12.28) for different times should have a “similar” spatial distribution of temperatures. Therefore,  $z$  and  $t$  must appear in a particular combination, just like solutions of the plane wave equation require that distance  $x$  and time  $t$  always appear in the combination  $t - x/v$ ,  $v$  being the wave velocity. In the case of the diffusion equation, the quantity  $L(t) = \sqrt{kt}$  represents a characteristic *thermal diffusion distance*, thereby it is reasonable to assume that if  $\theta = \theta(z, t)$  is a solution to (12.28), then it will be a function of the dimensionless ratio  $z/\sqrt{kt}$  rather than of an arbitrary combination of the variables  $z$  and  $t$ . To simplify the results, it is convenient to define a dimensionless *similarity variable*  $\eta$  as follows:

$$\eta(z, t) \equiv \frac{z}{2\sqrt{kt}} \quad (12.30)$$

Formally, the similarity implies that temperature distribution at time  $t$  can be obtained from the distribution at time  $t'$  by stretching the distance  $z$  by the square root of  $t'/t$ . In fact, the transformation  $z \rightarrow z' = z\sqrt{t'/t}$  gives  $\eta' = z'/(2\sqrt{kt'}) = z/(\sqrt{t'/t}/(2\sqrt{kt'})) = z/(2\sqrt{kt}) = \eta$ . Now let us rewrite the diffusion equation in terms of  $\eta$ . By the chain rule, we have:

$$\frac{\partial \theta}{\partial t} = \frac{d\theta}{d\eta} \frac{\partial \eta}{\partial t} = -\frac{\eta}{2t} \frac{d\theta}{d\eta} \quad (12.31)$$

$$\frac{\partial \theta}{\partial z} = \frac{d\theta}{d\eta} \frac{\partial \eta}{\partial z} = \frac{1}{2\sqrt{kt}} \frac{d\theta}{d\eta} \quad (12.32)$$

$$\frac{\partial^2 \theta}{\partial z^2} = \frac{1}{2\sqrt{kt}} \frac{d^2 \theta}{d\eta^2} \frac{\partial \eta}{\partial z} = \frac{1}{4kt} \frac{d^2 \theta}{d\eta^2} \quad (12.33)$$

Substituting into (12.28) gives the following ordinary linear differential equation:

$$\frac{d^2 \theta}{d\eta^2} + 2\eta \frac{d\theta}{d\eta} = 0 \quad (12.34)$$

With respect to  $\eta$ , the boundary conditions assume the form:

$$\lim_{\eta \rightarrow \infty} \theta(\eta) = 0; \quad \theta(0) = 1 \quad (12.35)$$

Now let us set:

$$\phi \equiv \frac{d\theta}{d\eta} \quad (12.36)$$

Substituting into (12.34) reduces this equation to a first-order differential equation:

$$\frac{d\phi}{d\eta} + 2\eta\phi = 0 \quad (12.37)$$

The solution is:

$$\phi(\eta) = ae^{-\eta^2} \quad (12.38)$$

Therefore, the general solution for  $\theta$  is:

$$\theta(\eta) = a \int_0^\eta e^{-\xi^2} d\xi + b \quad (12.39)$$

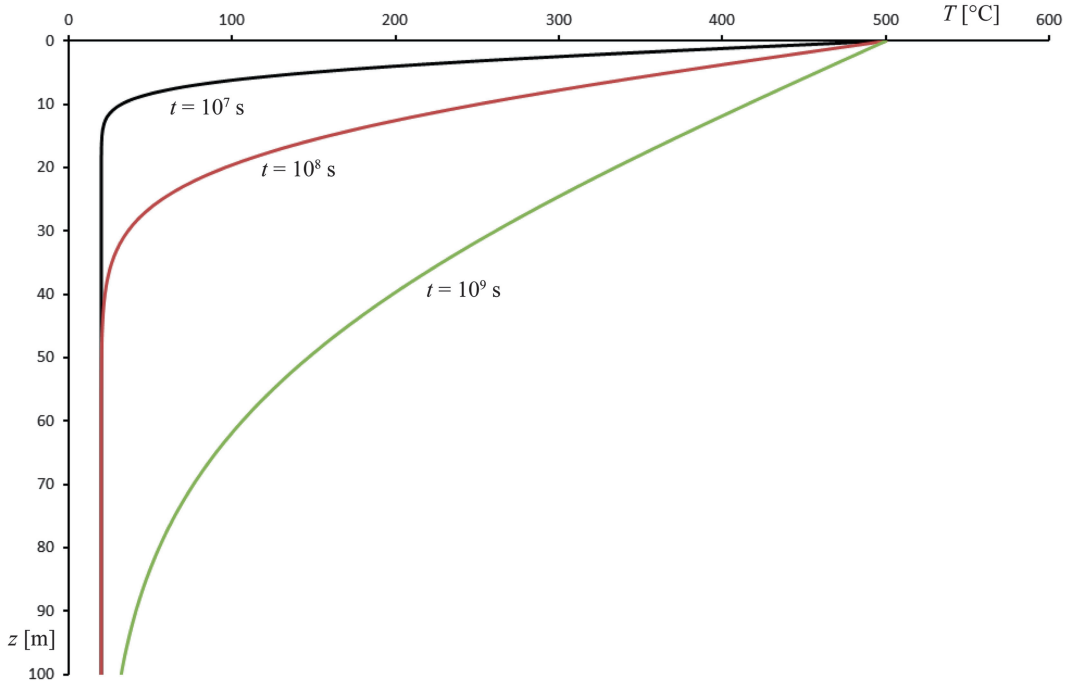
Finally, substituting the boundary conditions allows to express the solution to (12.34) in terms of error function:

$$\theta(\eta) = 1 - \frac{2}{\sqrt{\pi}} \int_0^\eta e^{-\xi^2} d\xi \equiv 1 - \text{erf}(\eta) = \text{erfc}(\eta) \quad (12.40)$$

This solution can be easily converted into a temperature distribution using (12.27) and (12.30). We have:

$$T(z, t) = (T_0 - T_i) \text{erfc}\left(\frac{z}{2\sqrt{kt}}\right) + T_i \quad (12.41)$$

A plot of  $T$  vs  $z$  is shown in Fig. 12.7 for a sudden increase of the surface temperature ( $T_0 > T_i$ ). The near-surface region where the



**Fig. 12.7** Distribution of temperature after instantaneous heating from 20 to 500 °C at the surface of a half-space after 10, 100, and 1,000 s. The assumed thermal diffusivity is  $\kappa = 1 \text{ m}^2 \text{ s}^{-1}$

temperature variations are significant is known as the *thermal boundary layer* (TBL). Turcotte and Schubert (2002) define the thickness of this region as the depth at which  $\theta = 0.1$ . Clearly, such a depth changes with time insofar as the half-space heats or cools after the initial surface variation. By the definition (12.27), the condition  $\theta = 0.1$  means that the actual variation of temperature with respect to the initial value is 10 % of the temperature change at the half-space surface. Substituting the value  $\theta = 0.1$  into (12.40) gives a corresponding value,  $\eta_T$ , for the similarity variable  $\eta$ :

$$\eta_T \equiv \text{erfc}^{-1}(0.1) \cong 1.16 \quad (12.42)$$

Therefore, the TBL thickness at time  $t$  will be given by:

$$z_T(t) = 2\eta_T \sqrt{\kappa t} \cong 2.32 \sqrt{\kappa t} \quad (12.43)$$

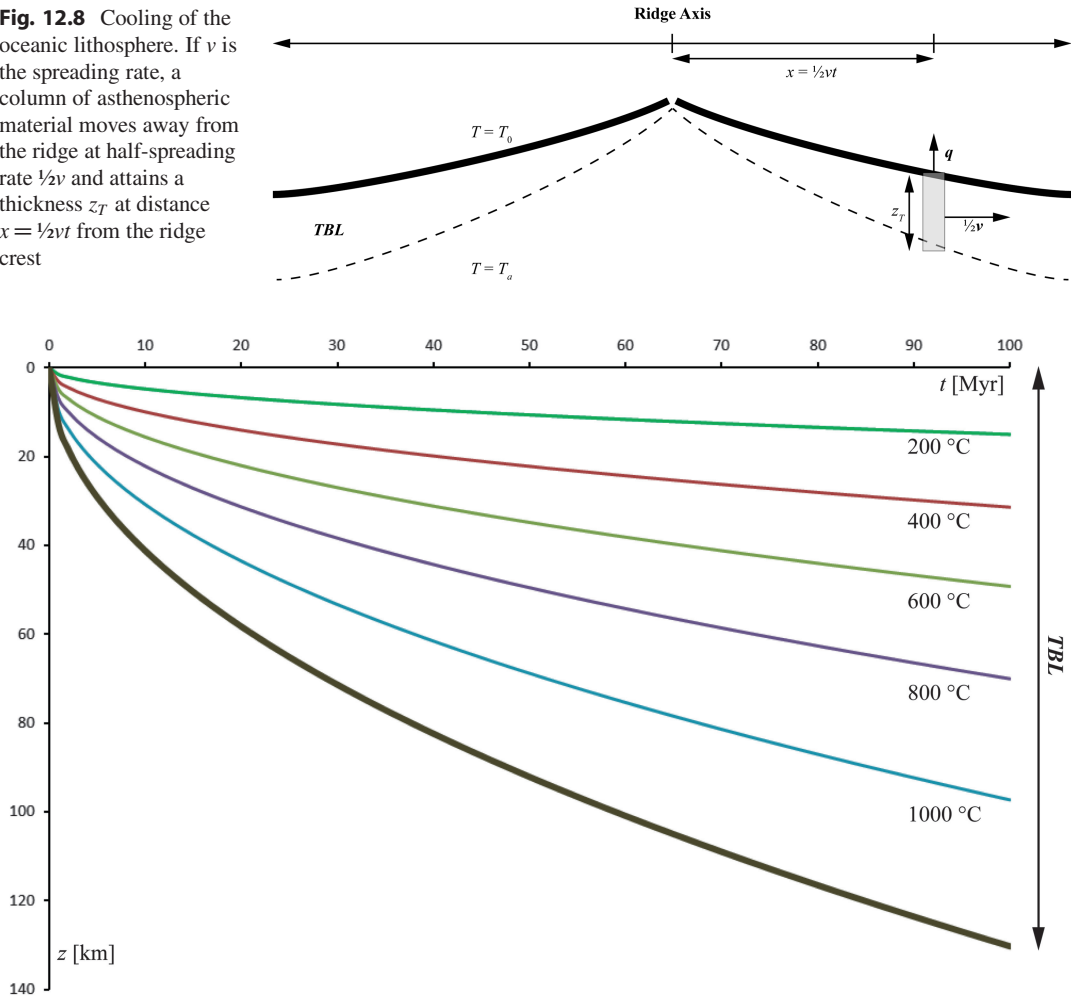
The surface heat flux corresponding to the cooling or heating law (12.41) can be easily obtained by differentiation. We have:

$$q_0 = -q(0, t) = k \left. \frac{\partial T}{\partial z} \right|_{z=0} = \frac{k(T_i - T_0)}{\sqrt{\pi \kappa t}} \quad (12.44)$$

## 12.4 Cooling of the Oceanic Lithosphere

The solution (12.41) to the diffusion equation can be adapted to describe the formation of the oceanic crust by cooling of MORB produced at a mid-ocean ridge after contact with oceanic seawater. Furthermore, in Sect. 1.3 we have seen that when fertile and wet asthenosphere melts at a spreading ridge by adiabatic decompression, the residual column of asthenospheric material leaving the melting regime is also dragged horizontally away from the ridge axis and cools by conductive loss of heat. At any time, we can divide this column into an upper part, where the potential temperature is fallen below the asthenosphere  $T_P$  ( $\sim 1,280$  °C), and a lower hotter zone, which has not yet lost a significant amount of

**Fig. 12.8** Cooling of the oceanic lithosphere. If  $v$  is the spreading rate, a column of asthenospheric material moves away from the ridge at half-spreading rate  $\frac{1}{2}v$  and attains a thickness  $z_T$  at distance  $x = \frac{1}{2}vt$  from the ridge crest



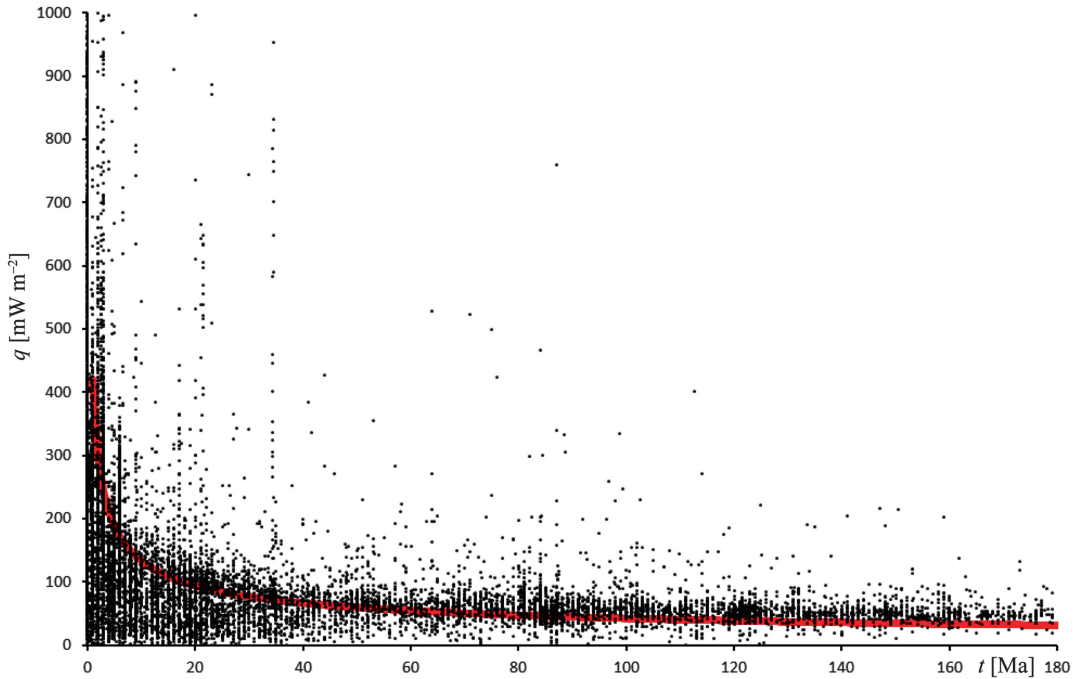
**Fig. 12.9** Isotherms of the oceanic lithosphere, determined inverting (12.45) and assuming  $\kappa = 1 \text{ mm}^2 \text{ s}^{-1}$ ,  $T_0 = 10^\circ \text{C}$ , and  $T_a = 1,280^\circ \text{C}$ . The shape of the TBL at any time (*thick black line*) has been calculated using (12.43)

heat. The conductive TBL where  $T < T_P$  defines the oceanic lithosphere. Let  $T_a$  be the initial temperature of the asthenosphere that leaves a melting regime, and let us assume that seawater instantaneously cools and maintains the surface of the residual column to the temperature  $T_0$ . In this instance, neglecting the horizontal components of heat conduction, we can apply the solution found in the previous section. Figure 12.8 illustrates an idealized cross-section through the ridge crest of a cooling oceanic plate.

If  $v$  is the spreading rate and  $x$  is the offset of an asthenospheric column from the ridge crest, then the temperature distribution at any depth  $z$  below the sea floor can be written as follows:

$$\begin{aligned} T(z, t) &= (T_0 - T_a) \operatorname{erfc}\left(\frac{z}{2\sqrt{2\kappa x/v}}\right) + T_a \\ &= (T_a - T_0) \operatorname{erf}\left(\frac{z}{2\sqrt{2\kappa x/v}}\right) + T_0 \end{aligned} \quad (12.45)$$

Figure 12.9 shows some isotherms beneath the ocean surface vs the ocean floor age. We can use (12.43) to estimate the thickness of the oceanic lithosphere of any age. This quantity is also displayed in Fig. 12.9 and coincides with the depth to the oceanic LAB minus the ocean floor depth.



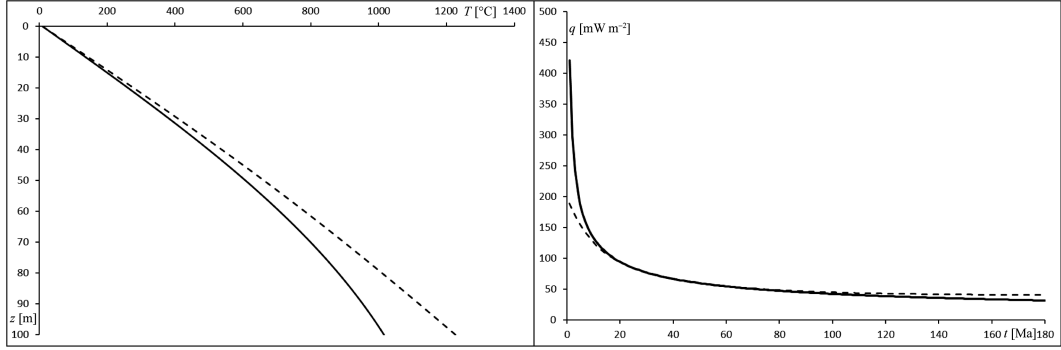
**Fig. 12.10** Heat flux as a function of the age of the ocean floor. Black dots are observed data included in the global data base maintained by the International Heat Flow Commission (see <http://www.heatflow.und.edu/index2.html>).

The theoretical trend based on the HSC model is displayed as a *red line*. This trend line was determined using (12.44) and assuming  $\kappa = 1 \text{ mm}^2 \text{ s}^{-1}$ ,  $T_0 = 10^\circ \text{C}$ ,  $T_a = 1,280^\circ \text{C}$ , and  $k = 3.3 \text{ W m}^{-1} \text{ K}^{-1}$ .

To estimate the suitability of the half-space cooling (HSC) model, we can compare the predicted surface heat flux  $q_0$  at any sea floor age  $t$  with observed values. Figure 12.10 shows the distribution of global marine heat flow data on oceanic crust provided by the *International Heat Flow Commission* and the predicted HSC heat flux given by (12.44). We note that there is considerable scatter in the observed data, even taking into account that they have been extracted from a global catalogue. A major source of scattering is associated with hydrothermal circulation through the oceanic crust and the consequent heat loss. However, the agreement between the heat flow data and the HSC trend is acceptable after  $\sim 60 \text{ Ma}$ , although the theoretical trend appears to decrease a little bit more rapidly. In fact, in old oceanic basins the observed heat flow apparently levels at  $\sim 53 \text{ mW m}^{-2}$ . A comparison with measurements performed only in thick sedimentary successions, which possibly block hydrothermal circulation and heat loss, would confirm more

clearly this misfit (Turcotte and Schubert 2002). This suggests that beyond some sea floor age the rate of conductive cooling is to some extent compensated by additional heat production. However, a more serious problem in Fig. 12.10 is represented by the apparent misfit between the observed heat flow and the theoretical trend on young oceanic lithosphere. In fact, the former appears to be significantly lower than that predicted by the HSC model.

An improvement over the HSC model, which was originally developed by Davis and Lister (1974), can be obtained taking into account that the lithosphere is heated from below as a consequence of mantle convection. In the plate cooling model (PCM) of Parsons and Sclater (1977), the oceanic lithosphere forms by cooling of an asthenospheric plate having finite thickness  $h$  and an isothermal lower boundary. Such a lower isothermal boundary prevents continued cooling and thickening of the oceanic lithosphere for older ages. The parameter  $h$  is fixed and chosen



**Fig. 12.11** *Left:* Oceanic geotherm of the HSC model (solid line) and of the PCM (dashed line) for a 100 Ma old lithosphere. *Right:* Heat flux as a function of the age

of the ocean floor in the two models. The maximum plate thickness is assumed to be  $h = 105$  km (McKenzie et al. 2005; Afonso et al. 2007)

as coincident with the effective thickness of the oldest lithosphere. If  $T_a$  is the temperature of the upper asthenosphere, then the boundary conditions can be written as follows:

$$\begin{aligned} \lim_{z \rightarrow h} T(z, t) &= T_a; \\ \lim_{z \rightarrow 0} T(z, t) &= T_0; \text{ for any } t > 0 \end{aligned} \quad (12.46)$$

$$T(z, 0) = T_a; \text{ for any } 0 \leq z \leq h \quad (12.47)$$

Therefore, the plate is assumed to have uniform temperature  $T_a$  at  $t=0$ , and at any successive time the upper and lower boundaries are maintained at temperatures  $T_0$  and  $T_a$  respectively. The solution for  $T$  (Carslaw and Jaeger 1959) has the form of an infinite series:

$$\begin{aligned} T(z, t) &= T_0 + (T_a - T_0) \\ &\times \left[ \frac{z}{h} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \exp\left(-\frac{\kappa n^2 \pi^2 t}{h^2}\right) \sin\left(\frac{n \pi z}{h}\right) \right] \end{aligned} \quad (12.48)$$

We note that for  $\kappa t \gg h^2$  the series in (12.48) tends to zero, thereby at large times a linear steady-state geotherm is attained:

$$T(z, t) \sim T_0 + (T_a - T_0) \frac{z}{h} \text{ as } t \rightarrow \infty \quad (12.49)$$

Conversely, for  $\kappa t \ll h^2$  the solution (12.48) assumes the form:

$$\begin{aligned} T(z, t) &\cong T_0 + (T_a - T_0) \\ &\times \left[ \frac{z}{h} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin\left(\frac{n \pi z}{h}\right) \right]; \text{ for } t \rightarrow 0 \end{aligned} \quad (12.50)$$

It is possible to show that the solution for  $t \rightarrow 0$  gives geotherms that do not differ significantly from those of the HSC model.

Regarding the surface heat flux, it can be obtained by (12.48) applying Fourier's law for  $z=0$ :

$$q_0(t) = \frac{k(T_a - T_0)}{h} \left[ 1 + 2 \sum_{n=1}^{\infty} \exp\left(-\frac{\kappa n^2 \pi^2 t}{h^2}\right) \right] \quad (12.51)$$

Also in this case for large times, such that  $\kappa t \gg h^2$ , a steady-state value is attained. This is given by:

$$q_0(t) \sim \frac{k(T_a - T_0)}{h} \text{ as } t \rightarrow \infty \quad (12.52)$$

which is significantly different from (12.44). A comparison of geotherm and heat flux predicted by the HSC model and the corresponding curves in the PCM is shown in Fig. 12.11. As required by (12.49), the PCM geotherm of old

lithosphere flattens and tends to a straight line. At 100 km depth the predicted temperature in the PCM is  $\sim 210^\circ\text{C}$  higher. Regarding the surface heat flow, the PCM flux curve shows a better fit to the heat flow data at the extremes of the age range. A detailed analysis of the PCM parameters has shown that this model can account quite satisfactorily for the observed heat flow (Stein and Stein 1992).

A direct consequence of the cooling of the oceanic lithosphere is represented by the progressive increase of its density. Let  $\rho = \rho(z, t)$  be the rock density at depth  $z$  and time  $t$ . This quantity is an intensive state variable that varies with temperature  $T$  and pressure  $P$ . In the case of decreasing temperature, density increases because the volume of a rock body decreases by thermal contraction.

In general, the thermodynamic relation that determines the change of volume associated with variations of temperature and pressure is:

$$dV = \left( \frac{\partial V}{\partial P} \right)_T dP + \left( \frac{\partial V}{\partial T} \right)_P dT \quad (12.53)$$

The derivatives in this expression are properties of the material that are usually expressed through the coefficient of thermal expansion,  $\alpha$ , which is defined by (1.3), and the *isothermal compressibility*  $\beta$ , given by:

$$\beta = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T \quad (12.54)$$

Substituting (12.3) and (12.54) into (12.53) gives the following expression for  $dV$ :

$$dV = V(-\beta dP + \alpha dT) \quad (12.55)$$

This expression determines the variation of the extensive variable  $V$  associated with changes of  $P$  and  $T$ . It is usually convenient to express volume variations in terms of an intensive variable rather than an extensive one. To this purpose, we introduce the *specific volume*  $v \equiv 1/\rho$ , which represents the volume per unit mass. The relative variations of  $v$  are linked to variations of density

and to variations of volume by the following simple equation:

$$\frac{dv}{v} = \frac{dV}{V} = -\frac{d\rho}{\rho} \quad (12.56)$$

Therefore, (12.55) can be rewritten in terms of density as follows:

$$d\rho = \rho(\beta dP - \alpha dT) \quad (12.57)$$

When the rock body can change freely its volume after a temperature variation, the pressure  $P$  is invariant, so that (12.57) assumes the form:

$$\frac{d\rho}{\rho} = -\alpha dT \quad (12.58)$$

Conversely, when the rock is confined, so that its volume does not change ( $dv = 0$ ), the variations of temperature and pressure are related by the following equation:

$$\beta dP - \alpha dT = 0 \quad (12.59)$$

In the oceanic lithosphere, Parsons and Sclater (1977) estimated that the thermal expansion coefficient  $\alpha$  assumed the value  $\alpha = 3.28 \times 10^{-5} \text{ K}^{-1}$ , while the more recent best-fitting value obtained by Stein and Stein (1992) is  $\alpha = 3.1 \times 10^{-5} \text{ K}^{-1}$ . We can use this estimate and the HSC model isotherms of Fig. 12.9 to calculate the horizontal gradient of density in the oceanic lithosphere. For example, Fig. 12.9 shows that at 70 km depth the distance between the 1,000 and 800  $^\circ\text{C}$  isotherms is  $\Delta t = 48$  Myrs. Therefore, assuming a spreading rate  $v = 30$  mm/year, we obtain a horizontal distance  $\Delta x = \frac{1}{2}v\Delta t = 720$  km between the two isotherms. This implies a horizontal gradient of temperature  $\partial T/\partial x \cong 0.28^\circ\text{C/km}$  at 70 km depth. Consequently, using (12.58) and assuming that  $\alpha$  does not change significantly with temperature, we can easily integrate (12.58) obtaining a total variation of density of  $\sim 21 \text{ kg/m}^3$  in 720 km. Regarding the isothermal compressibility, it increases with temperature and assumes values between  $\beta = 0.80 \times 10^{-12} \text{ Pa}^{-1}$  (at  $T \approx 700 \text{ K}$ ) and  $\beta = 1.20 \times 10^{-12} \text{ Pa}^{-1}$  at  $T \approx 2,750 \text{ K}$  (Arafin et al. 2008).

If we consider the lithosphere as an elastic solid layer “floating” over the fluid (in rheological sense) asthenosphere, we will easily realize that the lateral variability of density in the oceanic lithosphere must determine a hydrostatic imbalance that can be compensated locally only by vertical displacements and shear. This phenomenon is known as *thermal isostasy* and arises from Archimedes’ principle of hydrostatic equilibrium. To understand the application of this important principle to the asthenosphere, we recall that fluids cannot support static shear stress, thereby the equation of hydrostatic equilibrium reads:

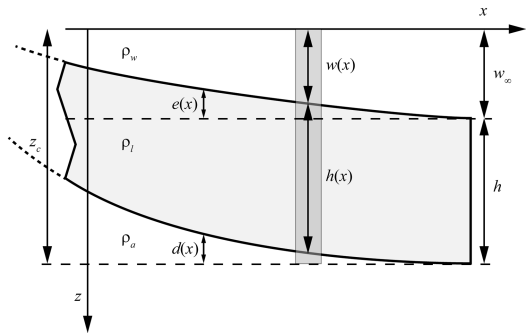
$$\frac{\partial \tau_{ij}}{\partial x_j} = -\frac{\partial P}{\partial x_i} = -\rho g_i \quad (12.60)$$

where  $\mathbf{g} = g\mathbf{k}$  is the gravity field vector. Integrating this equation, we see that the hydrostatic pressure  $P$  at any depth  $z$  does not depend from  $x$  and  $y$  and is simply the weight of the column of rock having height  $z$  and unit area. In the asthenosphere, this law must hold even when the column is formed partly by sea water, partly from oceanic MORBs, and partly from peridotite.

We have:

$$P(z) = g \int_0^z \rho(x, y, z) dz \quad (12.61)$$

It should be noted that although the density  $\rho$  in (12.61) in general depends from  $x$  and  $y$ , by (12.60) the pressure will be constant along any horizontal plane. Now let us consider the hydrostatic pressure at depth  $h$  below the sea floor, where  $h$  is the maximum thickness of the lithosphere in the PCM. Assuming that  $z = 0$  at the ocean surface and that  $w_\infty$  is the sea floor depth at large distance from the ridge, we can use the depth  $z_c = h + w_\infty$  as a reference depth, or *compensation depth*, for applying the isostasy principle, which then reads:  $P(z_c) = \text{const.}$  Let us assume that the coordinate  $x$  measures distances from the spreading ridge, so that direction  $y$  is



**Fig. 12.12** Thermal isostasy of the oceanic lithosphere. Thermal subsidence causes an increase of the ocean floor depth.  $\rho_a$ ,  $\rho_l$ , and  $\rho_w$ , are respectively the densities of the asthenosphere, the lithosphere, and the sea water. Columns of material having unit cross-section over the compensation depth  $z_c$  (dark grey rectangle) must have equal weight

along the ridge axis and can be ignored. Furthermore, let  $h(x)$  and  $w(x)$  be respectively the thickness of the lithosphere at offset  $x$  from the spreading ridge and the corresponding depth of the sea floor (Fig. 12.12). As shown in Fig. 12.9, in so far as the age of the ocean floor increases, a column of height  $h$  will include a larger fraction of dense lithosphere and a corresponding smaller fraction of the less dense asthenosphere. Furthermore, as  $x$  increases, at any depth below the sea floor the lithosphere will have increasing density because of thermal contraction. Consequently, the increased weight of a column of material with unit cross-section and height  $h$  above depth  $z_c$  must be compensated by a larger amount of water between the column and the sea surface, that is by a greater sea floor depth (Fig. 12.12).

At any time, the depth to the sea floor at distance  $x$  from the ridge must ensure the invariance of  $P(z_c)$ . Therefore, isostatic equilibrium requires subsidence of the oceanic lithosphere in so far as its age increases. If  $e(x)$  is the elevation of the ocean floor with respect to the asymptotic depth  $w_\infty$  and  $d(x)$  is the displacement of the bottom of the lithosphere above the compensation depth  $z_c$ , then the pressure  $P(z_c)$  at any offset  $x$  is given by:

$$\begin{aligned}
 P(z_c) &= g \left[ \rho_w w(x) + \rho_a d(x) + \int_0^{h(x)} \rho_l(x, \zeta) d\zeta \right] \\
 &= g \left[ \rho_w (w_\infty - e(x)) + \rho_a d(x) \right. \\
 &\quad \left. + \rho_{l0} \int_0^{h(x)} (1 - \alpha T(x, \zeta)) d\zeta \right]
 \end{aligned} \tag{12.62}$$

where  $\rho_{l0}$  is the lithosphere density at room temperature ( $T_0$ ) and the integral is performed over the lithosphere thickness at offset  $x$ . The condition of isostatic equilibrium can be expressed equating the  $P(z_c)$  at offset  $x$  to the asymptotic value:

$$\begin{aligned}
 &\rho_w (w_\infty - e(x)) + \rho_a d(x) \\
 &\quad + \rho_{l0} \int_0^{h(x)} (1 - \alpha T(x, \zeta)) d\zeta \\
 &= \rho_w w_\infty + \rho_{l0} \int_0^h (1 - \alpha T(\infty, \zeta)) d\zeta
 \end{aligned} \tag{12.63}$$

Parsons and Sclater (1977) used this condition to determine the depth to the ocean floor. They obtained the following expression for the elevation  $e(x)$  above the asymptotic depth:

$$\begin{aligned}
 e(x) &= \frac{4\alpha\rho_{l0}(T_a - T_0)h}{(\rho_{l0} - \rho_w)\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \\
 &\quad \times \exp \left[ - \left( R - \sqrt{R^2 + (2n+1)^2\pi^2} \right) x \right]
 \end{aligned} \tag{12.64}$$

where  $R$  is the Peclet number of the oceanic lithosphere and represents the ratio of heat advection rate to heat diffusion rate (see Sect. 1.3). If  $v$  is the full spreading rate, this dimensionless quantity is given by:

$$R = \frac{vh}{2\kappa} \tag{12.65}$$

For large offsets  $x$ , the depth to the sea floor assumes values close to following asymptotic value:

$$w_\infty = w_0 + \frac{\alpha\rho_a(T_a - T_0)h}{2(\rho_a - \rho_w)} \tag{12.66}$$

where  $w_0$  is the depth to the ridge crest. This expression gives the equilibrium depth of old oceanic basins. Assuming  $w_0 = 2.6$  km,  $\alpha = 3.1 \times 10^{-5} \text{ K}^{-1}$ ,  $\rho_a = 3,300 \text{ kg kgm}^{-3}$ ,  $T_a - T_0 = 1,270^\circ\text{C}$ ,  $h = 100$  km, and  $\rho_w = 1,000 \text{ kg m}^{-3}$ , we obtain from (12.66) an asymptotic depth  $w_\infty = 5.4$  km. In the case of young crust, the elevation above the asymptotic depth has the approximate expression:

$$\begin{aligned}
 e(x) &\cong \frac{\alpha\rho_{l0}(T_a - T_0)h}{2(\rho_{l0} - \rho_w)} \\
 &\quad - \frac{2\alpha\rho_{l0}(T_a - T_0)}{(\rho_{l0} - \rho_w)} \sqrt{\frac{2\kappa x}{\pi v}}; \quad \sqrt{\frac{2\kappa x}{v}} \ll h
 \end{aligned} \tag{12.67}$$

Therefore, the depth to the ocean floor will increase with the square root of the distance from the ridge or, equivalently, with the square root of the age. These results have been confirmed with high accuracy by the observed ocean floor bathymetry. A good fit was obtained by Stein and Stein (1992), who proposed the following reference model, based on the PCM, for the young and old ocean floor depths:

$$w(t) = \begin{cases} 2600 + 365\sqrt{t}; & t < 20 \text{ Ma} \\ 5651 - 2473e^{-0.0278t}; & t \geq 20 \text{ Ma} \end{cases} \tag{12.68}$$

## 12.5 Driving Mechanism of Plate Tectonics: Slab Pull and Ridge Push

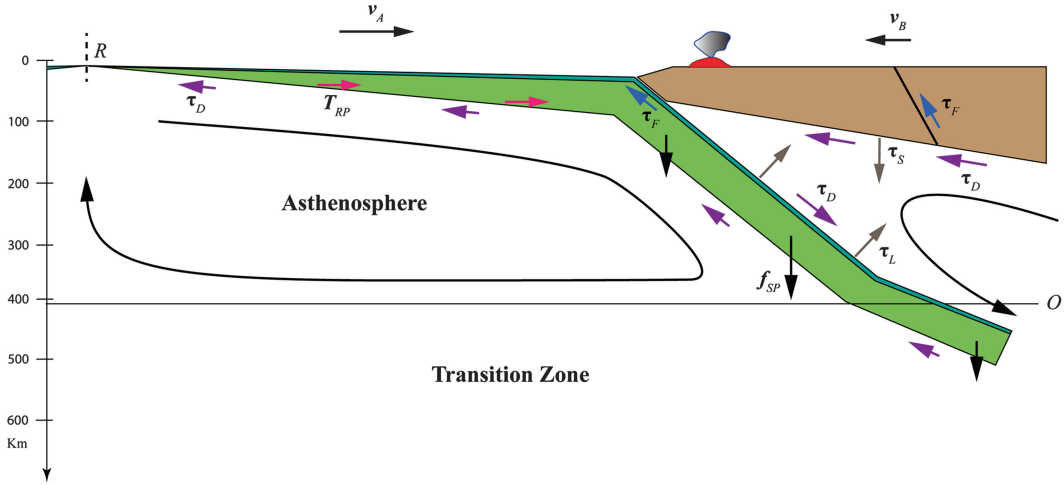
In the previous section, we have seen that the thermodynamics of the oceanic lithosphere, which includes both vertical diffusion and horizontal advection of heat, determines lateral variations of density and thickness. Although

these variations give rise to thermal subsidence, they cannot be considered as the ultimate cause of the instability that leads to subduction initiation of old oceanic lithosphere. In fact, the most reliable picture of the thermal history of oceanic plates, which is given by PCM, predicts asymptotic thermal and isostatic equilibrium as  $t \rightarrow \infty$ . In addition, a detailed petrological and geophysical analysis of the density structure of the oceanic lithosphere has shown that the depth-averaged value of  $\rho$  at  $t = 90$  Ma (including 7 km thick crust with lower density  $\rho = 2,900 \text{ kg m}^{-3}$ ) is between 3,310 and 3,312  $\text{kg m}^{-3}$  (Afonso et al. 2007). In this instance, the density contrast with respect to the surrounding asthenosphere would be  $\Delta\rho = 35.5 \text{ kg m}^{-3}$ , which is less than the commonly assumed value, but more interestingly the predicted depth-averaged density would be slightly lower than the density of the asthenosphere immediately below the compensation depth  $z_c$  ( $\rho_a \cong 3,330 \text{ kg m}^{-3}$ ). Therefore, a hypothetical gravitational instability of the old oceanic lithosphere could hardly explain the initiation of subduction, a process that is still poorly known. However, differently from the unsubducted lithosphere, slabs in the upper mantle are cooler than the surrounding mantle and have positive density contrasts up to  $\sim 200 \text{ kg m}^{-3}$ , although the portions subducted at depths close to the 670 km discontinuity may attain hydrostatic equilibrium (Ganguly et al. 2009). It is interesting to note that the model of Ganguly et al. (2009) predicts that a fragment of subducted lithosphere may have *positive* (i.e., upward directed) buoyancy in the uppermost lower mantle. In fact, in the case of old slabs the endothermic phase transitions of ringwoodite into perovskite and magnesio-wüstite occur at depths significantly greater than the 670 km discontinuity. In this instance, we have that the  $\sim 9\%$  negative density anomaly associated with the mineralogical contrast relative to the surrounding mantle, which gives a positive contribution to buoyancy, would prevail over the negative contribution associated with the thermal anomaly (see Sect. 1.6).

The increased density of the subducted lithosphere is mostly due to the metamorphism of the

oceanic crust at high temperatures and pressures and to the effect of phase transitions in the upper mantle (see Sect. 1.6). In fact, shortly after downward bending, the hydrated basalts and gabbros of the oceanic crust are converted to their high-pressure eclogitic phases, with release of substantial amounts of  $\text{H}_2\text{O}$  and a consequent increase of the crustal density to values in excess of  $3,500 \text{ kg m}^{-3}$  in a few Myrs (Ahrens and Schubert 1975; Kirby et al. 1996; Peacock and Wang 1999). Regarding the effect of phase transitions, we have seen in Chap. 1 that the exothermic phase transition of the olivine phase in peridotite to wadsleyite increases the density by  $\sim 6\%$ . Therefore, the olivine of a slab segment located just above the 410 km depth discontinuity will be subject to premature phase transition to wadsleyite. Consequently, the density contrast and the negative slab buoyancy will increase further at this depth. Schubert and Turcotte (1971) estimated that the total body force exerted on a descending slab due to the shallower phase transition was nearly as large as the force on the slab due to thermal contraction. Tassara et al. (2006) in a study about the Nazca plate estimated an average density contrast  $\Delta\rho = \rho_l - \rho_a \cong 90 \text{ kg m}^{-3}$  in the asthenosphere, while Ganguly et al. (2009) propose even larger values of  $\Delta\rho$ , in the case of old plates, in the deep asthenosphere just above the transition zone. In summary, our present knowledge of the complex thermochemical processes that accompany the penetration of slabs into the mantle allows to say that once started, the “subduction factory” can proceed autonomously, driven by the negative buoyancy and passive sinking of the slabs, at least down to the transition zone.

Today most geoscientists accept the idea that the pull exerted by the subducting slabs on the tectonic plates is the dominant force driving plate motions. This concept was originally proposed by Richter (1973) on the basis of theoretical arguments. However, it grew into a widespread theory after that an empirical analysis about the relative importance of the different torques exerted on tectonic plates confirmed that the negative buoyancy of slabs played a major role (Forsyth and Uyeda 1975). In their analysis, Forsyth and Uyeda (1975)



**Fig. 12.13** Major stresses and forces exerted on tectonic plates. Black arrows slab pull ( $f_{sp}$ ); violet arrows viscous drag ( $\tau_D$ ); red arrows ridge push ( $T_{RP}$ ); blue arrows friction stress ( $\tau_F$ ); brown arrows lift ( $\tau_L$ ) and suction

( $\tau_S$ ). Velocities  $v_A$  and  $v_B$  are relative to the top transition zone ( $O$ ). It is assumed that the ridge  $R$  is at rest with respect to the transition zone frame

considered the *slab pull* as a hydrostatic body force having magnitude  $f_{sp}(z) = [\rho_l(z) - \rho_m(z)]g$  (by Archimedes' principle),  $\rho_m$  and  $g$  being the mantle density and the gravity acceleration, respectively.

An inventory of stresses and forces exerted on tectonic plates is illustrated in Fig. 12.13. According to Forsyth and Uyeda (1975), they can be divided in *driving forces* determining plate motions and *resistive forces* that oppose them. The most important elements of the former class are the slab pull,  $f_{sp}$ , exerted by downgoing slabs and the torque about the centre of the Earth arising from the lateral variations of thickness of the oceanic lithosphere, which have been discussed in the previous section and are associated with its progressive cooling. The latter is often referred to as *ridge push*,  $N_{RP}$ , although the key factor determining this torque is lithospheric thickening, not ridge elevation (Lister 1975; Hager and O'Connell 1981; Harper 1984). According to Harper (1984), the basal traction generated by ridge push at a location  $\mathbf{r}$  is given by:

$$T_{RP}(\mathbf{r}) = g\rho_a\alpha(T_a - T_0)\kappa\nabla t \equiv p\nabla t(\mathbf{r}) \quad (12.69)$$

where  $\nabla t$  is the spatial gradient of ocean floor age at  $\mathbf{r}$ . Therefore, the traction associated with plate thickening is higher in the case of slow-spreading ridges. This expression allows to calculate the total torque exerted on an oceanic plate as a consequence of cooling. Integrating over the surface  $S$  of the plate and applying Stokes's theorem (see Appendix 1), we obtain a total torque:

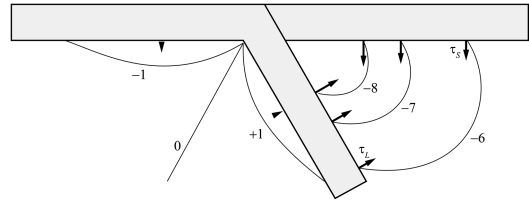
$$\begin{aligned} N_{RP} &= \int_S \mathbf{r} \times \mathbf{T}_{RP} dS = -pR \int_S \nabla t \times d\mathbf{S} \\ &= -pR \oint_{C(S)} t d\mathbf{r} \end{aligned} \quad (12.70)$$

where  $R$  is the Earth's radius,  $C(S)$  is the boundary of  $S$ , and the integral path is clockwise. The last integral in (12.70) shows that although ridge push is ultimately the result of hydrostatic forces, it can be calculated as if it were a boundary force, and this is effectively the way in which it is usually considered. The term "ridge push" comes from its appearance as a force exerted by ridges, because clearly the youngest ocean floor age always occurs at spreading centres.

Now let us come to the resistive forces illustrated in Fig. 12.13. There are essentially two classes of forces that oppose plate motions. The most important of them is represented by the basal drag,  $\tau_D$ , exerted by the asthenosphere, which is always opposite to the vector of *relative* velocity of the lithosphere with respect to the asthenosphere. In the next chapter, we shall prove that the magnitude of this shear stress increases linearly with relative velocity, asthenosphere viscosity, and horizontal pressure gradient in the asthenosphere. It is generally agreed that the basal drag is larger beneath the continents (e.g., Forsyth and Uyeda 1975), but some authors argue that in the case of large continental masses with deep roots this force could become even dominant. For example, in his “continental undertow” model Alvarez (2010) explains the continued continental collision along the Alpine-Himalayan belt by the active drag exerted by the asthenosphere. The other class of resistive forces is represented by frictional stress along strike-slip or convergent plate boundaries. This class also includes friction resistance associated with slip of bending lithosphere beneath accretionary wedges. The magnitude of these forces is small (Forsyth and Uyeda 1975), thereby they can be generally ignored in numerical modelling.

A class of forces that was not considered by Forsyth and Uyeda (1975) but has great importance in the total force balance is represented by the normal forces on the upper and lower surfaces of slabs due to dynamic pressure variations in the surrounding mantle. In reality, Forsyth and Uyeda (1975) included in their comparative analysis another force that is also related to dynamic pressure variations in the asthenosphere. This is the *suction* stress,  $\tau_s$ , exerted on the overriding plate (Figs. 12.13 and 12.14) by the low-pressure field existing beneath the accretionary wedge.

As pointed out by Tovish et al. (1978), the action of a *hydrodynamic lift*  $\tau_L$  on the subducting lithosphere (Fig. 12.13) is necessary to explain why subduction angles are much smaller than  $90^\circ$ , despite the gravitational torque exerted on slabs tends to align them with the vertical. Just as in the case of the ridge push, a quantitative study of hydrodynamic lifting requires fluid dynamics



**Fig. 12.14** Isobars and pressure distribution in a subduction zone (From Tovish et al. 1978). The unit of pressure is  $(v/(2\eta_3 h))^{1/3}$ , where  $v$  is the velocity of convergence and  $\eta_3$  is the non-linear viscosity of olivine

concepts that will be discussed in the next chapter. For the moment, it is interesting to mention the main result of the theoretical modelling performed by Tovish et al. (1978), which is illustrated in Fig. 12.14. The isobar field associated with the *corner flows* in a subduction zone shows low pressure in the oceanic corner, increasing from small suction at the base of the unsubducted plate to small compression along the lower part of the slab. Conversely, the magnitude of the pressure field is considerably higher in the arc corner and determines strong suction of the upper plate and slab lift, both increasing toward the corner.

## Problems

1. Build a continental lithosphere geotherm using a layered crustal model, assuming 10 km upper crust, 10 km middle crust, and 15 km lower crust. Use reasonable values for the radiogenic heat rates of each layer and for the other parameters. In particular, use data from Hofmeister (1999) for the thermal conductivity;
2. Determine the thickness of the magnetic crustal layers in continental and oceanic regions;
3. Determine the vertical slip rate along a fracture zone;
4. Write a computer program that converts an ocean floor age grid into a basement paleo-depth grid for any assigned time  $t$  in the geologic past. Any point having age  $t' < t$  is removed from the output grid assigning a NODATA\_value -9999.0;

5. Determine the shear stress of a 100 Ma old and 100 km thick continental margin, which deforms by simple shear, assuming a vertical contact with the continental lithosphere. Repeat the exercise for a 20 Ma old continental margin and comment the differences;

## References

- Afonso JC, Ranalli G, Fernández M (2007) Density structure and buoyancy of the oceanic lithosphere revisited. *Geophys Res Lett* 34:L10302. doi:10.1029/2007GL029515
- Ahrens TJ, Schubert G (1975) Rapid formation of eclogite in a slightly wet mantle. *Earth Planet Sci Lett* 27(1): 90–94
- Alvarez W (2010) Protracted continental collisions argue for continental plates driven by basal traction. *Earth Planet Sci Lett* 296:434–442
- Arafin S, Singh RN, George AK, Al-Lazki A (2008) Thermoelastic and thermodynamic properties of harzburgite—an upper mantle rock. *J Phys Chem Solids* 69(7):1766–1774
- Carslaw HS, Jaeger JC (1959) *Conduction of heat in solids*, 2nd edn. Oxford University Press, Oxford, 510 pp
- Clauser C, Huenges E (1995) Thermal conductivity of rocks and minerals. In: *Rock physics and phase relations, a handbook of physical constants*. AGU Reference Shelf 3, p 105–126
- Davies JH (2013) Global map of solid earth surface heat flow. *Geochem Geophys Geosyst* 14:4608–4622. doi:10.1002/ggge.20271
- Davis EE, Lister CRB (1974) Fundamentals of ridge crest topography. *Earth Planet Sci Lett* 21(4):405–413
- Forsyth D, Uyeda S (1975) On the relative importance of the driving forces of plate motion. *Geophys J Int* 43(1):163–200
- Ganguly J, Freed AM, Saxena SK (2009) Density profiles of oceanic slabs and surrounding mantle: integrated thermodynamic and thermal modeling, and implications for the fate of slabs at the 660 km discontinuity. *Phys Earth Planet Int* 172(3): 257–267
- Hager BH, O'Connell RJ (1981) A simple global model of plate dynamics and mantle convection. *J Geophys Res* 86(B6):4843–4867. doi:10.1029/JB086iB06p04843
- Harper JF (1984) Mantle flow due to internal vertical forces. *Phys Earth Planet Int* 36(3):285–290
- Hofmeister AM (1999) Mantle values of thermal conductivity and the geotherm from phonon lifetimes. *Science* 283(5408):1699–1706
- Kelvin WT (1864) On the secular cooling of the earth. *Trans R Soc Edinb* XXIII:167–169
- Kirby S, Engdahl RE, Denlinger R (1996) Intermediate-depth intraslab earthquakes and arc volcanism as physical expressions of crustal and uppermost mantle metamorphism in subducting slabs. In: Bebout GE, Scholl DW, Kirby SH, Platt JP (eds) *Subduction top to bottom*. American Geophysical Union, Washington, DC. doi:10.1029/GM096p0195
- Korenaga J (2003) Energetics of mantle convection and the fate of fossil heat. *Geophys Res Lett* 30(8):1437. doi:10.1029/2003GL016982
- Korenaga J (2008) Urey ratio and the structure and evolution of Earth's mantle. *Rev Geophys* 46:RG2007. doi:10.1029/2007RG000241
- Lister CRB (1975) Gravitational drive on oceanic plates caused by thermal contraction. *Nature* 257(5528): 663–665
- McKenzie D, Jackson J, Priestley K (2005) Thermal structure of oceanic and continental lithosphere. *Earth Planet Sci Lett* 233(3):337–349
- Parsons B, Sclater JG (1977) An analysis of the variation of ocean floor bathymetry and heat flow with age. *J Geophys Res* 82(5):803–827
- Peacock SM, Wang K (1999) Seismic consequences of warm versus cool subduction metamorphism: examples from southwest and northeast Japan. *Science* 286(5441):937–939
- Richter F (1973) Dynamical models for sea-floor spreading. *Rev Geophys Space Phys* 11:223–287
- Schubert G, Turcotte DL (1971) Phase changes and mantle convection. *J Geophys Res* 76(5):1424–1432. doi:10.1029/JB076i005p01424
- Stacey FD, Davis PM (2008) *Physics of the earth*. Cambridge University Press, Cambridge, UK, 532 pp
- Stein CA, Stein S (1992) A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature* 359(6391):123–129
- Tassara A, Götze HJ, Schmidt S, Hackney R (2006) Three-dimensional density model of the Nazca plate and the Andean continental margin. *J Geophys Res* 111:B09404. doi:10.1029/2005JB003976
- Tovish A, Schubert G, Luyendyk BP (1978) Mantle flow pressure and the angle of subduction: non-Newtonian corner flows. *J Geophys Res* 83(B12):5892–5898. doi:10.1029/JB083iB12p05892
- Turcotte DL, Schubert G (2002) *Geodynamics*, 2nd edn. Cambridge University Press, Cambridge, UK, 456 pp

## Abstract

With Chap. 13, we move from the lithosphere to the underlying mantle. In fact, this chapter introduces the important theme of mantle dynamics, in particular thermal convection and asthenosphere currents. Navier-Stokes and energy balance equations are derived and discussed, along with the classic Boussinesq approximation.

## 13.1 Continuity Equation

In this chapter, we are going to introduce the fluid behaviour of the Earth's mantle, in particular of the asthenosphere, and the influence of mantle flows on plate tectonics. As we have seen in Chap. 7, the fluid behaviour of solids is described by rheological constitutive equations that link stresses to strain rates rather than strains. Therefore, differently from seismology, where displacements and infinitesimal strains are the basic kinematic variables, a formulation of the laws that determine the long-term dynamics of the mantle will require a kinematic framework based on velocity fields and strain rates. In fluid dynamics, just as in the more general context of continuum mechanics (see Sect.2.1), the material is ideally subdivided into small parcels (or particles) of volume  $dV$  and it is assumed that the intensive variables of the system, such as velocity, density, temperature, and pressure, change continuously from point to point throughout the material. There are two different approaches to the formulation of the equations of

motion and conservation laws in fluid dynamics. In the *Eulerian* formulation, the changes of any intensive variable through time are considered at arbitrary *fixed* locations  $\mathbf{r}$ . For example, the fluid velocity  $\mathbf{v}(\mathbf{r}, t)$ , which represents the fundamental kinematic variable, is viewed as the velocity of the parcel that travels through a location  $\mathbf{r} \equiv (x, y, z)$  at time  $t$ , while its time derivative  $\partial \mathbf{v} / \partial t$  represents the variation of velocity between the particles that travel through  $\mathbf{r}$  at times  $t$  and  $t + dt$ . Therefore, in this representation the coordinates of  $\mathbf{r}$  are expressed in an inertial reference frame and the velocity is a function of four independent variables:  $\mathbf{v} = \mathbf{v}(x, y, z, t)$ . Conversely, in the *Lagrangian* representation we consider a single particle  $P$ , which is assigned a path  $\mathbf{r} = \mathbf{r}(t)$  from an initial location  $\mathbf{r}_0$ . In this instance, any change of the intensive variables is referred to a frame that is moving with  $P$ . For example, we can consider the variations of temperature or density of  $P$  while it is moving between two locations. Substituting the parametric equations  $\mathbf{r} = \mathbf{r}(t)$  into the velocity field  $\mathbf{v}$  gives a velocity vector  $\mathbf{v}(t) = d\mathbf{r}/dt = \mathbf{v}(\mathbf{r}(t), t)$  that depends only from time. In the context of the Lagrangian

representation, this quantity represents the velocity of the parcel  $P$  at time  $t$ . This approach is particularly useful if we want to formulate Newton's second law of motion for a fluid particle. In fact, if  $\rho(t)$  is the particle density at time  $t$ , then  $\rho(t)\mathbf{v}(t)$  represents its *momentum density*. Therefore, in the context of fluid dynamics, Newton's second law states that the rate of change of the momentum density,  $\rho\mathbf{v}$ , must be equal to the net force per unit volume,  $\mathbf{f}$ , exerted on a fluid particle  $P$ :

$$\rho \frac{d\mathbf{v}}{dt} = \mathbf{f} \quad (13.1)$$

Let  $q = q(x, y, z, t)$  be a scalar field in the Eulerian representation and consider the equivalent Lagrangian variable  $q(t) = q(x(t), y(t), z(t), t)$  for an assigned path  $\mathbf{r}(t) \equiv (x(t), y(t), z(t))$ . The total derivative  $dq/dt$ , which represents the rate of change of  $q$  following the fluid, is called the *material derivative* (or the *substantive derivative*) of  $q$ . By the chain rule, we have that this quantity can be expressed in terms of Eulerian variables:

$$\begin{aligned} \frac{dq}{dt} &= \frac{d}{dt} q(x(t), y(t), z(t), t) \\ &= \frac{\partial q}{\partial x} \frac{dx}{dt} + \frac{\partial q}{\partial y} \frac{dy}{dt} + \frac{\partial q}{\partial z} \frac{dz}{dt} + \frac{\partial q}{\partial t} \\ &= \frac{\partial q}{\partial t} + \mathbf{v} \cdot \nabla q \end{aligned} \quad (13.2)$$

This formula allows to determine the rate of change of a Lagrangian variable from the spatial and temporal variations of the equivalent Eulerian quantity. We are now ready to consider one of the most important differential equations of fluid dynamics, which is an expression of the mass conservation law. Let us consider an arbitrary region of space,  $\mathbf{R}$ , crossed by the fluid, fixed with respect to an Eulerian coordinate system. Some fluid enters this volume through its surface  $S(\mathbf{R})$ , while other fluid moves out. Clearly, at any point on  $S(\mathbf{R})$  only the component of  $\mathbf{v}$  along the normal direction  $\mathbf{n}$  contributes to fluid transfer in or out of  $\mathbf{R}$ . Therefore, if  $m = m(t)$  is the mass of fluid in  $\mathbf{R}$  at time  $t$ , then the outward mass flux per unit time through  $S(\mathbf{R})$ , which determines the total rate of mass change,  $dm/dt$ , is given by:

$$\frac{dm}{dt} = - \oint_{S(\mathbf{R})} \rho \mathbf{v} \cdot d\mathbf{S} \quad (13.3)$$

However, the total rate of mass change in  $\mathbf{R}$  can be also expressed as follows:

$$\frac{dm}{dt} = \frac{d}{dt} \int_{\mathbf{R}} \rho dV = \int_{\mathbf{R}} \frac{\partial \rho}{\partial t} dV \quad (13.4)$$

Therefore, applying Gauss' theorem (see Appendix I) to (13.3) and equating to (13.4), we obtain the identity:

$$\int_{\mathbf{R}} \left( \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) \right) dV = 0 \quad (13.5)$$

Since this identity holds for any arbitrary region  $\mathbf{R}$ , it is equivalent to:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (13.6)$$

This equation is known as the *continuity equation* and represents a local differential form of the mass conservation law. It should be noted that in (13.6) both the density  $\rho$  and the velocity  $\mathbf{v}$  are Eulerian variables. A Lagrangian version of this equation can be obtained differentiating the product  $\rho \mathbf{v}$  in (13.6) and using (13.2). We obtain:

$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} = 0 \quad (13.7)$$

Therefore, when the Lagrangian density  $\rho$  of any fluid particle remains constant through time, so that  $d\rho/dt = 0$ , then the velocity field is solenoidal:

$$\nabla \cdot \mathbf{v} = 0 \quad (13.8)$$

In this instance, the fluid parcels may deform and rotate, but they do not expand or shrink. Consequently, the fluid is said to be *incompressible*. In terms of Eulerian variables, the continuity equation assumes now the form:

$$\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho = 0 \quad (13.9)$$

This equation shows that the density may locally change, even in the case of incompressible fluids. The hypothesis of incompressibility has been widely used in the modelling of mantle flow. It implies that the volume  $dV$  of each mantle parcel remains invariant, thereby it is assumed that changes in pressure do not determine neither adiabatic compressional heating nor adiabatic extensional cooling during mantle convection.

## 13.2 Navier-Stokes Equations

Now we are going to formulate the equations of motion for a fluid in the framework of the Eulerian representation. To this purpose, we can start from the Cauchy momentum Eq. (7.50), taking into account that the acceleration of a volume element must be calculated as the material derivative of the velocity,  $\mathbf{a} = d\mathbf{v}/dt$ , and that gravity is the unique relevant body force in the context of mantle dynamics. Therefore, if  $\boldsymbol{\tau} = \boldsymbol{\tau}(\mathbf{r}, t)$  is the stress field and  $\mathbf{f} = \rho\mathbf{g}$  represents the gravitational body force density, applying (13.2) we have that in index notation the Eulerian version of the equations of motion assumes the form:

$$\rho \left[ \frac{\partial v_i}{\partial t} + \frac{\partial v_i}{\partial x_j} v_j \right] = \frac{\partial \tau_{ij}}{\partial x_j} + \rho g_i \quad (13.10)$$

These equations are completely general (except for the assumption of a specific body force field) and do not depend from a particular rheology of the material. To be used in the solution of geodynamical problems, they must be combined with a constitutive rheological equation that specifies the relation existing between stress, kinematic variables, which now are represented by strain rates, intrinsic parameters of the material, and state variables (see Sect. 7.4). In the case of an incompressible Newtonian fluid (see Sect. 7.4), it is possible to show that the constitutive equation reads (e.g., Ranalli 1995):

$$\tau_{ij} = -p\delta_{ij} + 2\eta\dot{\epsilon}_{ij} = -p\delta_{ij} + \eta \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \quad (13.11)$$

where  $\eta$  is the viscosity and  $p$  is the thermodynamic pressure. As the fluid is incompressible, by (13.8) we have that the trace of the strain rate tensor is zero. Therefore, taking the trace of (13.11) we obtain:

$$p = -\frac{1}{3}\tau_{kk} \quad (13.12)$$

This is an interesting relation that links the thermodynamic pressure of incompressible fluids, which must satisfy an equation of state that involves temperature and density, to an invariant of the stress tensor. In particular (13.12), establishes the equivalence between thermodynamic pressure and mean mechanical pressure in the case of incompressible fluids.

Finally, combining (13.11) and (13.10) and taking into account of (13.8) gives the equations of motion for an incompressible Newtonian fluid with uniform viscosity:

$$\rho \left[ \frac{\partial v_i}{\partial t} + \frac{\partial v_i}{\partial x_j} v_j \right] = -\frac{\partial p}{\partial x_i} + \eta \frac{\partial^2 v_i}{\partial x_j^2} + \rho g_i \quad (13.13)$$

Equations (13.13) are known as *Navier-Stokes equations* for an incompressible fluid. The first two terms at the right-hand side of these equations represent surface forces (per unit volume) exerted on a fluid particle. They are, respectively, the *pressure force*, and the *viscous force*. More general equations of motion for Newtonian fluids can be written releasing the incompressibility constraint. The most general form of the constitutive equation describing a Newtonian fluid reads (e.g., Schubert et al. 2004):

$$\tau_{ij} = -p\delta_{ij} + 2\eta\dot{\epsilon}_{ij} + \lambda\dot{\epsilon}_{kk}\delta_{ij} \quad (13.14)$$

where  $\lambda$  is called *second viscosity*. Note that  $\tau'_{ij} = 2\eta\dot{\epsilon}_{ij} + \lambda\dot{\epsilon}_{kk}\delta_{ij}$  represents the analog of the deviatoric stress introduced in Chap. 7 (see Eq. 7.26). The average pressure associated with the stress tensor is:

$$\begin{aligned}
\bar{p} &\equiv -\frac{1}{3}\tau_{kk} = p - \frac{2}{3}\eta\dot{\epsilon}_{kk} - \lambda\dot{\epsilon}_{kk} \\
&= p - \left(\lambda + \frac{2}{3}\eta\right)\dot{\epsilon}_{kk} \equiv p - k_B\dot{\epsilon}_{kk}
\end{aligned}
\quad (13.15)$$

The quantity  $k_B$  is referred to as *bulk viscosity* and represents a measure of dissipation after volumetric changes. When  $k_B = 0$ , thereby  $\lambda = -2\eta/3$ , the thermodynamic pressure  $p$  coincides with the average mechanical pressure  $\bar{p}$ . This is known as *Stokes' condition*. As mentioned above,  $\bar{p} = p$  also when  $\dot{\epsilon}_{kk} = \nabla \cdot \mathbf{v} = 0$ , that is, when the fluid is incompressible. Substitution of the constitutive Eqs. (13.14) into (13.10) allows to write the most general form of the equations of motion for a Newtonian fluid:

$$\begin{aligned}
\rho \left[ \frac{\partial v_i}{\partial t} + \frac{\partial v_i}{\partial x_j} v_j \right] &= -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \\
&\times \left[ \eta \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) + \lambda \frac{\partial v_k}{\partial x_k} \delta_{ij} \right] + \rho g_i
\end{aligned}
\quad (13.16)$$

In many cases, the term containing the bulk viscosity  $k_B$  in (13.15) is considered very small and set to zero. This is sometimes expressed by saying that the time rate of change in relative volume (i.e., the trace of the strain rate tensor) is negligible, so that the volume change is approximately elastic. In this hypothesis, the Stokes condition holds and (13.16) reduces to:

$$\begin{aligned}
\rho \left[ \frac{\partial v_i}{\partial t} + \frac{\partial v_i}{\partial x_j} v_j \right] &= -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \\
&\times \left[ \eta \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - \frac{2}{3} \frac{\partial v_k}{\partial x_k} \delta_{ij} \right) \right] + \rho g_i
\end{aligned}
\quad (13.17)$$

A further simplification can be obtained assuming an incompressible flow ( $\partial v_k / \partial x_k = 0$ ). This hypothesis gives:

$$\begin{aligned}
\rho \left[ \frac{\partial v_i}{\partial t} + \frac{\partial v_i}{\partial x_j} v_j \right] &= -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \\
&\times \left[ \eta \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \right] + \rho g_i
\end{aligned}
\quad (13.18)$$

We note that this form of the equations of motion is different from that of Navier-Stokes Eq. (13.13), despite the fluid is considered incompressible in both cases. The reason is that Eq. (13.13) holds only under the additional hypothesis that the viscosity  $\eta$  is constant.

Navier-Stokes equations and the other variants of the equations of motion for Newtonian fluids are second-order non-linear partial differential equations in the velocity  $\mathbf{v}$ . Such a non-linearity originates instability, chaos, and turbulence in the velocity fields and has represented a major problem in the application of these equations to the study of fluid dynamics. Therefore, much of our knowledge about the physics of fluids comes from the direct observation of natural phenomena or from laboratory experiments. In reality, few analytical solutions of these equations have been found so far, thereby it was only with the advent of computer science and the design of high-speed computer architectures that meaningful (non-trivial) numerical solutions were found. Chorin (1968) was the first to propose a solution to Eq. (13.13) using a finite-difference approach. Today the technique used by Chorin is known as the *time-dependent method*, which consists into a discretization of the time  $t$  and an iteration over time towards a steady-state flow field that is approached after a large number of time steps. Furthermore, a new scientific discipline was born, *computational fluid dynamics* (CFD), which combines advanced computer algorithms, numerical techniques, and theoretical fluid physics principles to perform *numerical experiments*, that is, graphical computer simulations of fluid flow that represent an inexpensive “virtual laboratory” for the study of fluid dynamics. A good introduction to this interesting subject can be found in the books of Ferziger and Perić (2002), Chung (2002), and Versteeg and Malalasekera (2007). In the more specific field of mantle and core dynamics, several open-source CFD tools have been developed during the last years, which can be used to test models of mantle flow or geodynamo. Most of them are freely available from the *Computational Infrastructure for Geodynamics* (CIG), an organization that supports the developing and maintaining of

computational geophysics software (available at: <http://www.geodynamics.org/>).

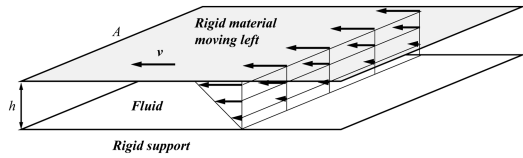
### 13.3 Energy Balance

The Navier-Stokes equations (or other more complex versions of the equations of motion for Newtonian fluids) and the continuity Eq. (13.6) form a system of four differential equations in the five unknowns  $p$  (thermodynamic pressure),  $(v_x, v_y, v_z)$  (velocity), and  $\rho$  (density). Therefore, at least an additional equation is needed to solve the system (either numerically or analytically). We note that thermodynamic pressure and density are not independent each other, because they must satisfy an *equation of state* together with the temperature  $T$ :

$$f(\rho, p, T) = 0 \quad (13.19)$$

This introduces an additional equation but also an extra unknown, the temperature  $T$ . With a total of six unknowns, we now need to solve a system of six differential equations. However, in addition to mass conservation, momentum conservation, and the state Eq. (13.19), it must be satisfied the law of *conservation of energy*, which then completes the set of equations that are needed to solve any geodynamic problem. Now we are going to consider in detail the equation corresponding to this conservation law.

The thermal structure of the mantle is essentially determined by the convective transport of heat. This concept is easily proved noting that if the continental or oceanic geotherms continued downwards below the LAB, a large part of the asthenosphere would be molten. Therefore, the temperature in the sub-lithospheric mantle must increase with depth approximately along an adiabat. Assuming that convection is sufficiently vigorous, mantle rocks that are carried to higher depth experience larger hydrostatic pressure by the overlying material and are compressed adiabatically, that is, without substantial conductive heat transfer to or from the surrounding mantle. Accordingly, their temperature increases only by adiabatic heating. The opposite occurs when hot



**Fig. 13.1** Shearing of a fluid bounded by two rigid plates in relative motion with velocity  $v$ . Arrows show the velocity field in the fluid, which varies between  $v$  and zero

mantle rocks are carried to shallow depth. In this case, the rocks experience a decrease of pressure and cool adiabatically in so far as they travel towards the LAB.

The adiabatic temperature gradient in the upper mantle has been determined in Sect. 1.3 (Eq. 1.8). However, as mentioned in Sect. 13.1, if the mantle material is considered incompressible, variations of hydrostatic pressure cannot change the volume of a small patch of convecting asthenosphere. Consequently, in this hypothesis the adiabatic temperature gradient is zero. To have an idea of the error associated with this approximation, we recall that the temperature gradient determined in Sect. 1.3 for the uppermost asthenosphere was  $(\partial T / \partial z)_S \cong 0.5^\circ \text{K km}^{-1}$ . At depth exceeding 300 km, a more appropriate value is  $0.3^\circ \text{K km}^{-1}$ , but this is still a significant variation of temperature with depth. Therefore, the approximation appears justified only if we limit our attention on upper mantle small-scale convection.

The most general equation describing the local energy balance in the fluid mantle must account for the transport of heat both by conduction and by advection, as well as for the effect of frictional heating associated with deformation. Therefore, we are looking for an equation that generalizes the heat diffusion Eq. (12.15), which only describes non-steady conduction of heat. To understand the contribution of frictional heating to the energy balance, let us consider a fluid sheet of thickness  $h$  and area  $A$ , which is sheared by the relative motion of two rigid plates as shown in Fig. 13.1. The volume of the fluid sheet is  $V = Ah$ . The fluid at contact with the fixed plate has velocity zero, while at the upper boundary the fluid velocity coincides with the velocity  $v$

of the overlying plate. This is a consequence of a fundamental empirical law of fluid dynamics that is known as *no-slip boundary condition* (e.g., Pozrikidis 2009), which states that in a fluid the tangential component of the velocity field is continuous across a solid boundary. Accordingly, the average strain rate through a cross-section of the fluid sheet in Fig. 13.1 will be:  $\dot{\epsilon} = v/h$ . If  $\tau$  is the shear stress that the overlying plate exerts on the fluid, then the total force is  $F = \tau A$ , thereby the power dissipated by shearing is:  $\dot{W} = Fv = \tau A h \dot{\epsilon} = \tau V \dot{\epsilon}$ . In general, the power dissipated in a fluid by viscous deformation is referred to as *viscous dissipation* and is associated with an irreversible conversion of mechanical energy into temperature increases.

If  $w$  is the work per unit volume associated with the viscous flow, than the viscous dissipation function,  $\Phi$ , has the expression:

$$\Phi = \dot{w} = \tau'_{ij} \frac{\partial v_i}{\partial x_j} \quad (13.20)$$

where  $\tau'$  is the deviatoric stress tensor. Substituting the deviatoric part of expression (13.14) gives:

$$\Phi = \left[ \eta \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) + \lambda \frac{\partial v_k}{\partial x_k} \delta_{ij} \right] \frac{\partial v_i}{\partial x_j} \quad (13.21)$$

Therefore, both dynamic and second viscosities lead to dissipation in a fluid. Now we are ready to formulate the law of energy conservation for a fluid parcel. Differently from the heat diffusion Eq. (12.15), the energy balance will take into account of the advection of material, thereby the partial derivative of the temperature will be substituted by a material derivative. In general, the conservation of energy of a mass of fluid requires that the rate of change of the kinetic energy,  $\dot{K}$ , plus the rate of change of the internal energy,  $\dot{U}$ , be equal to the mechanical power input,  $\dot{W}$ , plus the rate of heat produced within or entering the body,  $\dot{Q}$ :

$$\dot{K} + \dot{U} = \dot{W} + \dot{Q} \quad (13.22)$$

The rate of change of the kinetic energy can be obtained taking the total derivative of the volume integral (2.19). If  $\psi$  is an arbitrary scalar field, then the material derivative of the volume integral of  $\psi$  over a region  $\mathbf{R}$  is given by:

$$\begin{aligned} \frac{d}{dt} \int_{\mathbf{R}} \psi(\mathbf{r}, t) dV &= \int_{\mathbf{R}} \left[ \frac{d\psi}{dt} + \psi(\mathbf{r}, t) \nabla \cdot \mathbf{v} \right] dV \\ &= \int_{\mathbf{R}} \left[ \frac{\partial \psi}{\partial t} + \frac{\partial}{\partial x_i} (\psi v_i) \right] dV \end{aligned} \quad (13.23)$$

Therefore, using the continuity Eq. (13.7), for the kinetic energy we obtain:

$$\dot{K} = \frac{d}{dt} \frac{1}{2} \int_{\mathbf{R}} \rho v^2 dV = \int_{\mathbf{R}} \rho v_k \dot{v}_k dV \quad (13.24)$$

Let us consider now the internal energy  $U$ , which can be written as follows:

$$U(t) = \int_{\mathbf{R}} \rho(\mathbf{r}, t) u(\mathbf{r}, t) dV \quad (13.25)$$

where  $u$  is the internal energy per unit mass. Using (13.23) we obtain for the rate of change of the internal energy:

$$\begin{aligned} \dot{U} &= \frac{d}{dt} \int_{\mathbf{R}} \rho u dV = \int_{\mathbf{R}} \left[ \frac{d}{dt} (\rho u) + \rho u \nabla \cdot \mathbf{v} \right] dV \\ &= \int_{\mathbf{R}} [u \dot{\rho} + \rho \dot{u} + \rho u \nabla \cdot \mathbf{v}] dV = \int_{\mathbf{R}} \rho \dot{u} dV \end{aligned} \quad (13.26)$$

The total energy content of the region of fluid  $\mathbf{R}$  may change as a consequence of flow through its surface  $S(\mathbf{R})$ , heat conduction, work done by the gravity field or by surface tractions, and radioactive heat production. The mechanical power input arises from surface forces applied along  $S(\mathbf{R})$  and from the gravity force exerted on each volume element in  $\mathbf{R}$ . Let us consider first the surface force exerted on a surface element  $dS$  with orientation  $\mathbf{n}$ . If  $\mathbf{T}(\mathbf{n})$  is the traction, then the power input on  $dS$  is given by:

$$\begin{aligned} d\dot{W} &= \mathbf{T}(\mathbf{n}) \cdot \mathbf{v} dS = v_i \tau_{ij} n_j dS \\ &= (\boldsymbol{\tau} \mathbf{v}) \cdot \mathbf{n} dS = (\boldsymbol{\tau} \mathbf{v}) \cdot d\mathbf{S} \end{aligned} \quad (13.27)$$

where we have applied Cauchy's theorem (7.3) and the symmetry condition (7.2) for the stress tensor. Therefore, using Gauss' theorem (see Appendix I) we see that the power done by surface forces on  $S(\mathbf{R})$  is given by:

$$\dot{W} = \oint_{S(\mathbf{R})} (\boldsymbol{\tau} \mathbf{v}) \cdot d\mathbf{S} = \int_{\mathbf{R}} \nabla \cdot (\boldsymbol{\tau} \mathbf{v}) dV \quad (13.28)$$

The total mechanical power on  $\mathbf{R}$  is then calculated as follows:

$$\begin{aligned} \dot{W} &= \int_{\mathbf{R}} [\nabla \cdot (\boldsymbol{\tau} \mathbf{v}) + \rho \mathbf{g} \cdot \mathbf{v}] dV \\ &= \int_{\mathbf{R}} \left[ \frac{\partial \tau_{ij}}{\partial x_i} v_j + \tau_{ij} \frac{\partial v_j}{\partial x_i} + \rho g_j v_j \right] dV \\ &= \int_{\mathbf{R}} \left[ \left( \frac{\partial \tau_{ij}}{\partial x_i} + \rho g_j \right) v_j + \tau_{ij} \frac{\partial v_j}{\partial x_i} \right] dV \\ &= \int_{\mathbf{R}} \left[ \left( \frac{\partial \tau_{jk}}{\partial x_k} + \rho g_j \right) v_j + \tau_{ij} \frac{\partial v_j}{\partial x_i} \right] dV \\ &= \int_{\mathbf{R}} \left[ \dot{v}_j v_j \rho + \tau_{ij} \frac{\partial v_j}{\partial x_i} \right] dV \end{aligned} \quad (13.29)$$

where we have used the Lagrangian version of the equations of motion (13.10). Finally, the heat power associated with the heat flow  $q$  into  $\mathbf{R}$  and with the radiogenic heat  $H$  can be obtained easily generalizing (12.10) through inclusion of the contribution of  $H$ :

$$\dot{Q} = \int_{\mathbf{R}} [k \nabla^2 T + \rho H] dV \quad (13.30)$$

The energy balance and conservation law can be formulated by substituting the expressions (13.24), (13.26), (13.29), and (13.30) into (13.22) and taking into account that the equation holds for any arbitrary region  $\mathbf{R}$ . Therefore:

$$\rho \dot{u} = \tau_{ij} \frac{\partial v_j}{\partial x_i} + k \nabla^2 T + \rho H \quad (13.31)$$

The first term at the right-hand side of (13.31) can be expressed in terms of viscous dissipation function and pressure power input. In fact, by the symmetry of the stress tensor we have that:

$$\begin{aligned} \tau_{ij} \frac{\partial v_j}{\partial x_i} &= \tau_{ji} \frac{\partial v_j}{\partial x_i} = \tau_{ij} \frac{\partial v_i}{\partial x_j} = \left( -p \delta_{ij} + \tau'_{ij} \right) \frac{\partial v_i}{\partial x_j} \\ &= -p \frac{\partial v_k}{\partial x_k} + \Phi \end{aligned} \quad (13.32)$$

Therefore, the conservation law (13.31) can be rewritten as follows:

$$\rho \dot{u} = -p \nabla \cdot \mathbf{v} + \Phi + k \nabla^2 T + \rho H \quad (13.33)$$

The internal energy density  $u$  in (13.33) is not independent from the temperature  $T$  and the pressure  $p$ , thereby it is useful to find a form of the energy conservation law such that  $u$  is substituted by an expression of  $T$  and  $p$ . To this purpose, we can use the first law of thermodynamics, which states that in absence of mass exchange the infinitesimal variation of internal energy,  $du$ , of a region is the sum of the energy per unit mass absorbed in the form of heat,  $\delta q$ , and the infinitesimal work done by the surrounding on the system,  $\delta w$ :

$$du = \delta q + \delta w \quad (13.34)$$

Although irreversible processes occur in a fluid, for example viscous dissipation, a simple expression for  $du$  can be obtained only assuming that the transformation is quasi-static (i.e., reversible), thereby frictional heating and other irreversible processes are considered of secondary importance. In this hypothesis, the first law of thermodynamics can be rewritten as follows:

$$du = \delta q - p d\mathbf{v} \quad (13.35)$$

where  $\mathbf{v} = 1/\rho$  is the *specific volume* (volume per unit mass). The heat per unit mass adsorbed by the system can be expressed as follows:

$$\begin{aligned}
\delta q &= \left( \frac{\delta q}{dT} \right)_p dT + \left( \frac{\delta q}{dp} \right)_T dp \\
&= c_p dT + T \left( \frac{\partial s}{\partial p} \right)_T dp \\
&= c_p dT - \frac{\alpha T}{\rho} dp
\end{aligned} \quad (13.36)$$

where  $c_p$  is the specific heat per unit mass at constant pressure (see Eq. 1.5),  $\alpha$  is the coefficient of thermal expansion (Eq. 1.3), and  $s$  is the *specific entropy* (entropy per unit mass). In this expression, we have assumed again that the process is reversible, so that  $\delta q = Tds$  by the second law of thermodynamics. Substituting into (13.35) gives:

$$du = c_p dT - \frac{\alpha T}{\rho} dp - p dv \quad (13.37)$$

Now, by the continuity Eq. (13.7), the material derivative of the specific volume  $v$  can be calculated as follows:

$$\frac{dv}{dt} = \frac{d}{dt} \left( \frac{1}{\rho} \right) = -\frac{1}{\rho^2} \frac{d\rho}{dt} = \frac{1}{\rho} \nabla \cdot \mathbf{v} \quad (13.38)$$

Therefore, by (13.37) the material derivative of the internal energy per unit mass will be given by:

$$\frac{du}{dt} = c_p \frac{dT}{dt} - \frac{\alpha T}{\rho} \frac{dp}{dt} - \frac{1}{\rho} p \nabla \cdot \mathbf{v} \quad (13.39)$$

Finally, substituting into (13.33) gives the final canonical form of the energy balance equation:

$$\rho c_p \dot{T} - \alpha T \dot{p} = \Phi + k \nabla^2 T + \rho H \quad (13.40)$$

This equation holds in the hypothesis that the coefficient of thermal conductivity  $k$  is constant throughout the mantle. In the more general case of variable  $k$ , the law of conservation of energy will assume the following form:

$$\rho c_p \dot{T} - \alpha T \dot{p} = \Phi + \frac{\partial}{\partial x_i} \left( k \frac{\partial T}{\partial x_i} \right) + \rho H \quad (13.41)$$

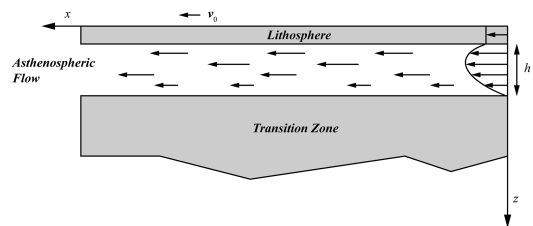
## 13.4 Flow in the Asthenosphere

Now we are going to consider some simple applications of the equations of fluid dynamics to the study of the Earth's mantle. In particular, the objective is to provide physical insights regarding the behavior of the upper mantle, which has direct influence on plate motions, despite the extreme computational difficulty associated with the numerical solution of these equations in more realistic applications. In the following, we will focus our attention on the Earth's asthenosphere, and we will search analytic solutions of the Navier-Stokes equations in simple situations, such that thermodynamic considerations can be neglected. Let us assume that a one-dimensional steady flow of asthenosphere material, considered as a Newtonian incompressible fluid, exists in a space of constant thickness  $h$  between the lithosphere and the transition zone (Fig. 13.2).

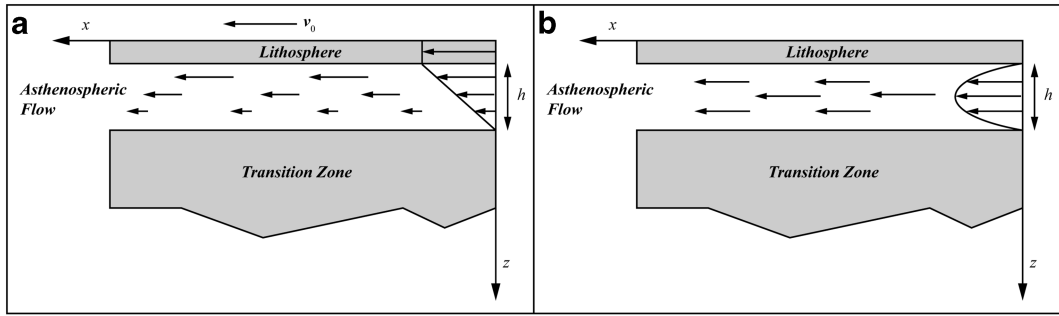
We also assume that the lithosphere moves at constant velocity  $v_0$  relative to the transition zone (Fig. 13.2). As discussed in the previous section, the no-slip boundary condition requires that the velocity field vectors just below the LAB has magnitude  $v_0$ , while the velocity is zero at the lower boundary with the transition zone. With the conventions used in Fig. 13.2, a velocity vector  $\mathbf{v}$  has components  $\mathbf{v} \equiv (v_x(z), 0, 0)$  for  $0 \leq z \leq h$ . Therefore, the stress tensor has components:

$$\boldsymbol{\tau} = \begin{bmatrix} -p & 0 & \eta \partial v_x / \partial z \\ 0 & -p & 0 \\ \eta \partial v_x / \partial z & 0 & -p \end{bmatrix} \quad (13.42)$$

In this expression, the pressure  $p$  is assumed to be a function of  $x$  and  $z$  only:  $p = p(x, z)$ .



**Fig. 13.2** A Poiseuille-Couette viscous flow in the asthenosphere (in the case of *active drag*)



**Fig. 13.3** Couette flow (a) and Poiseuille flow (b)

Furthermore, for a stationary incompressible flow it results:  $\partial v_x / \partial x = \partial v_x / \partial t = 0$ . Accordingly, the Navier-Stokes equations reduce to the following simple equations:

$$\begin{cases} 0 = -\frac{\partial p}{\partial x} + \eta \frac{\partial^2 v_x}{\partial z^2} \\ 0 = -\frac{\partial p}{\partial z} + \rho(z)g \end{cases} \quad (13.43)$$

The second equation follows from the hypothesis that the velocity field is horizontal. It simply says that the variations of pressure with depth are hydrostatic, thereby  $\partial p / \partial x$  is independent from  $z$  and the pressure field has the form:

$$p(x, z) = f(x) + g \int_0^z \rho(z) dz \quad (13.44)$$

Therefore, the first equation can be rewritten as an ordinary second-order differential equation:

$$\frac{d^2 v_x}{dz^2} = \frac{1}{\eta} \frac{\partial p}{\partial x} \quad (13.45)$$

Finally, using the no-slip boundary conditions:  $v_x(0) = v_0$  and  $v_x(h) = 0$ , we have the following simple solution:

$$v_x(z) = \frac{1}{2\eta} \frac{\partial p}{\partial x} z^2 - \left( \frac{v_0}{h} + \frac{h}{2\eta} \frac{\partial p}{\partial x} \right) z + v_0 \quad (13.46)$$

The solution (13.46) states that a velocity profile through the asthenosphere has a parabolic shape, granted that the assumptions done are valid. This kind of flow and its variants are known

as *Poiseuille-Couette flows* and have been used by several authors, especially in recent years, to model flows in the asthenosphere (e.g., Schubert and Turcotte 1972; Parmentier and Oliver 1979; Conrad et al. 2010; Höink and Lenardic 2010; Höink et al. 2011; Natarov and Conrad 2012). When  $\partial p / \partial x \neq 0$ , the flow is said to be *pressure driven*, whereas for  $\partial p / \partial x = 0$ , the magnitude of  $v_x$  decreases linearly with depth (Fig. 13.3a) and we say that the flow is a *Couette flow*. Finally, when a non-zero horizontal pressure gradient exists but  $v_0 = 0$ , then the velocity profile is parabolic and symmetric with respect to the plane  $z = h/2$ . This kind of pressure-driven flow is called a *Poiseuille flow* (Fig. 13.3b) and is always in the direction of *decreasing* pressure.

By (13.46), we note that the effect of an increasing viscosity is a tendency towards a Couette flow for any assigned pressure gradient:

$$v_x(z) \sim \left(1 - \frac{z}{h}\right) v_0 \text{ as } \eta \rightarrow \infty \quad (13.47)$$

It is useful to calculate the net areal flux per unit area through a vertical cross-section in the asthenosphere. This quantity coincides with the average horizontal velocity in the asthenospheric channel and is given by:

$$\Phi_a = \frac{1}{h} \int_0^h v_x(z) dz = -\frac{h^2}{12\eta} \frac{\partial p}{\partial x} + \frac{1}{2} v_0 \quad (13.48)$$

Assuming  $h = 300$  km, an average viscosity  $\eta = 10^{20}$  Pa s, and a pressure gradient of 10 kPa  $\text{km}^{-1}$ , we have that the pressure-driven contribu-

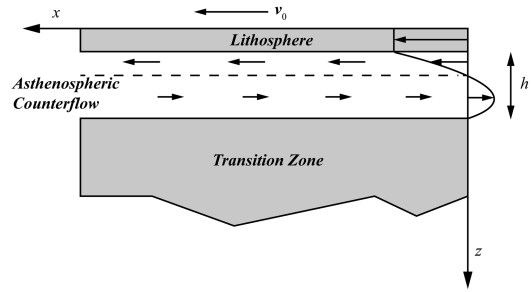
tion to the net areal flux, which coincides with the first term at the right-hand side of (13.48), is  $\sim 2.50 \times 10^{-10} \text{ ms}^{-1} = 8 \text{ mm year}^{-1}$ . With a plate velocity  $v_0 = 50 \text{ mm year}^{-1}$ , approximately  $1/4$  of the total areal flux would come from pressure gradients, but for  $\partial p/\partial x = 1 \text{ kPa}$  this contribution would be reduced to  $\sim 3\%$ . Therefore, granted that the average viscosity assumed for the asthenosphere is correct, the pressure-driven flow becomes important only when the horizontal gradient of pressure is significant. Substantial pressure gradients in the asthenosphere have been hypothesized since the 1990s (e.g., Phipps Morgan et al. 1995). More recently, lateral pressure variations of  $7,500\text{--}8,000 \text{ Pa km}^{-1}$  have been reported for the East Pacific Rise region (Conder et al. 2002), and possibly in excess of  $100 \text{ kPa km}^{-1}$  along the Tonga trench (Conder and Wiens 2007). An interesting consequence of the asthenospheric flow is represented by the drag,  $\tau_D$ , exerted on the overlying tectonic plates. This is easily calculated from (13.42) and (13.46):

$$\tau_D = \tau_{xy}(x, 0) = \eta \left. \frac{\partial v_x}{\partial z} \right|_{z=0} = -\frac{v_0 \eta}{h} - \frac{h}{2} \frac{\partial p}{\partial x} \quad (13.49)$$

Using the same parameters of the example above, we see that the viscous drag term  $v_0 \eta/h$  gives a contribution of  $\sim 0.53 \text{ MPa}$  to the total drag, while with a pressure gradient of  $10 \text{ kPa km}^{-1}$  the pressure-driven component is three times larger:  $(h/2)\partial p/\partial x \sim 1.5 \text{ MPa}$ . Therefore, for  $\eta = 10^{20} \text{ Pa s}$ , substantial variations of pressure in the asthenosphere generate active drag of tectonic plates. The expression (13.49) shows that when  $v_0$  and  $\partial p/\partial x$  have opposite sign (as in Fig. 13.2), the active pressure-driven drag opposes the passive viscous stress, thereby an equilibrium velocity exists such that  $\tau_D = 0$ . This is given by:

$$v_{eq} = -\frac{h^2}{2\eta} \frac{\partial p}{\partial x} \quad (13.50)$$

When  $\partial p/\partial x = 10 \text{ kPa km}^{-1}$  and using the same asthenosphere thickness and average viscosity of the examples above, we have that  $v_{eq} = 4.5 \times 10^{-9} \text{ ms}^{-1} = 142 \text{ mm year}^{-1}$ . This



**Fig. 13.4** Poiseuille-Couette flow and counterflow (this is an example of *passive drag*)

is the velocity (relative to the transition zone reference frame) at which a tectonic plate, initially at rest, would be accelerated by an asthenospheric pressure-driven flow. When  $v_0$  and  $\partial p/\partial x$  have the same sign, the pressure-driven drag reinforces the viscous stress and the flow assumes the shape illustrated in Fig. 13.4. This is called the *asthenospheric counterflow* (Schubert and Turcotte 1972; Chase 1979) and represents a hypothetical situation that does not find confirmation in the observed pattern of gravity anomalies and dynamic topography (e.g., Turcotte and Schubert 2002).

Another interesting feature of the simple model discussed above is the predicted amount of shear heating. From (13.20) and (13.42) we have that the viscous dissipation function is given by:

$$\begin{aligned} \Phi(x, z) &= \tau'_{ij} \frac{\partial v_i}{\partial x_j} = \eta \left( \frac{\partial v_x}{\partial z} \right)^2 \\ &= \eta \left[ \frac{1}{\eta} \left( z - \frac{h}{2} \right) \frac{\partial p}{\partial x} - \frac{v_0}{h} \right]^2 \end{aligned} \quad (13.51)$$

In the case of a Couette flow this expression reduces to:

$$\Phi = \eta \frac{v_0^2}{h^2} \quad (13.52)$$

Now let us consider the law of conservation of energy (13.41), which now assumes the simple form:

$$\Phi + k \frac{\partial^2 T}{\partial z^2} = 0 \quad (13.53)$$

Substituting (13.52) gives:

$$\eta \frac{v_0^2}{h^2} + k \frac{\partial^2 T}{\partial z^2} = 0 \quad (13.54)$$

This equation can be integrated immediately assigning the boundary conditions:  $T(0) = T_0$  and  $T(h) = T_a$ . The result is:

$$T(z) = T_0 + \frac{z}{h} \left( T_a - T_0 + \frac{\eta v_0^2}{2k} \right) - \frac{\eta v_0^2}{2k} \frac{z^2}{h^2} \quad (13.55)$$

It is useful to rewrite this expression in terms of dimensionless temperature ratio. We have:

$$\theta(z) \equiv \frac{T(z) - T_0}{T_a - T_0} = \frac{z}{h} \left( 1 + \frac{\eta v_0^2 / (2k)}{T_a - T_0} \right) - \frac{z^2}{h^2} \left( \frac{\eta v_0^2 / (2k)}{T_a - T_0} \right) \quad (13.56)$$

Therefore, in the case of a pure Couette flow, the temperature distribution in the asthenosphere is governed by the single dimensionless parameter  $(\eta v_0^2 / (2k)) / (T_a - T_0)$ . The dimensionless quantity:

$$\text{Pr} \equiv \frac{\eta}{\rho \kappa} \quad (13.57)$$

is a characteristic of the fluid and represents the ratio of the momentum diffusivity (or *kinematic viscosity*)  $\nu \equiv \eta / \rho$  to the thermal diffusivity  $\kappa$ . It is called *Prandtl number* and says how much rapidly a fluid diffuses its momentum relative to the heat diffusion rate. Another dimensionless parameter that describes a fluid property is the *Eckert number*, which represents the ratio of kinetic to thermal energy:

$$E \equiv \frac{v_0^2}{c_p (T_a - T_0)} \quad (13.58)$$

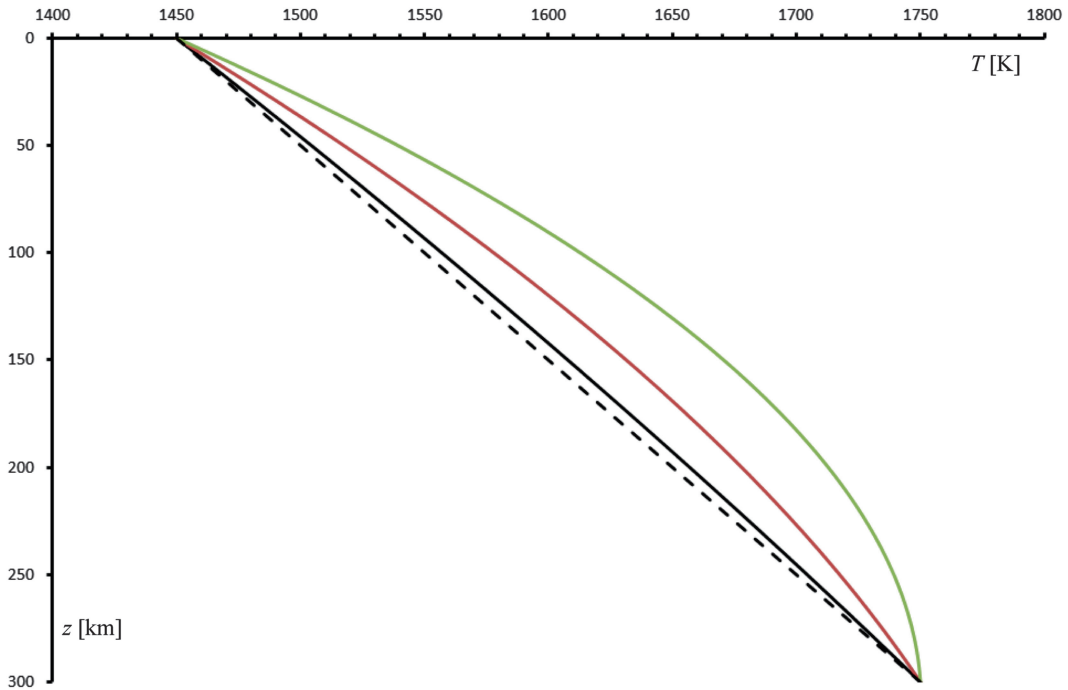
Combining the definitions (13.57) and (13.58), we see that the temperature distribution in the asthenosphere is governed by the dimensionless parameter  $\frac{1}{2}\text{Pr}E$ :

$$\theta(z) = \frac{z}{h} \left( 1 + \frac{1}{2} \text{Pr} E \right) - \frac{z^2}{h^2} \left( \frac{1}{2} \text{Pr} E \right) \quad (13.59)$$

Plots of the expected temperatures in the asthenosphere for various plate velocities are shown in Fig. 13.5. When  $v_0 = 0$  ( $\text{Pr}E = 0$ ), there is no frictional heating and the geotherm is a straight line. The temperatures in excess of this linear trend are associated with viscous dissipation. In terms of dimensionless quantities, the excess temperature due to frictional heating is given by:

$$\delta\theta(z) = \frac{1}{2} \text{Pr} E \frac{z}{h} \left( 1 - \frac{z}{h} \right) \quad (13.60)$$

The excess temperature has a maximum  $\delta\theta_{\max} = (1/8)\text{Pr}E$  for  $z/h = 1/2$ . For example, for  $v_0 = 50 \text{ mm year}^{-1}$  ( $\text{Pr}E = 0.21$ ) we would have  $\delta T_{\max} = 7.86 \text{ K}$  at  $z = 150 \text{ km}$  below the LAB. In this example, the maximum increase of temperature due to shear heating would be only  $\sim 2.6 \%$  of the temperature difference between the base and the top of the asthenospheric channel. However, if the average viscosity in the asthenosphere were one order of magnitude greater, say  $\eta = 10^{21} \text{ Pa s}$ , a fast moving plate that travels at  $v_0 = 100 \text{ mm year}^{-1}$  would trigger an asthenospheric flow with  $\text{Pr}E = 8.38$ . Accordingly, the maximum excess temperature would be  $\sim 314 \text{ K}$  and we would have downward heat loss at the base of the asthenosphere! This simple thermodynamic consideration suggests that an appropriate value of the average viscosity in the asthenosphere should not exceed  $10^{20} \text{ Pa s}$ . This value is confirmed by recent accurate estimates based on surface wave tomography and seismic anisotropy (Conrad et al. 2007; Conrad and Behn 2010), which give an average upper mantle viscosity  $\eta_{um} = 0.5\text{--}1.0 \times 10^{21} \text{ Pa s}$ , and a value of asthenosphere viscosity  $\eta = 0.5\text{--}1 \times 10^{20} \text{ Pa s}$ . Consequently, there is strong evidence that active asthenospheric drag of tectonic plates associated with pressure-driven flows represents a real mechanism explaining the non-equilibrium states discussed in Sect. 6.7. The lateral variations of pressure that



**Fig. 13.5** Predicted temperatures in the asthenosphere in the case of Couette flow. Here  $z$  is the depth starting from the LAB and the curves refer to the case  $v_0 = 0$  (dashed line),  $v_0 = 50 \text{ mm year}^{-1}$  (solid black line),

$v_0 = 100 \text{ mm year}^{-1}$  (red line), and  $v_0 = 150$  (green line). The other parameters are:  $T_0 = 1,450 \text{ K}$ ,  $T_a = 1,750 \text{ K}$ ,  $c_p = 1 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ,  $k = 4 \text{ W m}^{-1} \text{ K}^{-1}$ ,  $\rho = 3,450 \text{ kg m}^{-3}$ ,  $\eta = 10^{20} \text{ Pa s}$ , and  $h = 300 \text{ km}$

originate pressure-driven flows are associated with a variety of sources of upper mantle inhomogeneity. The first, obvious, possibility is represented by thermo-chemical discontinuities at the COBs or close to the cratonic roots. This *edge-driven convection* is a form of small-scale mantle convection that explains several interesting geological phenomena (King and Anderson 1998 and refs. therein), such as surface topography anomalies, uplift of rift flanks, and the anomalous igneous activity along some COBs (see Sect. 1.3). Another, much more important, source of upper mantle inhomogeneity is represented by mantle plumes, which could be responsible for large-scale flows in the mantle and the acceleration of continents of the dimensions of India (Cande and Stegman 2011; Becker and Faccenna 2011). It is likely that in the next future we will be able to improve our knowledge about these interesting

phenomena through a better comprehension of the spatial distribution and kinematics of mantle heterogeneity.

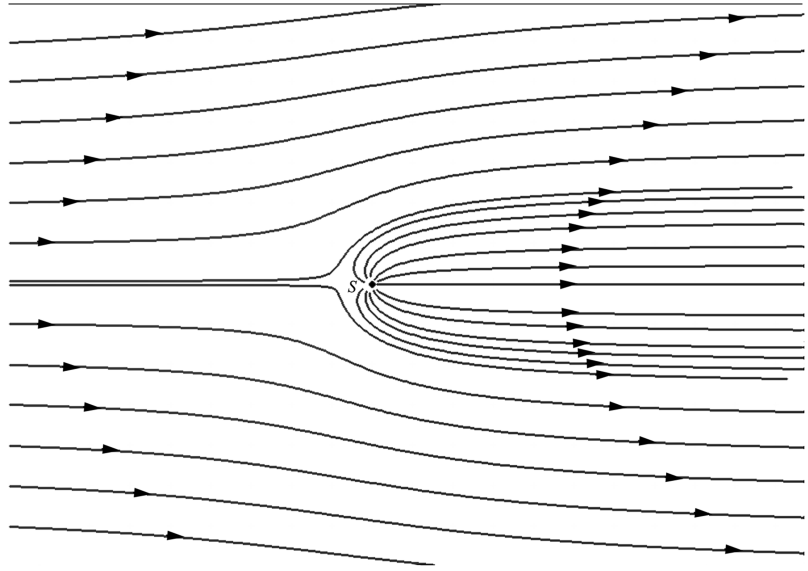
### 13.5 Stream Functions and Two-Dimensional Flows

Let  $\mathbf{v} = \mathbf{v}(\mathbf{r}, t)$  be the instantaneous velocity field at a *fixed* time  $t$  and let  $\mathbf{r}_0$  be an arbitrary space point. Starting from  $\mathbf{r}_0$ , we can build a parametric curve  $\mathbf{r} = \mathbf{r}(\zeta)$  such that:

$$\frac{d\mathbf{r}}{d\zeta} = \mathbf{v}(\mathbf{r}(\zeta), t) \quad (13.61)$$

A curve based on (13.61) is referred to as a *streamline*. Velocity vectors are always tangent to these lines, which are commonly used to illustrate the instantaneous pattern of velocity in a fluid

**Fig. 13.6** Parabolic streamline pattern, resulting from a superposition of a linear flow and a source  $S$ . It is clearly visible a stagnation point close to  $S$ . This pattern is often used as a model of the interaction between a mantle plume and the large-scale asthenospheric flow (Sleep 1987, 1990)



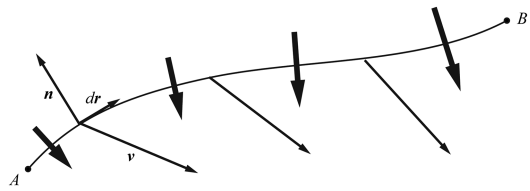
(Fig. 13.6). The kinematic pattern at any given time  $t$  may include *stagnation points*, that is, points where  $\mathbf{v} = 0$ , either in the fluid interior or along its boundary. Two or more streamlines may meet at a stagnation point, as illustrated in Fig. 13.6. In general, streamlines can be closed lines, extend to infinity, or terminate at stagnation points. It should be noted that the path  $\mathbf{r} = \mathbf{r}(t)$  of a fluid parcel does not coincide, in general, with any streamline unless the motion is steady. A flow field is said to be *two-dimensional* when the velocity  $\mathbf{v}(\mathbf{r}, t)$  is at any point normal to some fixed direction. In this instance, it is always possible to choose a Cartesian reference frame such that  $\mathbf{v}(\mathbf{r})$  has components  $(u(\mathbf{r}), 0, v(\mathbf{r}))$  at any point  $\mathbf{r}$ . Two-dimensional velocity fields in fluid materials have some interesting properties that greatly simplify the solution of the Navier-Stokes equations when the fluid can be considered incompressible.

For a two-dimensional incompressible fluid, the mass conservation Eq. (13.8) reduces to:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial z} = 0 \quad (13.62)$$

This condition is clearly satisfied if  $u$  and  $v$  are derivatives of a scalar field  $\psi = \psi(x, z, t)$ :

$$u = \frac{\partial \psi}{\partial z} ; \quad v = -\frac{\partial \psi}{\partial x} \quad (13.63)$$



**Fig. 13.7** Areal flux through a path between two points A and B

The function  $\psi$  is termed the *stream function* and represents a valuable tool for solving a number of problems in geodynamics. By (13.63), we see that the stream function associated with any two-dimensional flow is not unique, because an arbitrary constant may be added to  $\psi$  that generates the same velocity field  $(u, v)$ . Now let us consider an arbitrary path in the  $(x, z)$  plane between two points A and B (Fig. 13.7). At any point along this curve, the versor normal to the tangent vector  $d\mathbf{r} = (dx, dz)$  is given by:

$$\mathbf{n} = \left( \frac{dy}{dr}, -\frac{dx}{dr} \right) \quad (13.64)$$

The areal flux across the path is defined as the line integral of the normal component of the velocity along the path. By (13.63) and (13.64), this quantity is given by:

$$\begin{aligned}
 \Phi(A, B) &= \int_A^B (\mathbf{v} \times \mathbf{n}) \, dr = \int_A^B (u \, dz - v \, dx) \\
 &= \int_A^B \left( \frac{\partial \psi}{\partial z} dz + \frac{\partial \psi}{\partial x} dx \right) \\
 &= \int_A^B \nabla \psi \cdot d\mathbf{r} = \int_A^B d\psi = \psi(B) - \psi(A)
 \end{aligned} \tag{13.65}$$

Therefore, the change of the stream function between two points is equal to the areal flow rate across *any* path linking the two points. In the case that the path coincides with a streamline, there is no flux through the curve, because by definition the velocity is always tangent to a streamline. Consequently, a fundamental property of the stream function is that it is invariant along streamlines. Another important property is associated with the *vorticity* of the velocity field. In general, this is a vector field defined as follows:

$$\boldsymbol{\omega} = \nabla \times \mathbf{v} \tag{13.66}$$

In the case of case of a two-dimensional field, this is a vector in the invariant  $y$  direction:

$$\boldsymbol{\omega} = -j \left( \frac{\partial v}{\partial x} - \frac{\partial u}{\partial z} \right) \equiv j \, \omega_y \tag{13.67}$$

Therefore, substituting the definitions (13.63) it results:

$$\omega_y = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} = \nabla^2 \psi \tag{13.68}$$

This expression implies that when the stream function is a harmonic function, then the flow is irrotational, thereby  $\mathbf{v}$  is a potential field. Now let us consider the Navier-Stokes Eq. (13.13) and assume that the inertial term  $\rho \dot{\mathbf{v}}$  can be neglected because of the very low accelerations. In this instance, introducing the variable  $P(x, z) \equiv p(x, z) - \rho g z$ , we have that in the two-dimensional case the equations can be rewritten as follows:

$$\begin{cases} -\frac{\partial P}{\partial x} + \eta \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} \right) = 0 \\ -\frac{\partial P}{\partial z} + \eta \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial z^2} \right) = 0 \end{cases} \tag{13.69}$$

Substituting the stream function gives:

$$\begin{cases} -\frac{\partial P}{\partial x} + \eta \left( \frac{\partial^3 \psi}{\partial x^2 \partial z} + \frac{\partial^3 \psi}{\partial z^3} \right) = 0 \\ -\frac{\partial P}{\partial z} - \eta \left( \frac{\partial^3 \psi}{\partial x^3} + \frac{\partial^3 \psi}{\partial z^2 \partial x} \right) = 0 \end{cases} \tag{13.70}$$

We can eliminate the pressure term from these equations and obtain a single differential equation for  $\psi$  taking the partial derivative of the first equation with respect to  $z$  and the partial derivative of the second equation with respect to  $x$ . Then, subtracting the resulting equations gives:

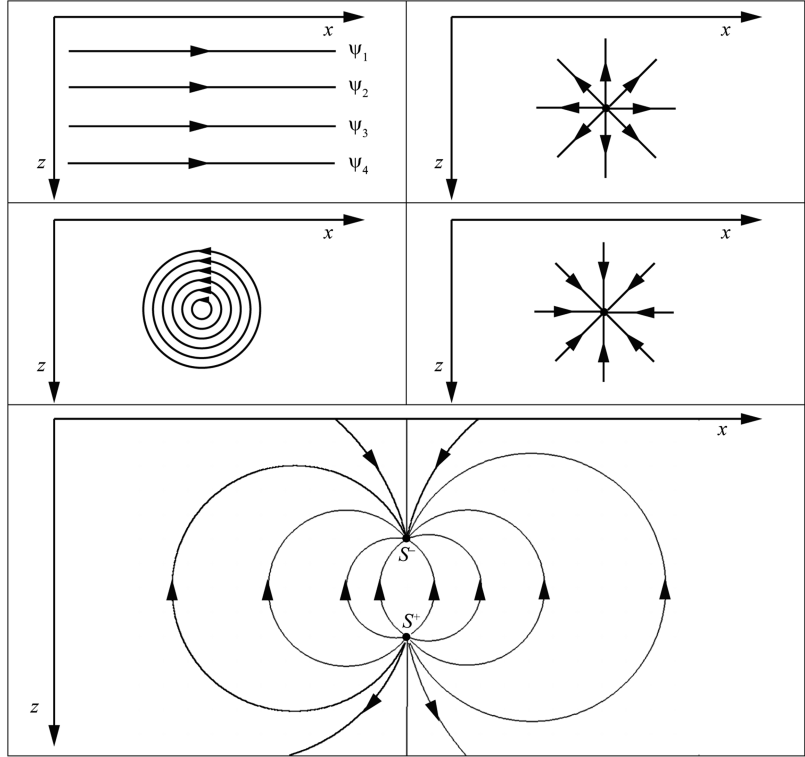
$$\frac{\partial^4 \psi}{\partial x^4} + 2 \frac{\partial^4 \psi}{\partial x^2 \partial z^2} + \frac{\partial^4 \psi}{\partial z^4} = \nabla^4 \psi = 0 \tag{13.71}$$

Therefore, the stream function is biharmonic. It is common practice in geodynamics to display flow fields through sets of streamlines, which are usually chosen in such a way that the variation  $\delta\psi$  between neighbouring streamlines is constant. With this convention, in so far as the distance  $d$  between two streamlines increases, the velocity decreases, because  $v \approx \delta\psi/d$ . Finally, it is often useful to express the components of a two-dimensional velocity field in polar coordinates  $(r, \theta)$ . In this instance, the relation between stream function and velocity components  $(u_r, u_\theta)$  is:

$$u_r = \frac{1}{r} \frac{\partial \psi}{\partial \theta} ; u_\theta = -\frac{\partial \psi}{\partial r} \tag{13.72}$$

Two-dimensional modelling through stream functions has been widely used in geodynamics because of its simplicity and because some complex flows can be represented by superposition of elementary model flows. The most simple model is a *uniform flow* in which the streamlines are parallel each other (Fig. 13.8). *Source* and *sink flows* are purely radial outward or inward flows, respectively. The velocity is everywhere

**Fig. 13.8** Basic flow fields: uniform flow (top-left), source monopole (top-right), vortex flow (middle-left), sink (middle-right), and doublet flow (bottom)



radial and constant along circles about the source or the sink. In a *vortex flow* the streamlines are concentric circles about a point (Fig. 13.8). Finally, *doublet flows* result from the combination of a source  $S^+$  and a sink  $S^-$  with equal strength. In this case, the streamlines are circles passing through  $S^+$  and  $S^-$ , as illustrated in Fig. 13.8. It is a simple exercise to find the stream function associated with these basic flows. In the case of a uniform flow in the  $x$  direction, we have:

$$\frac{\partial \psi}{\partial z} = u ; \quad \frac{\partial \psi}{\partial x} = 0 \quad (13.73)$$

Integrating these equations gives:

$$\psi = vz \quad (13.74)$$

In polar coordinates, this expression assumes the form:

$$\psi(r, \theta) = vr \sin \theta \quad (13.75)$$

Let us consider now source flows. In this case, the velocity  $v = v(r) = u_r$  and we can assume

that the source emits fluid isotropically at a steady volumetric flowrate  $Q$ . Therefore, for any circle  $C(r)$  about the source we must have:

$$Q = \oint_{C(r)} v d\ell = 2\pi r u_r \quad (13.76)$$

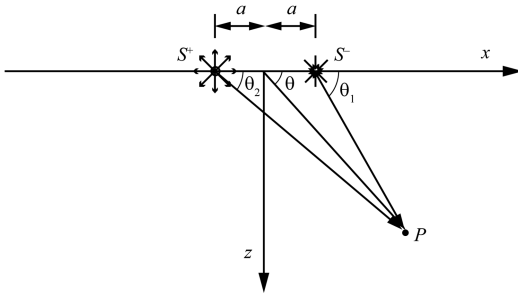
Accordingly, the velocity field in polar coordinates will be given by:

$$u_r = \frac{Q}{2\pi r} ; \quad u_\theta = 0 \quad (13.77)$$

Finally, by (13.72) we have for the stream function:

$$\psi = \frac{Q}{2\pi} \theta \quad (13.78)$$

In the case of a sink flow, (13.78) holds with a strength  $Q < 0$ . Now let us turn to the vortex flows, in which the radial pressure gradient is always zero. Clearly,  $u_r = 0$  for this class of flows, and the velocity then depends only on the distance from the vortex center. Therefore, in this



**Fig. 13.9** Method for calculating the stream function of a doublet flow at a point  $P$  by superposition of elemental source and sink stream functions

case  $\psi$  does not depend from  $\theta$  by (13.72). In the simplest case, the tangential component of  $v$  is given by:

$$u_\theta = \frac{K}{r} \quad (13.79)$$

where  $K$  is a constant whose sign discriminates between clockwise and counterclockwise vortices.

Solving (13.72) from this velocity field gives:

$$\psi = -K \ln r \quad (13.80)$$

Finally, let us consider a doublet flow. As illustrated in Fig. 13.9, the stream function results from the superposition of the individual stream functions associated with the source and with the sink:

$$\psi = \frac{Q}{2\pi} (\theta_2 - \theta_1) \quad (13.81)$$

Taking the tangent of (13.81) gives:

$$\tan\left(\frac{2\pi\psi}{Q}\right) = \tan(\theta_2 - \theta_1) = \frac{\tan \theta_2 - \tan \theta_1}{1 + \tan \theta_1 \tan \theta_2} \quad (13.82)$$

From elementary geometry considerations we have that:

$$\tan \theta_1 = \frac{r \sin \theta}{r \cos \theta - a} ; \tan \theta_2 = \frac{r \sin \theta}{r \cos \theta + a} \quad (13.83)$$

Therefore, substituting these expressions into (13.82) and taking the arctangent gives, for small values of the parameter  $a$ :

$$\psi = \frac{Q}{2\pi} \tan^{-1} \left( \frac{2ar \sin \theta}{r^2 - a^2} \right) \cong \frac{Qar \sin \theta}{\pi (r^2 - a^2)} \quad (13.84)$$

Finally, taking the limit as  $a \rightarrow 0$  and  $Q \rightarrow \infty$ , we obtain:

$$\psi = \frac{K \sin \theta}{r} \quad (13.85)$$

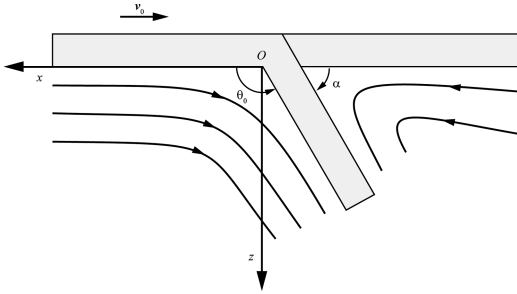
where the parameter  $K$  is the *doublet strength*. As mentioned above, the most important feature of the basic flows in fluid dynamics is that more complex and realistic flows can be constructed by superposition of the elemental stream functions associated with the basic flows. A more in-depth introduction to this topic can be found in Batchelor (2000). Here we mention the so-called *Rankine half-body*, which results from superposition of a uniform flow of velocity  $v_0$  and a source of strength  $Q$ . Assuming that the source  $S$  is placed at the origin of the reference frame, the stream function of this flow can be obtained summing the functions (13.75) and (13.78):

$$\psi = v_0 r \sin \theta + \frac{Q}{2\pi} \theta \quad (13.86)$$

The corresponding flow is illustrated in Fig. 13.6 and has a parabolic pattern. This model originally was introduced to explain the shape of the broad topographic swell associated with the Hawaiian hotspot (Sleep 1987, 1990), but has been successfully employed to represent in general the interaction of mantle plumes with the asthenospheric flow (e.g., Ribe and Christensen 1994; Walker et al. 2005). By (13.72) we obtain for the velocity components:

$$u_r = v_0 \cos \theta + \frac{Q}{2\pi r} ; u_\theta = -v_0 \sin \theta \quad (13.87)$$

For  $(r, \theta) = (Q/2\pi v_0, \pi)$  both components of the velocity vanish and we have a stagnation



**Fig. 13.10** Corner flows and streamlines at a subduction zone

point. The stream function at this location is given by:  $\psi = \frac{1}{2}Q$ . The corresponding streamline has a parabolic shape and is called the *stagnation streamline*. It indeed forms the outline of an object, the Rankine half-body. Streamlines within the Rankine half-body do not belong to the large-scale flow surrounding the source, as it is evident from Fig. 13.6.

Now we are going to consider another interesting application of the techniques based on stream function analysis. In Sect. 12.5, we mentioned the corner flows that form at subduction zones (Fig. 13.10). Here we shall solve the equations of motion in the case of a Newtonian incompressible flow and assuming that the contribution of the inertial terms is negligible in the Navier-Stokes equations. In this instance, we have seen that the equations of motion reduce to the biharmonic Eq. (13.71) for  $\psi$ . In plane polar coordinates, this equation assumes the form:

$$\nabla^4 \psi = \nabla^2 \left( \frac{\partial^2 \psi}{\partial r^2} + \frac{1}{r} \frac{\partial \psi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \psi}{\partial \theta^2} \right) = 0 \quad (13.88)$$

Let us assume that the overriding plate of a subduction system is at rest with respect to the top transition zone. In this instance, the hinge line is at rest and we can set up a reference frame as indicated in Fig. 13.10. The boundary conditions of this model are:

$$\begin{cases} u_\theta(r, 0) = -\frac{\partial \psi}{\partial r} \Big|_{\theta=0} = 0 \\ u_r(r, 0) = \frac{1}{r} \frac{\partial \psi}{\partial \theta} \Big|_{\theta=0} = -v_0 \end{cases} \quad (13.89)$$

Furthermore,

$$\begin{cases} u_\theta(r, \theta_0) = -\frac{\partial \psi}{\partial r} \Big|_{\theta=\theta_0} = 0 \\ u_r(r, \theta_0) = \frac{1}{r} \frac{\partial \psi}{\partial \theta} \Big|_{\theta=\theta_0} = v_0 \end{cases} \quad (13.90)$$

These boundary conditions are compatible with a solution by separation of variables of the form:

$$\psi(r, \theta) = r f(\theta) \quad (13.91)$$

Substituting into (13.88) gives:

$$\begin{aligned} \nabla^4 \psi &= \nabla^2 \left( \frac{1}{r} f(\theta) + \frac{1}{r} \frac{d^2 f}{d\theta^2} \right) \\ &= \frac{1}{r^3} \left( f(\theta) + 2 \frac{d^2 f}{d\theta^2} + \frac{d^4 f}{d\theta^4} \right) = 0 \end{aligned} \quad (13.92)$$

The general solution for  $f$  is:

$$f(\theta) = a \sin \theta + b \cos \theta + c \theta \sin \theta + d \theta \cos \theta \quad (13.93)$$

The constants  $a, b, c$ , and  $d$  must be chosen so that the boundary conditions (13.89) and (13.90) are satisfied. Therefore, we must have:

$$f(0) = 0 ; f(\theta_0) = 0 \quad (13.94)$$

Furthermore, by (13.91) we have that  $u_r = f'(\theta)$ . Consequently,

$$f'(0) = -v_0 ; f'(\theta_0) = v_0 \quad (13.95)$$

From the condition  $f(0) = 0$  we have immediately:  $b = 0$ . Similarly, from the first of the conditions (13.95) we obtain:  $a + d = -v_0$ . Therefore, we are left with two linear equations in the unknown parameters  $c$  and  $d$ .

The solution for  $a, c$ , and  $d$  is then:

$$\begin{aligned} a &= -\frac{v_0 \theta_0}{\theta_0 + \sin \theta_0} ; c = \frac{v_0 (1 + \cos \theta_0)}{\theta_0 + \sin \theta_0} ; \\ d &= -\frac{v_0 \sin \theta_0}{\theta_0 + \sin \theta_0} \end{aligned} \quad (13.96)$$

Finally, the complete expression for the stream function for the corner flow is:

$$\psi = r \left[ -\frac{v_0 \theta_0}{\theta_0 + \sin \theta_0} \sin \theta + \theta \left( \frac{v_0 (1 + \cos \theta_0)}{\theta_0 + \sin \theta_0} \times \sin \theta - \frac{v_0 \sin \theta_0}{\theta_0 + \sin \theta_0} \cos \theta \right) \right] \quad (13.97)$$

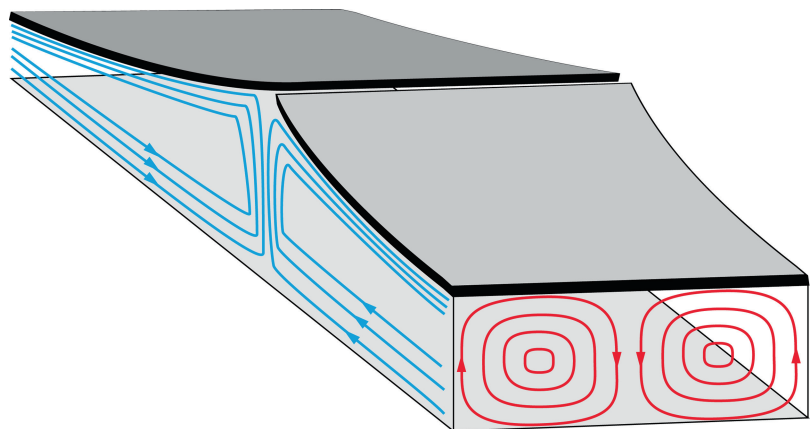
The streamlines of this function are shown in Fig. 13.10. It is interesting to note that both components of the velocity field generated by the stream function (13.97) are independent from  $r$ . Lastly, we mention that a very similar solution exists also in the case of non-Newtonian fluids (Tovish et al. 1978).

### 13.6 Rayleigh-Bénard Convection

*Rayleigh-Bénard convection*, or *thermal convection*, is a natural convection that occurs in a fluid layer that is heated from below and cooled from above. In these conditions, a regular pattern of convection cells develops between the top and the bottom of the layer, because the hot light fluid near the lower boundary tends to ascend, while cooled fluid near the top becomes denser than the average and sinks. At the scale of the upper mantle Richter and Parsons (1975), proved that small-scale convective rolls superimpose the large-scale upper mantle flow determining the complex flow pattern illustrated in Fig. 13.11. To describe quantitatively this process, it is neces-

sary to start from the consideration that thermal convection ultimately is driven by density variations arising from thermal expansion or contraction. Consequently, we will be concerned with the problem to give a specific form to the equation of state (13.19), which determines the relationship between thermodynamic pressure, density, and temperature. Unfortunately, solids lack an equation of state that can be formulated on the basis of theoretical arguments, similar to the well-known ideal gas equation,  $PV = nRT$ . In general, each solid has its own equation of state based on empirical grounds, which specifies how the density changes as a function of temperature and pressure:  $\rho = \rho(T, p)$ . However, in the Earth's mantle there are also significant density changes associated with phase transitions or, possibly, with variations of chemical composition.

Despite these complications, most of the density variations in the Earth's mantle are due to hydrostatic compression. Therefore, starting from a hypothetical homogeneous mantle in hydrostatic equilibrium, where heat is transferred only by conduction, it should be possible to find an approximate solution of the equations that describe the onset of thermal instabilities and the formation of steady Rayleigh-Bénard rolls. Such an approximation exists and is called *Boussinesq approximation*. Although this method has been widely used in mantle geodynamics studies, its theoretical ground is often outlined using confusing arguments, both in articles and books. With the objective to delimitate precisely the field of



**Fig. 13.11** Large scale upper-mantle circulation (blue lines) and Rayleigh-Bénard rolls (in red). The total flow pattern in the upper mantle results from superposition of these two flows

applicability of this method, here we will follow the more rigorous approach of Spiegel and Veronis (1960) and Furbish (1997). Let us assume that pressure,  $p$ , temperature,  $T$ , and density,  $\rho$ , have expressions of the type:

$$\begin{aligned} p(\mathbf{r}, t) &= \bar{p} + p_0(z) + p'(\mathbf{r}, t) \\ T(\mathbf{r}, t) &= \bar{T} + T_0(z) + T'(\mathbf{r}, t) \\ \rho(\mathbf{r}, t) &= \bar{\rho} + \rho_0(z) + \rho'(\mathbf{r}, t) \end{aligned} \quad (13.98)$$

where  $\bar{p}$ ,  $\bar{T}$ , and  $\bar{\rho}$  are global spatial averages,  $p_0$ ,  $T_0$ , and  $\rho_0$  are variations about the mean in the pure conductive limit (that is, in hydrostatic conditions), and the primed variables denote fluctuations associated with fluid motion.

It is also assumed that these variables are defined in a fluid layer of thickness  $H$ , heated from below and maintained at temperature  $T_m$ , while the top is cooled from above and maintained at a lower temperature  $T_a$ . Now let us introduce *scale heights* for the state variables defined above. These parameters measure the distance over which the corresponding state variable changes by a factor  $e$  (the Napier's constant) along a vertical profile:

$$\begin{aligned} H_p &= \left| \frac{1}{\bar{p}} \frac{dp_0}{dz} \right|^{-1}; \quad H_T = \left| \frac{1}{\bar{T}} \frac{dT_0}{dz} \right|^{-1}; \\ H_\rho &= \left| \frac{1}{\bar{\rho}} \frac{d\rho_0}{dz} \right|^{-1} \end{aligned} \quad (13.99)$$

Our first approximation is to assume that  $H$  throughout the fluid is much less than the smallest scale height, and that the latter coincides with the density scale height:

$$H \ll H_\rho \quad (13.100)$$

This condition can be rewritten as follows:

$$\left| \frac{1}{\bar{\rho}} \frac{d\rho_0}{dz} \right| \ll \frac{1}{H} \quad (13.101)$$

Then, integrating from  $z=0$  to  $z=H$  we conclude that:

$$\frac{\Delta\rho_0}{\bar{\rho}} \equiv \varepsilon \ll 1 \quad (13.102)$$

where  $\Delta\rho_0$  is the maximum change of the static density variation  $\rho_0$  over the distance  $H$ . The quantity  $\varepsilon$  should be considered as the maximum acceptable error in simplifying the equations governing mantle convection, in particular thermal convection. Therefore, we shall neglect any term in these equations having the same order of magnitude of  $\varepsilon$ . Condition (13.102) is sufficient to produce simplified equations when the velocities have infinitesimal magnitude, for example at the onset of convection. However, in the context of non-linear dynamics it is also necessary to assume that the relative magnitude of the fluctuations associated with fluid motion does not exceed the static variation  $\varepsilon$ .

Thus, we require that:

$$\left| \frac{\rho'}{\bar{\rho}} \right| \leq O(\varepsilon) \quad (13.103)$$

In normal conditions, we expect that  $\rho' \ll \Delta\rho_0$ , thereby it should not be necessary to verify a posteriori that condition (13.103) is satisfied. Now let us consider the equation of state (13.19) of the fluid, which can be written in the form:

$$\rho = \rho(p, T) \quad (13.104)$$

Expanding  $\rho$  in Taylor series about the state  $(\bar{p}, \bar{T})$  yields:

$$\begin{aligned} \rho &= \bar{\rho} + \frac{\partial\rho}{\partial T}(T - \bar{T}) + \frac{\partial\rho}{\partial p}(p - \bar{p}) \\ &+ \frac{1}{2} \frac{\partial^2\rho}{\partial T^2}(T - \bar{T})^2 + \frac{1}{2} \frac{\partial^2\rho}{\partial T \partial p}(T - \bar{T})(p - \bar{p}) \\ &+ \frac{1}{2} \frac{\partial^2\rho}{\partial p^2}(p - \bar{p})^2 + \dots \end{aligned} \quad (13.105)$$

where it is intended that the derivatives are calculated at the mean state  $(\bar{p}, \bar{T})$ . Now let us define the average coefficient of thermal expansion,  $\bar{\alpha}$ , and the average isothermal compressibility,  $\bar{\beta}$ , as follows:

$$\bar{\alpha} \equiv -\frac{1}{\bar{\rho}} \frac{\partial\rho}{\partial T} \Big|_{T=\bar{T}, p=\bar{p}}; \quad \bar{\beta} \equiv \frac{1}{\bar{\rho}} \frac{\partial\rho}{\partial p} \Big|_{T=\bar{T}, p=\bar{p}} \quad (13.106)$$

Substituting into (13.105) gives:

$$\begin{aligned} \rho = \bar{\rho} & \left[ 1 - \bar{\alpha} (T - \bar{T}) + \bar{\beta} (p - \bar{p}) \right. \\ & + \frac{1}{2\bar{\rho}} \frac{\partial^2 \rho}{\partial T^2} (T - \bar{T})^2 \\ & + \frac{1}{2\bar{\rho}} \frac{\partial^2 \rho}{\partial T \partial p} (T - \bar{T}) (p - \bar{p}) \\ & \left. + \frac{1}{2\bar{\rho}} \frac{\partial^2 \rho}{\partial p^2} (p - \bar{p})^2 + \dots \right] \end{aligned} \quad (13.107)$$

Therefore,

$$\begin{aligned} \frac{\rho - \bar{\rho}}{\bar{\rho}} = & -\bar{\alpha} (T - \bar{T}) + \bar{\beta} (p - \bar{p}) \\ & + \frac{1}{2\bar{\rho}} \frac{\partial^2 \rho}{\partial T^2} (T - \bar{T})^2 + \dots \end{aligned} \quad (13.108)$$

Now, because  $|\rho_0| \leq \Delta \rho_0$ , it follows that  $\rho_0 / \bar{\rho} \leq O(\epsilon)$ . Consequently, by (13.103) it results:

$$\frac{\rho - \bar{\rho}}{\bar{\rho}} = \frac{\rho_0 + \rho'}{\bar{\rho}} \leq O(\epsilon) \quad (13.109)$$

This relation implies that also the right-hand side of (13.108) is of order  $O(\epsilon)$ . Therefore, we must have:

$$\begin{aligned} \frac{1}{2\bar{\rho}} \frac{\partial^2 \rho}{\partial T^2} (T - \bar{T})^2 & < O(\epsilon^2) ; \\ \frac{1}{2\bar{\rho}} \frac{\partial^2 \rho}{\partial T \partial p} (T - \bar{T}) (p - \bar{p}) & < O(\epsilon^2) \text{ etc.} \end{aligned} \quad (13.110)$$

Hence, to order  $\epsilon$  the expression (13.107) can be rewritten as follows:

$$\rho = \bar{\rho} \left[ 1 - \bar{\alpha} (T - \bar{T}) + \bar{\beta} (p - \bar{p}) \right] \quad (13.111)$$

This is a linearized version of the equation of state in the Boussinesq approximation. Substituting (13.98) into (13.111) gives:

$$\rho_0 + \rho' = -\bar{\alpha} \bar{\rho} (T_0 + T') + \bar{\beta} \bar{\rho} (p_0 + p') \quad (13.112)$$

In the static conductive state,  $p' = T' = \rho' = 0$ . Then:

$$\rho_0 = \bar{\rho} \left( -\bar{\alpha} T_0 + \bar{\beta} p_0 \right) \quad (13.113)$$

$$\rho' = \bar{\rho} \left( -\bar{\alpha} T' + \bar{\beta} p' \right) \quad (13.114)$$

Let us consider now the continuity Eq. (13.7). Substituting (13.98) and solving for the divergence of  $\mathbf{v}$  gives:

$$\nabla \cdot \mathbf{v} = -\frac{1}{\bar{\rho}} \left( 1 + \frac{\rho_0}{\bar{\rho}} + \frac{\rho'}{\bar{\rho}} \right)^{-1} \frac{d}{dt} (\rho_0 + \rho') \quad (13.115)$$

The factor within the brackets can be expanded into a geometric series, giving:

$$\begin{aligned} \left( 1 + \frac{\rho_0}{\bar{\rho}} + \frac{\rho'}{\bar{\rho}} \right)^{-1} &= 1 - \frac{\rho_0}{\bar{\rho}} - \frac{\rho'}{\bar{\rho}} + \left( \frac{\rho_0}{\bar{\rho}} + \frac{\rho'}{\bar{\rho}} \right)^2 \\ &\dots = 1 + O(\epsilon) \end{aligned} \quad (13.116)$$

Therefore, using (13.102) we conclude that:

$$\nabla \cdot \mathbf{v} = -\frac{\epsilon}{\Delta \rho_0} \frac{d}{dt} (\rho_0 + \rho') + O(\epsilon^2) \quad (13.117)$$

Consequently, to the first order in  $\epsilon$ ,

$$\nabla \cdot \mathbf{v} \cong 0 \quad (13.118)$$

This surprising result says that to the first order in  $\epsilon$  the velocity field has the same solenoidal property of velocity fields associated with incompressible fluids. Such conclusion should not lead to think that we are modelling the mantle as an incompressible fluid. Equation (13.118) only specifies a property of the velocity field in the Boussinesq approximation for *compressible* fluids. Now let us turn to the equations of motion. In absence of motion, we have that  $\rho' = p' = 0$ , thereby substitution of (13.98) into the vertical component of Navier–Stokes equations gives:

$$0 = -\frac{\partial p_0}{\partial z} + (\bar{\rho} + \rho_0) g \quad (13.119)$$

Now, assuming  $\eta$  and  $\lambda$  constants in the Navier-Stokes equations, substituting the expressions (13.98) for the pressure and the density at the right-hand side of (13.16), using the approximation (13.118), and applying (13.119) we obtain:

$$\rho \dot{\mathbf{v}} = -\nabla p' + \eta \nabla^2 \mathbf{v} + \rho' g \mathbf{k} \quad (13.120)$$

This equation still contains the full density  $\rho$ . Substituting (13.98) and applying the approximation (13.116) leads to the following simplified equation of motion:

$$\dot{\mathbf{v}} = -\frac{1}{\bar{\rho}} \nabla p' + \bar{\nu} \nabla^2 \mathbf{v} + \frac{\rho' \epsilon}{\Delta \rho_0} g \mathbf{k} \quad (13.121)$$

where  $\bar{\nu} \equiv \eta/\bar{\rho}$  is the average (constant) kinematic viscosity. In this equation, the buoyancy term is the only place where the infinitesimal factor  $\epsilon$  is retained. This is a consequence of the fact that convective motions arise from buoyancy forces that are associated with fluctuations in the density field. Therefore, the acceleration  $|\partial \mathbf{v}/\partial t|$  must have the same order of magnitude of the acceleration associated with buoyancy:

$$\left| \frac{\partial \mathbf{v}}{\partial t} \right| \sim \left| \frac{\rho' \epsilon}{\Delta \rho_0} g \right| \quad (13.122)$$

or,

$$\left| \frac{\partial \mathbf{v}/\partial t}{(\rho' \epsilon / \Delta \rho_0)} \right| \sim g \quad (13.123)$$

Therefore, in a convective system the acceleration of gravity is always much greater than  $|\partial \mathbf{v}/\partial t|$ . Equation (13.121) can be simplified further taking the vertical component. If  $v$  is the vertical component of the velocity, then:

$$\frac{\partial v}{\partial t} + \mathbf{v} \cdot \nabla v = -\frac{1}{\bar{\rho}} \frac{\partial p'}{\partial z} + \bar{\nu} \nabla^2 v + \frac{\rho' \epsilon}{\Delta \rho_0} g \quad (13.124)$$

Now let us consider the first and the last terms at the right-hand side of (13.124). Using (13.114) and (13.102) we obtain:

$$\begin{aligned} -\frac{1}{\bar{\rho}} \frac{\partial p'}{\partial z} + \frac{\rho' \epsilon}{\Delta \rho_0} g &= -\frac{1}{\bar{\rho}} \left( \frac{\partial p'}{\partial z} - \rho' g \right) \\ &= -\frac{1}{\bar{\rho}} \left( \frac{\partial p'}{\partial z} - \bar{\rho} \bar{\beta} p' g \right) - \bar{\alpha} T' g \\ &= -\frac{1}{\bar{\rho}} \left( \frac{\partial p'}{\partial z} - \frac{1}{D} p' \right) - \bar{\alpha} T' g \end{aligned} \quad (13.125)$$

where:

$$D \equiv \frac{1}{\bar{\rho} \bar{\beta} g} \quad (13.126)$$

has the dimensions of a length and can be regarded as the thickness of a fluid layer with constant density  $\bar{\rho}$  and hydrostatic pressure  $p$  that varies from zero to  $1/\bar{\beta}$ . Substituting (13.119) in (13.99) for  $H_p$  gives:

$$\begin{aligned} H_p &= \left| \frac{1}{\bar{p}} \frac{dp_0}{dz} \right|^{-1} = \left| \frac{1}{\bar{p}} (\bar{\rho} + \rho_0) g \right|^{-1} \\ &= \left| \frac{\bar{p}}{(\bar{\rho} + \rho_0) g} \right| = \frac{\bar{p}}{\bar{\rho} g} \frac{1}{1 + \rho_0/\bar{\rho}} = \frac{\bar{p}}{\bar{\rho} g} + O(\epsilon) \end{aligned} \quad (13.127)$$

Therefore substituting into (13.126),

$$D = \frac{H_p}{\bar{\rho} \bar{\beta}} \quad (13.128)$$

In the Earth's upper mantle,  $\bar{\beta} \sim 10^{-12} \text{ Pa}^{-1}$  and  $\bar{p} \sim 10^{10} \text{ Pa}$ , thereby  $\bar{p} \bar{\beta} \sim 10^{-2}$ . Furthermore, we have that in any case  $p'/H \leq \partial p'/\partial z$ . Thus, the quantity:

$$\frac{1}{D} p' = \left( \frac{\bar{\rho} \bar{\beta}}{H_p} \right) p' < \left( \frac{\bar{\rho} \bar{\beta}}{H} \right) p' < \frac{1}{H} p' \quad (13.129)$$

is negligible compared to  $\partial p'/\partial z$ . Consequently, the Eq. (13.121) are simplified to:

$$\dot{\mathbf{v}} = -\frac{1}{\bar{\rho}} \nabla p' + \bar{\nu} \nabla^2 \mathbf{v} - \bar{\alpha} T' g \mathbf{k} \quad (13.130)$$

These are the equations of motion in the Boussinesq approximation. The same considerations used to derive (13.130) can

be used to further simplify (13.114). In fact, the buoyancy term associated with pressure fluctuations  $p'$  in (13.125) is small relative to the contribution coming from thermal fluctuations  $T'$ . Therefore, (13.114) can be rewritten as follows:

$$\frac{\rho'}{\bar{\rho}} \cong -\bar{\alpha}T' \quad (13.131)$$

This is the simplified equation of state in the Boussinesq approximation. Now let us turn to the equation of conservation of energy (13.40). Neglecting the viscous dissipation and internal heat production terms, this equations assumes the form:

$$\rho c_p \dot{T} - \alpha T \dot{p} = k \nabla^2 T \quad (13.132)$$

Substituting (13.98) for  $T$ , using the definition (13.2) of material derivative, and dividing by  $\rho c_p$  gives:

$$\dot{T}' + v \frac{\partial T_0}{\partial z} - \frac{\alpha T}{\rho c_p} \left( v \frac{\partial p_0}{\partial z} + \dot{p}' \right) = \kappa \nabla^2 T \quad (13.133)$$

The variation of the hydrostatic pressure  $p_0$  with depth is simply  $\rho g$ . The other term in brackets represents the time variation of pressure fluctuation and can be neglected. The other variable quantities in (13.133) are the term  $\alpha T$  at the left-hand side and the diffusivity  $\kappa = k/(\rho c_p)$  at the right-hand side (assuming  $k$  constant). To order  $O(\epsilon)$ , by (13.116) we can set:

$$\kappa \cong \bar{\kappa} = \frac{k}{\rho c_p} \quad (13.134)$$

Regarding the term  $\alpha T$ , by hypothesis  $H_T \geq H_p \gg H$ . Accordingly,

$$\left| \frac{1}{\bar{T}} \frac{dT_0}{dz} \right| \ll \frac{1}{H} \quad (13.135)$$

Then, integrating from  $z=0$  to  $z=H$  we conclude that:

$$\frac{\Delta T_0}{\bar{T}} = O(\epsilon) \quad (13.136)$$

where  $\Delta T_0$  is the maximum change of the conductive temperature variation  $T_0$  over the distance  $H$ . In agreement with (13.103) we also assume that:

$$\left| \frac{T'}{\bar{T}} \right| \leq O(\epsilon) \quad (13.137)$$

Therefore, because  $|T_0| \leq \Delta T_0$ , it follows that  $T_0/\bar{T} \leq O(\epsilon)$ . Now, using (13.111) we can write, to order  $O(\epsilon)$ :

$$\begin{aligned} \alpha T &= -\frac{T}{\rho} \left( \frac{\partial \rho}{\partial T} \right)_p = \bar{\alpha} \bar{T} \frac{\bar{\rho}}{\rho} \\ &= \bar{\alpha} \bar{T} \frac{\left( 1 + \frac{T_0}{\bar{T}} + \frac{T'}{\bar{T}} \right)}{\left( 1 + \frac{\rho_0}{\bar{\rho}} + \frac{\rho'}{\bar{\rho}} \right)} \cong \bar{\alpha} \bar{T} \end{aligned} \quad (13.138)$$

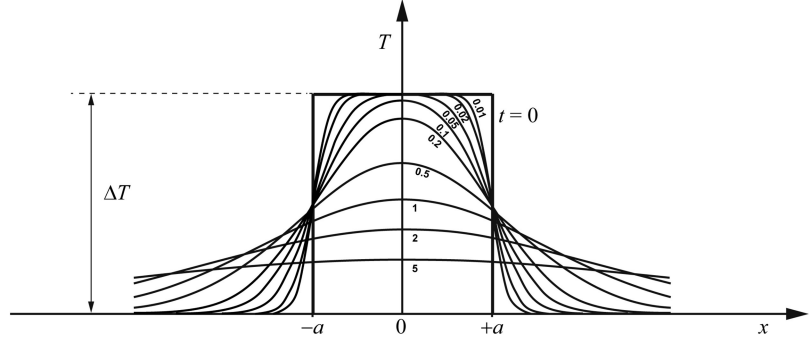
Substituting (13.138) and (13.134) into the energy Eq. (13.133) gives the final form of the conservation of energy in the Boussinesq approximation.

$$\dot{T}' + v \left( \frac{\partial T_0}{\partial z} - \frac{\bar{\alpha} \bar{T} g}{c_p} \right) = \bar{\kappa} \nabla^2 T' \quad (13.139)$$

The term in brackets at the left-hand side of this equation represents the static temperature gradient in excess of the adiabatic gradient (see Eq. 1.8). When using these equations, it should be ensured that the thickness  $H$  of the fluid layer be much less than the smallest among the scale heights (13.99). Furthermore, the fluctuations of temperature, pressure, and density induced by fluid motions should not exceed the total variations of these quantities under static conditions. We also note that temperature and velocity are coupled through the equations of motion (13.130) and the conservation of energy (13.139). This is a consequence of the fact that in thermal convection the velocity field is governed by temperature fluctuations, which in turn depend from velocity through the advection of heat.

Numerical solutions to the governing equations in the Boussinesq approximation have been found by several authors, among which we mention Turcotte and Oxburgh (1967), Richter

**Fig. 13.12** Time evolution of the temperature field within and around an infinite slab with initial thermal anomaly  $\Delta T$  and thickness  $2a$ . Numbers close to the curves are values of the dimensionless parameter  $\kappa t/a^2$



(1973), Richter and Parsons (1975), Jarvis and Peltier (1982), Bercovici et al. (1989), and Tackley (1998). The typical streamline pattern arising from these models is illustrated in Fig. 13.11. Here we will limit to do some simple physical considerations about the conditions for the onset of thermal convection in the upper mantle. Let us consider a spherical blob of radius  $a$  and characterized by a temperature excess or deficit  $\Delta T$  with respect to the surrounding mantle. By (12.58), such a thermal anomaly induces a density anomaly  $\Delta \rho = -\alpha \rho \Delta T$  that corresponds (by Archimedes' principle) to an additional buoyancy force given by:

$$\Delta f_b = -\alpha \rho \Delta T g \quad (13.140)$$

This body will sink or rise with increasing velocity until the viscous drag is balanced by the buoyancy (13.140). From this point on, the motion will proceed at constant velocity  $v_s$  given by:

$$v_s = -\frac{2}{9} a^2 \frac{\alpha \rho \Delta T}{\eta} g \quad (13.141)$$

This is a form of *Stokes' law* and  $v_s$  is referred to as the *terminal velocity*. Clearly, during its rise or fall the sphere either transfers or adsorbs heat by diffusion, thereby the thermal anomaly  $\Delta T$  cannot be a constant. To determine the velocity of this process, let us consider the simple case of an infinite slab with initial temperature anomaly  $\Delta T$ , placed in the region between  $x = -a$  and  $x = +a$  (Fig. 13.12). This is only an example of how the temperature can be distributed at time  $t = 0$ . In general, we have:

$$T(x, 0) = f(x) ; \quad -\infty < x < +\infty \quad (13.142)$$

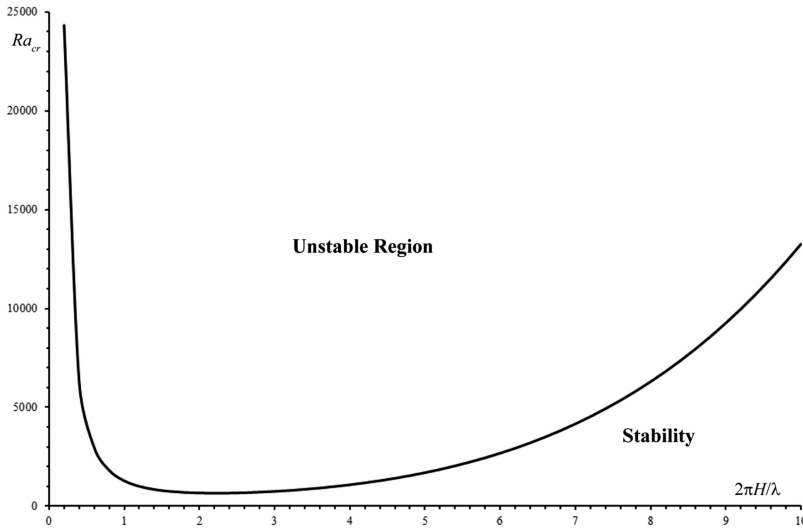
The solution to the diffusion Eq. (12.13) with the initial condition (13.142) is known as Laplace's solution:

$$T(x, t) = \frac{1}{2\sqrt{\pi \kappa t}} \int_{-\infty}^{+\infty} f(x') e^{-(x-x')^2/(4\kappa t)} dx' \quad (13.143)$$

Usually, the integral in (13.143) is evaluated numerically. However, when  $f(x)$  has the shape of a square pulse like that of Fig. 13.12, so that  $f(x) = 0$  for  $|x| > a$ , the integral reduces to a sum of two error functions (Carslaw and Jaeger 1959):

$$T(x, t) = \frac{1}{2} \Delta T \left[ \operatorname{erfc} \left( \frac{a-x}{2\sqrt{\kappa t}} \right) + \operatorname{erfc} \left( \frac{a+x}{2\sqrt{\kappa t}} \right) \right]; \quad -\infty < x < +\infty \quad (13.144)$$

The time evolution of the temperature field predicted by (13.144) is illustrated in Fig. 13.12. We note that for  $\kappa t/a^2 = 5$  the slab is not anymore thermically distinguishable from the surrounding region. Therefore, for  $t \geq \tau \equiv 5a^2/\kappa$  we can say that there is temperature equilibration. For example, for a 100 km thick slab sinking in the asthenosphere ( $a = 50$  km) we would have equilibration in  $\sim 340$  Myrs. Now let us come back to the anomalous blob that is rising or falling at velocity  $V_s$ . During the time interval  $\tau$ , the blob may travel a distance  $L = v_s \tau$ . Consequently, the free rising or falling of the blob is possible only for  $L \gg a$ , because if this condition is not



**Fig. 13.13** Critical Rayleigh number  $Ra_{cr}$  for the onset of Rayleigh-Bénard convection in a layer of thickness  $H$ , as a function of the dimensionless parameter  $2\pi H/\lambda$

satisfied the process of thermal equilibration is too fast for allowing motion under the combined effect of the buoyancy force and the viscous drag. This suggests that a single parameter can be used to measure the capacity of thermal anomalies to be transported by buoyancy forces through a fluid. In fact, using the expression (13.141) we see that the condition for thermal convection,  $L/a \gg 1$ , can be expressed as follows:

$$Ra = \frac{\alpha \rho a^3 g \Delta T}{\eta \kappa} = \frac{\alpha a^3 g \Delta T}{\nu \kappa} \gg 1 \quad (13.145)$$

where  $\nu$  is the kinematic viscosity and we have suppressed the multiplicative constant  $10/9 \approx 1$ . The dimensionless parameter  $Ra$  is called *Rayleigh number*. It represents the relative importance of the buoyancy forces with respect to viscous drag and rate of heat diffusion. However, the form (13.145) is not quite general to be useful, because here  $Ra$  depends from the blob radius  $a$ . Furthermore, it is still unclear how much greater than one should be  $Ra$  to trigger free thermal convection. In general, a linear stability analysis of the Boussinesq equations for a fluid layer of thickness  $H$ , heated from below and cooled from above, shows that a minimum *critical Rayleigh number*,  $Ra_{cr}$ , exists for the onset of thermal convection (e.g., Turcotte and Schubert 2002;

Ricard 2007). In this instance, the Rayleigh number and the critical value are defined as follows:

$$Ra = \frac{\overline{\alpha} \rho H^3 g \Delta T}{\eta \overline{\kappa}} = \frac{\overline{\alpha} H^3 g \Delta T}{\overline{\nu} \overline{\kappa}} \quad (13.146)$$

$$Ra_{cr} = \frac{\lambda^2 \left( \pi^2 + \frac{4\pi^2 H^2}{\lambda^2} \right)^3}{4\pi^2 H^2} \quad (13.147)$$

where  $\lambda$  is a characteristic wavelength of the thermal fluctuations. If  $Ra < Ra_{cr}$ , any fluctuation will decay with time, whereas for  $Ra > Ra_{cr}$  perturbations will grow exponentially with time. The critical Rayleigh number depends from the dimensionless parameter  $2\pi H/\lambda$  as illustrated in Fig. 13.13. For any fluctuation wavelength  $\lambda$ , if  $Ra$  lies above the curve, then the corresponding perturbation generates instability. Conversely, thermal convection is freezed for all wavelengths such that  $Ra < Ra_{cr}$ .

The curve of Fig. 13.13 shows that the stability curve has an absolute minimum. It is easy to find the value of  $2\pi H/\lambda$  such that the curve attains a minimum. This is given by:

$$\frac{2\pi H}{\lambda} = \frac{\pi}{\sqrt{2}} \quad (13.148)$$

which furnishes the following minimum for  $Ra_{cr}$ :

$$\min(Ra_{cr}) \cong 657.5 \quad (13.149)$$

An important feature of thermal convection is represented by the *aspect ratio* of the convective cells, which is the ratio of the horizontal width of the cells,  $w$ , to the vertical thickness  $H$  of the fluid layer. Linear stability analysis shows that the aspect ratio of the most rapidly growing fluctuations is  $w/H = \lambda/(2H) = \sqrt{2}$ . Therefore, upper mantle convective rolls should have a horizontal width of  $\sim 950$  km.

## Problems

1. Write and solve the equations of motion for a one-dimensional steady flow in the asthenosphere, considered as a two-layers Newtonian incompressible fluid. It is assumed that the upper asthenosphere has viscosity  $\eta_1 = 10^{20}$  Pa s and thickness  $h_1 = 200$  km, while the lower layer has viscosity  $\eta_2 = 10^{21}$  Pa s and thickness  $h_2 = 200$  km.  $v_0 = 100$  mm year<sup>-1</sup> is the velocity of the overlying lithosphere. Determine the depth of maximum velocity for a horizontal pressure gradient  $\partial p/\partial x = -10$  kPa km<sup>-1</sup>;
2. Determine the horizontal pressure gradient that was necessary to accelerate India to  $v_0 = 180$  mm year<sup>-1</sup> in the early Paleocene, assuming a 400 km thick asthenosphere and an average viscosity  $\eta = 10^{20}$  Pa s. Determine the time required to attain such a velocity  $v = 100$  mm/yr starting from  $v = 0$  mm year<sup>-1</sup>;
3. Let us assume that the lower boundary of a tectonic plate can be represented by a function  $z = f(x)$ . Assuming that no streamline has a cusp along this boundary, use the no-slip boundary condition to determine the boundary values of velocity for the asthenosphere along the LAB;
4. Repeat the previous exercise assuming that the lower boundary of a tectonic plate is represented by a surface  $z = f(x, y)$ ;
5. Determine the corner flows of a subduction zone assuming that the subducting plate is at rest with respect to the transition zone and that the overriding plate moves with velocity  $v_0$  in the positive  $x$  direction;
6. Prove the transformation Eq. (13.72);
7. Write the continuity equation for an incompressible fluid in polar coordinates;
8. Rewrite the equations of motion (13.69) in polar coordinates;
9. Find the velocity field  $(u, v)$  for a fluid that is moving in the positive (downward)  $z$  direction relative to a spherical object fixed at the origin of the reference frame. This is a Stokes' flow. To solve the problem, you must solve the equations found in exercises (7) and (8). Assume that the velocity field is a uniform field in the  $z$  direction with magnitude  $v$  as  $r \rightarrow \infty$ .

## References

- Batchelor GK (2000) An introduction to fluid dynamics. Cambridge University Press, Cambridge, UK, 615 pp
- Becker TW, Faccenna C (2011) Mantle conveyor beneath the Tethyan collisional belt. *Earth Planet Sci Lett* 310(3):453–461
- Bercovici D, Schubert G, Glatzmaier GA (1989) Three-dimensional spherical models of convection in the Earth's mantle. *Science* 244(4907):950–955
- Cande SC, Stegman DR (2011) Indian and African plate motions driven by the push force of the Réunion plume head. *Nature* 475:47–52. doi:10.1038/nature10174
- Carslaw HS, Jaeger JC (1959) Conduction of heat in solids, 2nd edn. Oxford University Press, Oxford, 510 pp
- Chase CG (1979) Asthenospheric counterflow: a kinematic model. *Geophys J R Astr Soc* 56:1–18
- Chorin AJ (1968) Numerical solution of the Navier-Stokes equations. *Math Comput* 22(104):745–762
- Chung TJ (2002) Computational fluid dynamics. Cambridge University Press, Cambridge, UK, 1012 pp
- Conder JA, Wiens DA (2007) Rapid mantle flow beneath the Tonga volcanic arc. *Earth Planet Sci Lett* 264(1):299–307
- Conder JA, Forsyth DW, Parmentier EM (2002) Asthenospheric flow and the asymmetry of the East Pacific Rise, MELT area. *J Geophys Res* 107(B12):2344. doi:10.1029/2001JB000807
- Conrad CP, Behn MD (2010) Constraints on lithosphere net rotation and asthenospheric viscosity from global mantle flow models and seismic anisotropy.

- Geochem Geophys Geosyst 11:Q05W05. doi:10.1029/2009GC002970
- Conrad CP, Behn MD, Silver PG (2007) Global mantle flow and the development of seismic anisotropy: differences between the oceanic and continental upper mantle. *J Geophys Res* 112, B07317. doi:10.1029/2006JB004608
- Conrad CP, Wu B, Smith EI, Bianco TA, Tibbetts A (2010) Shear-driven upwelling induced by lateral viscosity variations and asthenospheric shear: a mechanism for intraplate volcanism. *Phys Earth Planet Int* 178(3):162–175
- Ferziger JH, Perić M (2002) Computational methods for fluid dynamics, 3rd edn. Springer, Berlin, 423 pp
- Furbish DJ (1997) Fluid physics in geology. Oxford University Press, Oxford, 476 pp
- Höink T, Lenardic A (2010) Long wavelength convection, Poiseuille–Couette flow in the low-viscosity asthenosphere and the strength of plate margins. *Geophys J Int* 180(1):23–33
- Höink T, Jellinek AM, Lenardic A (2011) Viscous coupling at the lithosphere–asthenosphere boundary. *Geochem Geophys Geosyst* 12:Q0AK02. doi:10.1029/2011GC003698
- Jarvis GT, Peltier WR (1982) Mantle convection as a boundary layer phenomenon. *Geophys J Roy Astron Soc* 68(2):389–427
- King SD, Anderson DL (1998) Edge-driven convection. *Earth Planet Sci Lett* 160(3):289–296
- Natarov SI, Conrad CP (2012) The role of Poiseuille flow in creating depth-variation of asthenospheric shear. *Geophys J Int* 190(3):1297–1310
- Parmentier EM, Oliver JE (1979) A study of shallow global mantle flow due to the accretion and subduction of lithospheric plates. *Geophys J Royal Astr Soc* 57(1):1–21
- Phipps Morgan J, Morgan WJ, Zhang Y-S, Smith WHF (1995) Observational hints for a plume-fed, suboceanic asthenosphere and its role in mantle convection. *J Geophys Res* 100(B7):12753–12767. doi:10.1029/95JB00041
- Pozrikidis C (2009) Fluid dynamics, 2nd edn. Springer, Berlin, 773 pp
- Ranalli G (1995) Rheology of the earth, 2nd edn. Chapman & Hall, London, 413 pp
- Ribe NM, Christensen UR (1994) Three-dimensional modeling of plume–lithosphere interaction. *J Geophys Res* 99(B1):669–682. doi:10.1029/93JB02386
- Ricard Y (2007) Physics of mantle convection. In: Schubert G (ed) Treatise on geophysics, vol 7. Elsevier, Netherlands, pp 32–87
- Richter FM (1973) Convection and the large-scale circulation of the mantle. *J Geophys Res* 78(35):8735–8745. doi:10.1029/JB078i035p08735
- Richter FM, Parsons B (1975) On the interaction of two scales of convection in the mantle. *J Geophys Res* 80(17):2529–2541. doi:10.1029/JB080i017p02529
- Schubert G, Turcotte DL (1972) One-dimensional model of shallow-mantle convection. *J Geophys Res* 77(5):945–951. doi:10.1029/JB077i005p00945
- Schubert G, Turcotte DL, Olson P (2004) Mantle convection in the Earth and planets, 2nd edn. Cambridge University Press, Cambridge, 940 pp
- Sleep NH (1987) Lithospheric heating by mantle plumes. *Geophys J Roy Astron Soc* 91(1):1–11
- Sleep NH (1990) Hotspots and mantle plumes: some phenomenology. *J Geophys Res* 95(B5):6715–6736. doi:10.1029/JB095iB05p06715
- Spiegel EA, Veronis G (1960) On the Boussinesq approximation for a compressible fluid. *Astrophys J* 131:442–447
- Tackley PJ (1998) Three-dimensional simulations of mantle convection with a thermo-chemical basal boundary layer: D''? In: Gurnis M, Wyssession ME, Knittle E, Buffett BA (eds) The core–mantle boundary region, vol 28, Geodyn Series. AGU, Washington, DC, pp 231–253
- Tovish A, Schubert G, Luyendyk BP (1978) Mantle flow pressure and the angle of subduction: non-Newtonian corner flows. *J Geophys Res* 83(B12):5892–5898. doi:10.1029/JB083iB12p05892
- Turcotte DL, Oxburgh ER (1967) Finite amplitude convective cells and continental drift. *J Fluid Mech* 28(1):29–42
- Turcotte DL, Schubert G (2002) Geodynamics, 2nd edn. Cambridge University Press, Cambridge, UK, 456 pp
- Versteeg HK, Malalasekera W (2007) An introduction to computational fluid dynamics, 2nd edn. Pearson Education, Harlow, 503 pp
- Walker KT, Bokelmann GHR, Klemperer SL, Nyblade A (2005) Shear wave splitting around hotspots: evidence for upwelling-related mantle flow? In: Foulger GR, Natland JH, Presnall DC, Anderson DL (eds) Plates, plumes, and paradigms, Geological Society of America Special Paper 388., pp 171–192. doi:10.1130/2005.2388(11)

**Abstract**

This conclusive chapter introduces the Earth's gravity field and the concept of geopotential. The approach follows the potential field techniques presented in Chap. 4. The concepts of geoid and ellipsoid are discussed, along with the methods for processing gravity data. The chapter concludes with the important topic of dynamic topography, which links the geoid to mantle dynamics.

## 14.1 Gravity Field and Geopotential

In this chapter, we shall review some fundamental properties of the Earth's gravity field, which have a strong impact on plate tectonics research. Often, researchers involved in plate kinematics studies have a limited "contact area" with these topics, which essentially consists into the inspection of gravity anomaly maps with the purpose of better identifying fracture zone tracks, detect the presence of seamounts, or analyze the structural features of sedimentary basins. Conversely, a more in-depth understanding the Earth's gravity field is essential in many geodynamic studies, because the pressure gradients that are associated with asthenospheric flows and the upper mantle heterogeneities have a strong impact on the gravity anomalies at the Earth's surface and the shape of the Earth. In the following, we will apply the same potential field techniques discussed in Chap. 4 in the context of geomagnetism. Therefore, our starting point are the non-

relativistic equations of the Earth's gravity field  $\mathbf{g} = \mathbf{g}(\mathbf{r})$ :

$$\begin{cases} \nabla \cdot \mathbf{g} = 4\pi G \rho(\mathbf{r}) \\ \nabla \times \mathbf{g} = \mathbf{0} \end{cases} \quad (14.1)$$

where  $\rho = \rho(\mathbf{r})$  is the mass density and  $G$  is the gravitational constant:

$$G = 6.67259 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2} \quad (14.2)$$

The field equations (14.1) imply that  $\mathbf{g}$  is a potential field, thereby there exists a scalar function  $V = V(\mathbf{r})$  such that:

$$\mathbf{g} = \nabla V \quad (14.3)$$

The potential  $V = V(\mathbf{r})$  is called the *gravitational geopotential* and represents the fundamental quantity for the analysis of the field properties through spherical harmonic expansions. In fact, by (14.1) it satisfies Poisson's equation:

$$\nabla^2 V = 4\pi G \rho(\mathbf{r}) \quad (14.4)$$

Consequently,  $V$  is harmonic in any region  $\mathbf{R}$  where  $\rho = 0$ . According to the theorems proved in Sects. 4.7 and 4.8,  $V$  can be determined uniquely on the basis of a given set of boundary conditions. However, in geophysics it is also quite common calculating the potential or the field through direct integration, starting from the continuous mechanics version of Newton's law (see Eqs. 5.10 and 5.11). In the next sections, we shall use both these approaches to study the Earth's gravity field.

## 14.2 Spherical Harmonic Expansion of the Geopotential: The Geoid

A solution of Laplace's equation for the geopotential can be obtained using the techniques described in Sect. 4.8. The corresponding spherical harmonic expansion has the form:

$$V(r, \theta, \phi) = \frac{GM}{r} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^n \times \sum_{m=0}^n [a_{nm} \cos m\phi + b_{nm} \sin m\phi] \times P_{nm}(\cos \theta) ; r \geq R \quad (14.5)$$

where  $M$  is the Earth's mass,  $R$  is the Equatorial radius, and the coefficients  $a_{nm}$  and  $b_{nm}$  are called *Stokes' coefficients*. Just like in the case of the geomagnetic potential, specific orthogonality and normalization constraints are used for the surface spherical harmonics  $R_{nm}(\theta, \phi) \equiv P_{nm} \cos m\phi$  and  $S_{nm}(\theta, \phi) \equiv P_{nm} \sin m\phi$  (e.g., Heiskanen and Moritz 1993):

$$\int_0^{2\pi} d\phi \int_0^{\pi} R_{nm}(\theta, \phi) R_{sr}(\theta, \phi) \sin \theta d\theta = \frac{2\pi}{2n+1} \frac{(n+m)!}{(n-m)!} \delta_{ns} \delta_{mr} ; m \neq 0 \quad (14.6)$$

$$\int_0^{2\pi} d\phi \int_0^{\pi} S_{nm}(\theta, \phi) S_{sr}(\theta, \phi) \sin \theta d\theta = \frac{2\pi}{2n+1} \frac{(n+m)!}{(n-m)!} \delta_{ns} \delta_{mr} ; m \neq 0 \quad (14.7)$$

Furthermore,

$$\int_0^{2\pi} d\phi \int_0^{\pi} R_{nm}(\theta, \phi) S_{sr}(\theta, \phi) \sin \theta d\theta = 0 \text{ for any } n, m, r, s \quad (14.8)$$

$$\int_0^{2\pi} d\phi \int_0^{\pi} [R_{n0}(\theta)]^2 \sin \theta d\theta = \frac{4\pi}{2n+1} \quad (14.9)$$

The determination of the coefficients  $a_{nm}$  and  $b_{nm}$  in (14.5) can be accomplished multiplying both sides of this equation by some polynomial  $R_{rs}(\theta, \phi)$  or, respectively,  $S_{rs}(\theta, \phi)$ , and integrating over the unit sphere. Using the orthogonality relations (14.6), (14.7), (14.8), and (14.9) we obtain:

$$a_{nm} = \frac{2n+1}{2\pi} \frac{(n-m)!}{(n+m)!} \int_0^{2\pi} d\phi \int_0^{\pi} V(\theta, \phi) \times R_{nm}(\theta, \phi) \sin \theta d\theta ; m \neq 0 \quad (14.10)$$

$$a_{n0} = \frac{2n+1}{4\pi} \int_0^{2\pi} d\phi \int_0^{\pi} V(\theta, \phi) P_n(\cos \theta) d\theta \quad (14.11)$$

$$b_{nm} = \frac{2n+1}{2\pi} \frac{(n-m)!}{(n+m)!} \int_0^{2\pi} d\phi \int_0^{\pi} V(\theta, \phi) \times S_{nm}(\theta, \phi) \sin \theta d\theta ; m \neq 0 \quad (14.12)$$

These are the basic equations used to determine approximate values of the Stokes'

coefficients from observed values of gravity. The conventional normalization (14.6), (14.7), (14.8), and (14.9) is often substituted by the so-called *full normalization*, which allows a better manipulation of the spherical harmonics. In this instance, the surface harmonics  $R_{nm}(\theta, \phi)$  and  $S_{nm}(\theta, \phi)$  are replaced by the *fully normalized harmonics* defined as follows:

$$\left\{ \begin{array}{l} \bar{R}_{nm}(\theta, \phi) \\ \bar{S}_{nm}(\theta, \phi) \end{array} \right\} = \sqrt{2(2n+1) \frac{(n-m)!}{(n+m)!}} \times \left\{ \begin{array}{l} R_{nm}(\theta, \phi) \\ S_{nm}(\theta, \phi) \end{array} \right\} ; m \neq 0 \quad (14.13)$$

$$\bar{R}_{n0}(\theta) = \sqrt{2n+1} R_{n0}(\theta) = \sqrt{2n+1} P_n(\cos \theta) \quad (14.14)$$

With these definitions, the orthogonality relations (14.6), (14.7), (14.8), and (14.9) are then substituted by more canonical orthogonality relations:

$$\int_0^{2\pi} d\phi \int_0^\pi \bar{R}_{nm}(\theta, \phi) \bar{R}_{sr}(\theta, \phi) \sin \theta d\theta = 4\pi \delta_{ns} \delta_{mr} \quad (14.15)$$

$$\int_0^{2\pi} d\phi \int_0^\pi \bar{S}_{nm}(\theta, \phi) \bar{S}_{sr}(\theta, \phi) \sin \theta d\theta = 4\pi \delta_{ns} \delta_{mr} \quad (14.16)$$

Therefore, the average square of fully normalized harmonics over the unit sphere is unity, whether or not  $m$  is zero. Also the coefficients of the spherical harmonic expansion have now more simple expressions:

$$\bar{a}_{nm} = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi V(\theta, \phi) \bar{R}_{nm}(\theta, \phi) \sin \theta d\theta \quad (14.17)$$

$$\bar{b}_{nm} = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi V(\theta, \phi) \bar{S}_{nm}(\theta, \phi) \sin \theta d\theta \quad (14.18)$$

Stokes' coefficients are determined empirically from the analysis of orbital perturbations

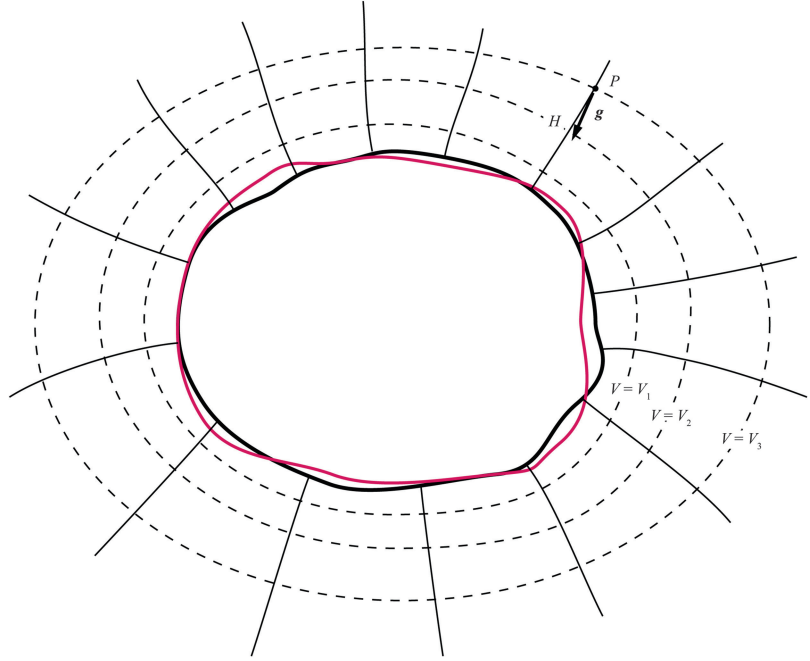
of artificial satellites and updated regularly using terrestrial data. Examples of recent geopotential models are GEM-T3 (Lerch et al. 1994), GRIM4 (Schwintzer et al. 1997), GRIM5 (Biancale et al. 2000), GGM02 (Tapley et al. 2005), and EGM2008 (Pavlis et al. 2012). These models can be used to build reliable representations of the Earth's gravity field in a variety of technical and scientific applications. Starting from a geopotential model, it is possible to build equipotential surfaces that locally define the vertical direction. Among the infinitely many equipotential surfaces of the Earth's gravity field, the one passing through the mean surface of the oceans (removing the effect of tides) is used as a reference surface for measuring elevation. This surface is the *geoid* and represents the mathematical figure of the Earth as determined by the density distribution in the Earth's interior and by rotation. The (usually) curved lines that intersect orthogonally any equipotential surface are called *plumb lines*. As illustrated in Fig. 14.1, given two values  $V_1$  and  $V_2$  for the potential, the corresponding equipotential surfaces  $V(\mathbf{r}) = V_1$  and  $V(\mathbf{r}) = V_2$  are not generally parallel each other. This fact leads to a curious paradox. Let us consider two near points at the Earth's surface,  $A$  and  $B$ . Now let us assume that a level placed at an intermediate location  $C$  indicates that the two points lie on the same horizontal surface. In this instance, they should have the same altitude.

However, the effective altitude of the two points relative to the geoid will be in most cases slightly different, depending on the shape of the plumb lines. In fact the altitude, more precisely the *orthometric height*  $H$ , of a point  $P$  is defined as the length of the plumb line between the actual location of  $P$  and the geoid. For this reason, a precise determination of orthometric altitudes requires in general a combination of gravity measurements and optical leveling.

## 14.3 Geoid and Ellipsoid

The shape of the geoid is determined by the distribution of masses in the Earth, especially from lateral density variations in the Earth's mantle.

**Fig. 14.1** Geoid (*thick black line*), topography (*red line*), plumb lines (*thin black lines*), and level surfaces of the gravity potential (*dashed lines*).  $H$  is the orthometric height of a point  $P$



A simple method for describing the geometry of the Earth's geoid is to specify its undulations with respect to a reference regular surface that would represent the overall shape of the Earth in absence of lateral density variations. To a first approximation, the Earth and other planetary bodies are rigid rotating objects having spherical symmetry. The main factor controlling the departure from spherical symmetry of a homogeneous deformable body that is rotating about a spin axis is the combined effect of the gravity field and the centrifugal force. Let us start from a homogeneous body that is at rest in an inertial reference frame. The equilibrium shape of this body is a sphere of radius  $R$ , and its gravity potential for  $r \geq R$  is given by:

$$V_0(\mathbf{r}) = \frac{GM}{r} \quad (14.19)$$

Now let us assume that this body is put in motion instantaneously, and that this motion consists of a rotation about a fixed spin axis  $\mathbf{n}$  with constant angular velocity  $\Omega$ . In this instance, a test unit mass at the surface of the body would be

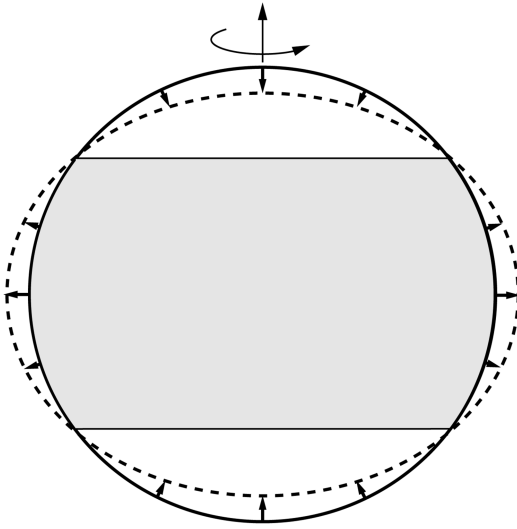
subject to the combined effect of the gravity force  $\mathbf{g}$  and the centrifugal force  $\mathbf{f}_c$ :

$$\mathbf{f}(\mathbf{r}) = \mathbf{g}(\mathbf{r}) + \mathbf{f}_c(\mathbf{r}) = \mathbf{g}(\mathbf{r}) - \Omega \times (\Omega \times \mathbf{r}) \quad (14.20)$$

where  $\Omega = \Omega \mathbf{n}$ . This force depends from the colatitude through a factor  $\sin\theta$ , thereby the surface of the body is not anymore an equipotential surface. Consequently, the body will start deforming to adapt its shape to the new level surfaces. For  $r = R$ , the potential  $U$  associated with the force field (14.20) is initially given by the sum of the gravity potential,  $V_0(R)$  plus the potential associated with the centrifugal force,  $W(\theta)$ :

$$\begin{aligned} U(\theta) &= V_0(R) + W(\theta) = \frac{GM}{R} + \frac{1}{2}\Omega^2 R^2 \sin^2\theta \\ &= \frac{GM}{R} - \frac{1}{3}\Omega^2 R^2 [P_2(\cos\theta) - 1] \end{aligned} \quad (14.21)$$

This expression suggests the shape that the body should acquire to adapt its surface to an equipotential surface of the combined field (14.20). In fact, by (14.21) we have that:



**Fig. 14.2** Mass redistribution (arrows) in a homogeneous deformable body once that it is put in motion about a spin axis. The grey zone indicates the area for which  $U(\theta) \geq \langle U \rangle$

$$U(0) = U(\pi) = \frac{GM}{R} ; U(\pi/2) = \frac{GM}{R} + \frac{1}{2}\Omega^2 R^2 \quad (14.22)$$

The excess potential along the Equator at distance  $r = R$  from the centre implies that the body must redistribute its mass, pushing material away from the centre along the equatorial belt and flattening at the poles, as illustrated in Fig. 14.2. Consequently, it will acquire an oblate form. However, such an adjustment will change the gravity potential  $V = V(\mathbf{r})$ , because the new shape does not have anymore a spherical symmetry. It is quite intuitive that a process would start such that the new potential first adds a zonal term  $P_2(\cos\theta)$ , which determines in turn an additional  $P_4(\cos\theta)$  term in  $U(\theta)$ , that controls now a new change of shape, and so on.

The final shape of the body will be that of an *oblate spheroid* whose potential is an infinite series of zonal harmonics of even degree:

$$U(r, \theta) = \frac{GM}{r} \left[ 1 - \sum_{n=1}^{\infty} \left( \frac{R}{r} \right)^{2n} J_{2n} P_{2n}(\cos\theta) \right] - \frac{1}{3}\Omega^2 r^2 [P_2(\cos\theta) - 1] \quad (14.23)$$

where the zonal coefficients  $J_{2n}$  are analogous to the Stokes' coefficients of the spherical harmonic expansion (14.5). If  $a$  and  $b$  are the equatorial and polar radii, respectively, then the *flattening*  $f$  will be given by:

$$f = \frac{a - b}{a} \quad (14.24)$$

To apply this model to the Earth, we must assume that the reference figure of the Earth is a *level spheroid*, that is, an ellipsoid of revolution which is an equipotential surface of a *normal gravity field*. In this case, the geoid representing the actual figure of the Earth will be an equipotential surface that deviates from the reference spheroid because of lateral heterogeneity in the mass distribution within the Earth. To express the flattening of the level spheroid in terms of Stokes' coefficients, let us consider the approximate equation of a spheroid, which allows to write the Earth's radius,  $R$ , as a function of the colatitude:

$$R(\theta) \cong a(1 - f\cos^2\theta) \quad (14.25)$$

Substituting  $r$  in the spherical harmonic expansion (14.23) by this expression, gives the constant potential  $U_0$  of the reference ellipsoid. To this purpose, we will consider the second-order approximation of the geopotential (14.23) with  $R = a$ , which is:

$$U_0 \cong \frac{GM}{R(\theta)} \left[ 1 - \frac{a^2}{R^2(\theta)} J_2 P_2(\cos\theta) \right] - \frac{1}{3}\Omega^2 R^2(\theta) [P_2(\cos\theta) - 1] \quad (14.26)$$

If our assumption that the reference figure of the Earth is a level spheroid is correct, then this equation will be satisfied for any colatitude  $\theta$ . By hypothesis, the flattening  $f$  is a small quantity. Therefore, the following approximations are applicable:

$$\begin{aligned}\frac{1}{R(\theta)} &\cong \frac{1}{a} (1 + f \cos^2 \theta); \quad \frac{1}{R^2(\theta)} \\ &\cong \frac{1}{a^2} (1 + 2f \cos^2 \theta); \quad R^2(\theta) \\ &\cong a^2 (1 - 2f \cos^2 \theta)\end{aligned}\quad (14.27)$$

Substituting into (14.26) and dividing both sides by  $a/GM$  gives:

$$\begin{aligned}\frac{aU_0}{GM} &= (1 + f \cos^2 \theta) \\ &\times \left[ 1 - \frac{J_2}{2} (1 + 2f \cos^2 \theta)(3 \cos^2 \theta - 1) \right] \\ &- \frac{m}{2} (1 - 2f \cos^2 \theta)(3 \cos^2 \theta - 1)\end{aligned}\quad (14.28)$$

where the quantity:

$$m \equiv \frac{\Omega^2 a^3}{GM} \quad (14.29)$$

represents the ratio between centrifugal potential and gravitational potential for a sphere and is termed *rotation parameter*. Neglecting all the quadratic terms in  $f^2$ ,  $mf$ , and  $fJ_2$  in (14.28) we obtain:

$$\frac{aU_0}{GM} = 1 - \frac{3}{2} J_2 \cos^2 \theta + f \cos^2 \theta + \frac{m}{2} - \frac{1}{2} m \cos^2 \theta \quad (14.30)$$

To obtain an equation that is independent from  $\theta$ , it is necessary that the sum of all terms in  $\cos^2 \theta$  be zero. Therefore, the following relation must be satisfied:

$$f = \frac{1}{2} (3J_2 + m) \quad (14.31)$$

Alternatively, assigning the flattening  $f$ , the parameter  $J_2$  must be calculated as follows:

$$J_2 = \frac{1}{3} (2f - m) \quad (14.32)$$

The Stokes' parameter  $J_2$  plays a fundamental role in geodesy and is termed *dynamic form factor* or *ellipticity coefficient*. It can be determined by observation of satellite orbits. In fact, by (14.23), to the second order the normal gravity potential at distance  $r$  from the Earth's centre is given by:

$$\begin{aligned}U(r, \theta) &\cong \frac{GM}{r} - \frac{GMa^2}{r^3} J_2 P_2(\cos \theta) \\ &- \frac{1}{3} \Omega^2 r^2 [P_2(\cos \theta) - 1]\end{aligned}\quad (14.33)$$

While the first term in (14.33) maintains a satellite on an elliptical orbit, the second one introduces a precession in the satellite orbit, at a rate that depends precisely on  $J_2$ . Therefore, measuring the precession rate gives a measure of the dynamic form factor. A recent estimate is:

$$J_2 = 1.081874 \times 10^{-3} \quad (14.34)$$

The *World Geodetic System 1984* (WGS84) represents the most recent consistent set of constants and model parameters for the definition of the normal gravity and the reference ellipsoid. In this reference system, the basic parameters assume the values listed in Table 14.1.

The WGS84 reference can be used to calculate the parameters of the theoretical normal gravity associated with the ellipsoid. It can be shown (e.g., Heiskanen and Moritz 1993) that the gravity field on the ellipsoid is given by:

$$\gamma(\psi) = \frac{a\gamma_a \cos^2 \psi + b\gamma_b \sin^2 \psi}{\sqrt{a^2 \cos^2 \psi + b^2 \sin^2 \psi}} \quad (14.35)$$

In this formula, which is known as *formula of Somigliana*,  $\psi$  is the *geodetic latitude*, that is, the

angle between the normal to the ellipsoid and the equatorial plane, while  $\gamma_a$  and  $\gamma_b$  are respectively the equatorial and polar normal gravity. In the WGS84, these parameters are given by:

$$\begin{aligned}\gamma_a &= 9.7803253359 \text{ ms}^{-2} \\ \gamma_b &= 9.8321849378 \text{ ms}^{-2}\end{aligned}\quad (14.36)$$

while  $b$  can be calculated using (14.24) and the reference values of Table 14.1:

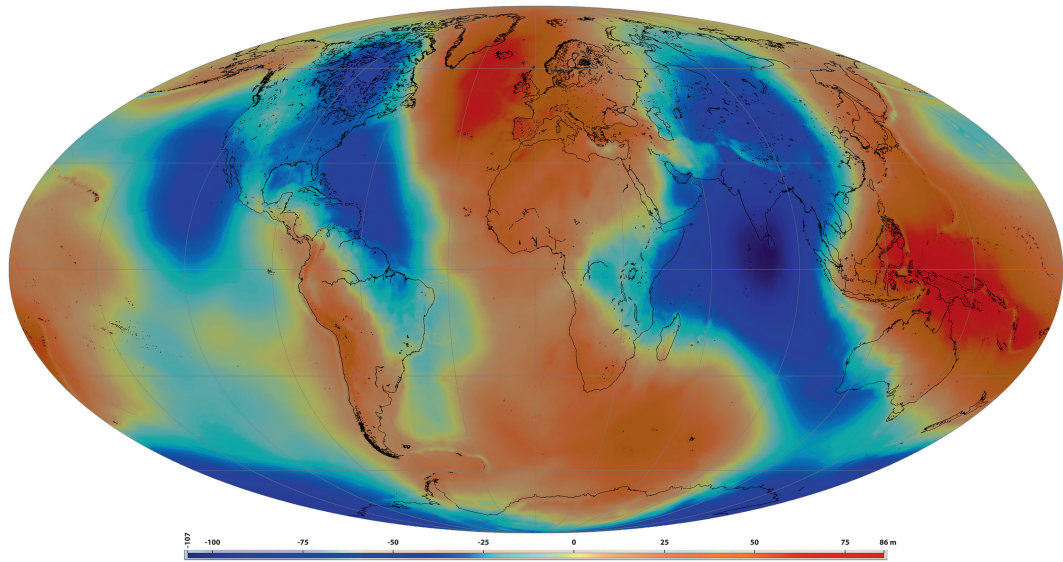
$$b = a(1 - f) = 6,356,752.3 \text{ m} \quad (14.37)$$

**Table 14.1** Basic reference values in the WGS84 reference frame

Quantity	Symbol	Value	Units
Semi-major axis	$a$	6,378,137.0	m
Geocentric gravitational constant (mass of Earth's atmosphere included)	$GM$	$3.986004418 \times 10^{14}$	$\text{m}^3/\text{s}^2$
Reciprocal of flattening	$1/f$	298.257223563	
Mean angular velocity	$\Omega$	$7,292,115 \times 10^{-11}$	$\text{s}^{-1}$
Theoretical normal potential of the ellipsoid (derived)	$U_0$	62,636,860.8497	$\text{m}^2 \text{ s}^{-2}$

The formula of Somigliana (14.35) predicts an average latitudinal variation of  $\gamma(\psi)$  of  $\sim 0.8 \text{ mGal/km}$  in N–S direction ( $1 \text{ Gal} = 10^{-2} \text{ m/s}^2$ ).

Now we are ready to give an intuitive representation of the Earth's figure through a comparison of the geoid with the reference ellipsoid. In fact, any departure from the geometrical simplicity of the ellipsoid of revolution must be associated with localized variations of density in the Earth's interior. In particular, a local excess of density in the mantle or in the lithosphere tends to increase the gravity in that area, possibly leading to an upward displacement of the geoid relative to the ellipsoid. Conversely, a mass deficit could cause a decrease of gravity and a consequent downward bending of the geoid. The method for calculating these *geoid undulations* (or *geoid anomalies*) is rather complex and goes beyond the scope of this book. The interested reader can find an exhaustive description of the technique in the classic book of Heiskanen and Moritz (1993), while a modern approach to the problem can be found in Rapp (1997). The undulations of the recent geoid model EGM2008 (Pavlis et al. 2012) are illustrated in Fig. 14.3. They range from  $-107 \text{ m}$  in the central Indian Basin to  $+86 \text{ m}$



**Fig. 14.3** Geoid undulations relative to the WGS84 ellipsoid in the EGM2008 model (Pavlis et al. 2012)

in the area north of Australia. With the exception of convergence zones, there is poor correlation of geoid undulations with plate boundaries.

However, in most cases subduction zones are associated with geoid highs that can be explained by the presence of dense slabs below the arcs. Significant examples are the Peru–Chile, Aleutine, and Tonga subduction zones (Fig. 14.3). More difficult is to explain the negative undulations of N. America and central Asia, although the geoid low over the Hudson Bay in Canada may be associated with postglacial rebound. In general, long-wavelength ( $>4,000$  km) anomalies of the geoid are mainly associated with large-scale density anomalies and flows in the lower mantle, while short-wavelength features correlate with lithospheric structure (Hager and Richards 1989 and references therein). Interestingly, Dziewonski et al. (1977) observed that the long-wavelength lows correlated with seismically fast, presumably cold and dense, regions of the lower mantle, while long-wavelength geoid highs correlated with seismically slow, possibly light, lower mantle features. Moreover, the geoid anomaly results to be positive for the regions surrounding many hot spots, although these features are located on low-density mantle. Clearly, this is the reverse of what would be expected, because in principle the geoid should exhibit a positive correlation with internal density anomalies. The solution of this apparent paradox can be found in a pioneer work of Pekeris (1935), who focused on thermal convection but showed that the geoid results from the combined effect of density distribution in the Earth's interior and a process today known as *dynamic topography*, which will be the subject of the last section of this chapter. It is responsible for topographic uplift over hot upwelling currents and subsidence over cold downwellings. Pekeris (1935) showed that the density anomalies generated by dynamic topography close to the Earth's surface give a contribution to the geoid that opposes and eventually overcomes the contribution of deep sources, giving geoid highs over low-density mantle upwellings and lows over high-density downwellings.

## 14.4 MacCullagh's Formula

The potential of a planetary mass, in particular of the Earth, can be calculated using the classic Poisson integral, which can be obtained either integrating (14.1) or simply extending Newton's gravity law to a continuous body. Using the geometry shown in Fig. 14.4, we have for the potential generated by a density distribution in a region  $\mathbf{R}$ :

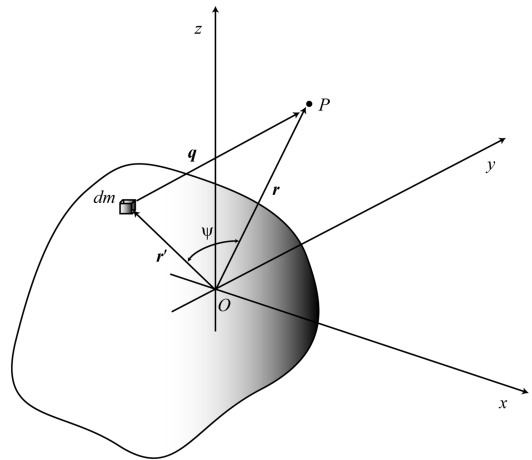
$$V(\mathbf{r}) = G \int_{\mathbf{R}} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad (14.38)$$

The term  $|\mathbf{r} - \mathbf{r}'|$  in the integral (14.38) can be expanded in power series. In fact, if  $z$  is a real variable such that  $-1 \leq z < 1$ , then McLaurin's expansion of the function  $f(z) = (1 - z)^{-1/2}$  gives:

$$\begin{aligned} f(z) = (1 - z)^{-1/2} &= 1 + \frac{1}{2}z + \frac{1}{2!} \frac{1}{2} \frac{3}{2} z^2 \\ &+ \frac{1}{3!} \frac{1}{2} \frac{3}{2} \frac{5}{2} z^3 + \dots ; \quad -1 \leq z < 1 \end{aligned} \quad (14.39)$$

Therefore, setting  $z \equiv 2x\zeta - \zeta^2$  we obtain that the function:

$$G(x, \zeta) = \frac{1}{\sqrt{1 - 2x\zeta + \zeta^2}} \quad (14.40)$$



**Fig. 14.4** Geometry for the calculation of Poisson's integral at a point  $P$

can be written as follows:

$$\begin{aligned}
 G(x, \zeta) &= 1 + \frac{1}{2}(2x\zeta - \zeta^2) + \frac{3}{8}(2x\zeta - \zeta^2)^2 \\
 &\quad + \frac{5}{16}(2x\zeta - \zeta^2)^3 + \dots \\
 &= 1 + x\zeta + \left(\frac{3}{2}x^2 - \frac{1}{2}\right)\zeta^2 \\
 &\quad + \left(\frac{5}{2}x^3 - \frac{3}{2}x\right)\zeta^3 + \dots \\
 &= \sum_{n=0}^{\infty} \zeta^n P_n(x) \quad (14.41)
 \end{aligned}$$

Consequently, the function  $G$  seems to have the capability to “generate” Legendre polynomials when it is expanded in power series. In this sense, it is called the *generating function* for these polynomials. Now, using (14.40) we see that at sufficiently large distance from the body it is possible to write:

$$\begin{aligned}
 |\mathbf{r} - \mathbf{r}'|^{-1} &= (r^2 + r'^2 - 2rr' \cos \psi)^{-1/2} \\
 &= \frac{1}{r} \left(1 + (r'/r)^2 - 2(r'/r) \cos \psi\right)^{-1/2} \\
 &= \frac{1}{r} G(\cos \psi, r/r') \quad (14.42)
 \end{aligned}$$

Therefore, from (14.41) we have that the Poisson integral (14.38) can be expressed as follows:

$$\begin{aligned}
 V(\mathbf{r}) &= G \int_{\mathbf{R}} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dx' dy' dz' \\
 &= \frac{G}{r} \sum_{n=0}^{\infty} \int_{\mathbf{R}} \rho(\mathbf{r}') \left(\frac{r'}{r}\right)^n \\
 &\quad \times P_n(\cos \psi) dx' dy' dz' \quad (14.43)
 \end{aligned}$$

Now let us apply these concepts to the Earth and assume that the reference frame is a geocentric frame, so that the origin coincides with the centre of mass of the Earth (see Sect. 2.3). The first three terms of the series (14.43) give a good approximation of the real field at large distances from the Earth. The  $n=0$  term is evidently the

potential (14.19) generated by a homogeneous sphere. Regarding the term for  $n=1$ , noting that  $r' \cos \psi = \hat{\mathbf{r}} \cdot \mathbf{r}'$ , we have:

$$\begin{aligned}
 V_1(\mathbf{r}) &= \frac{G}{r^2} \int_{\mathbf{R}} \rho(\mathbf{r}') r' \cos \psi dx' dy' dz' \\
 &= \hat{\mathbf{r}} \cdot \frac{G}{r^2} \int_{\mathbf{R}} \rho(\mathbf{r}') \mathbf{r}' dx' dy' dz' = 0 \quad (14.44)
 \end{aligned}$$

In fact, the last integral in (14.44) represents the centre of mass of the Earth. Finally, let us consider the contribution for  $n=3$ . Recalling (2.20) and (2.24), we have:

$$\begin{aligned}
 V_2(\mathbf{r}) &= \frac{G}{2r^3} \int_{\mathbf{R}} \rho(\mathbf{r}') r'^2 (3\cos^2 \psi - 1) dx' dy' dz' \\
 &= \frac{G}{r^3} \int_{\mathbf{R}} \rho(\mathbf{r}') r'^2 dx' dy' dz' \\
 &\quad - \frac{3G}{2r^3} \int_{\mathbf{R}} \rho(\mathbf{r}') r'^2 \sin^2 \psi dx' dy' dz' \\
 &= \frac{G}{r^3} \int_{\mathbf{R}} \rho(\mathbf{r}') r'^2 dx' dy' dz' \\
 &\quad - \frac{3G}{2r^3} \int_{\mathbf{R}} \rho(\mathbf{r}') r'^2 (1 - \cos^2 \psi) dx' dy' dz' \\
 &= \frac{G}{2r^3} (I_{xx} + I_{yy} + I_{zz}) - \frac{3G}{2r^3} \int_{\mathbf{R}} \rho(\mathbf{r}') \\
 &\quad \times [r'^2 - (\hat{\mathbf{r}} \cdot \mathbf{r}')^2] dx' dy' dz' \\
 &= \frac{G}{2r^3} [(A + B + C) - 3I(\hat{\mathbf{r}})] \quad (14.45)
 \end{aligned}$$

where  $A \equiv I_{xx}$ ,  $B \equiv I_{yy}$ , and  $C \equiv I_{zz}$  are the moments of inertia about the three coordinate axes and  $I(\hat{\mathbf{r}})$  is the moment of inertia about the axis of  $\mathbf{r}$ . Consequently, considering only the first three terms of the expansion (14.43) we have the following approximate *formula of MacCullagh* for the potential:

$$V(\mathbf{r}) \cong \frac{GM}{r} + \frac{G}{2r^3} [(A + B + C) - 3I(\hat{\mathbf{r}})] \quad (14.46)$$

This formula reveals the role of the moments of inertia for the determination of the Earth's gravity field. Assuming that the Earth can be approximated by an oblate spheroid and that the  $z$  axis coincides with the rotation axis, we have that  $A = B$  and the inertial tensor is diagonal, thereby by (2.26) the moment of inertia  $I(\hat{\mathbf{r}})$  can be expressed as follows:

$$\begin{aligned} I(\hat{\mathbf{r}}) &= A(\hat{r}_x^2 + \hat{r}_y^2) + C\hat{r}_z^2 \\ &= A(1 - \hat{r}_z^2) + C\hat{r}_z^2 = A + (C - A)\hat{r}_z^2 \\ &= A + (C - A)\cos^2\theta \end{aligned} \quad (14.47)$$

where  $\theta$  is the colatitude of  $\mathbf{r}$ . Therefore, assuming axial symmetry we have that MacCullagh's formula (14.46) reduces to:

$$V(\mathbf{r}) \cong \frac{GM}{r} - \frac{G}{r^3} (C - A) P_2(\cos\theta) \quad (14.48)$$

Comparing this expression with the spherical harmonic expansion (14.23), we obtain an expression for the first coefficients  $J_n$ :

$$J_0 = 1 ; J_1 = 0 ; J_2 = -\frac{C - A}{MR^2} \quad (14.49)$$

Therefore, MacCullagh's formula allows to determine the excess moment of inertia about the spin axis relative to the moment of inertia about an equatorial axis, associated with the Earth's flattening.

## 14.5 Gravity Measurements and Reduction of Gravity Data

Gravity measurements are made through *gravimeters* close to the Earth's surface, by aircraft, ships, or land surveys. Gravimeters may be based on precise spring balances, pendulums

or small bodies falling in a vacuum. In the latter case, a test mass is dropped inside a vacuum chamber for some centimeters and the trajectory is monitored accurately using a laser interferometer and an atomic clock. The precision of these instruments is of a few  $\mu\text{Gal}$  ( $10^{-8} \text{ ms}^{-2}$ ) in static conditions. In general, the observed gravity results from the superposition of several factors, which must be accurately separated to obtain a data set that can be interpreted geologically. In fact, the final objective of most gravity surveys is to determine *gravity anomalies* associated with short-wavelength lateral density variations close to the Earth's surface and to reconstruct the density distribution in the underground. Therefore, the contribution of these small-scale density anomalies must be isolated in the gravity measurements from other factors that influence the variability of the observed data. These factors include:

- The variations of latitude, which affect the formula of Somigliana (14.35);
- The effect of elevation above the sea level;
- The average crustal mass above the sea level;
- The lateral variations of altitude about the average mass above the sea level;
- Periodic tidal changes in the shape of the Earth;
- Variations of the centrifugal potential caused by the motion of the gravimeter, for example when the measurements are made by aircrafts or ships. This is called *Eötvös effect*;

The reduction of raw gravity data to a format that can be interpreted in terms of density anomalies is performed through a series of corrections. In small-scale studies, the difference between geoid and reference ellipsoid is usually neglected, thereby the orthometric altitude  $H$  relative to the geoid is assumed to be coincident with the *geometric altitude*  $h$  relative to the ellipsoid. While shipboard gravity data can be compared directly with the normal gravity  $\gamma_0$ , measurements performed at elevation  $h$  above the ellipsoid must be adjusted to take into account of the variation of gravity with altitude. The normal gravity at a small altitude  $h$  above the reference surface can be expressed as a Taylor's series expansion about

the distance  $R$  of the ellipsoid from the Earth's centre:

$$\gamma(R+h) = \gamma_0 + h \left. \frac{\partial \gamma}{\partial r} \right|_{r=R} + \frac{1}{2} h^2 \left. \frac{\partial^2 \gamma}{\partial r^2} \right|_{r=R} + \dots \quad (14.50)$$

Dropping high-order terms and rearranging gives:

$$\gamma_0 \cong \gamma(R+h) - h \left. \frac{\partial \gamma}{\partial r} \right|_{r=R} \quad (14.51)$$

To first order, the derivative in (14.51) can be estimated as follows:

$$\begin{aligned} \left. \frac{\partial \gamma}{\partial r} \right|_{r=R} &\cong \left. \frac{\partial}{\partial r} \left( \frac{GM}{r^2} \right) \right|_{r=R} = -2 \frac{GM}{r^3} \Big|_{r=R} \\ &= -\frac{2\gamma_0}{R} \end{aligned} \quad (14.52)$$

Therefore, substituting into (14.51) gives:

$$\gamma_0 \cong \gamma(R+h) + \frac{2\gamma_0 h}{R} \quad (14.53)$$

The quantity:

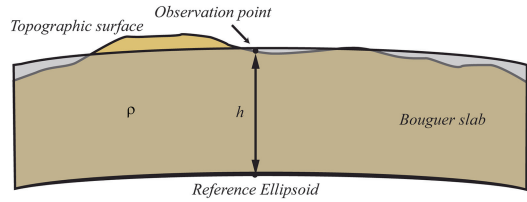
$$\delta\gamma_{FA} = \gamma_0 - \gamma(R+h) = \frac{2\gamma_0 h}{R} \cong 0.3086 \times 10^{-5} h \quad (14.54)$$

is called *free-air correction* and represents the adjustment that is required to take into account of the elevation of the observation point. A *free-air anomaly* is calculated from a raw gravity datum subtracting the theoretical gravity at altitude  $h$  from the observed value at geodetic coordinates  $(\psi, \phi)$  and elevation  $h$ :

$$\Delta g_{FA}(\psi, \phi, h) = g(\psi, \phi, h) - \gamma_h(\psi) \quad (14.55)$$

From (14.54), we have that these anomalies can be calculated using the following canonical expression:

$$\begin{aligned} \Delta g_{FA}(\psi, \phi, h) &= g(\psi, \phi, h) - \gamma_0(\psi) \\ &\quad + 0.3086 \times 10^{-5} h \end{aligned} \quad (14.56)$$



**Fig. 14.5** Bouguer slab for the determination of simple Bouguer anomalies

Free-air anomalies are usually employed in marine geophysics to analyze the tectonic structure of the sea floor, especially in oceanic regions, where they furnish accurate representations of spreading ridges, fracture zones, etc. (e.g., see Fig. 2.4). On land, these anomalies show strong correlation with topography, because they do not take into account of the presence of masses between the observation point and sea level. In some sense, free-air anomalies show the gravity anomalies that would be observed if all the masses were condensed within the geoid. The most simple correction that is performed to take into account of the presence of masses above the geoid is called *simple Bouguer correction*. Let us assume that the topography of a region is sufficiently smooth (Fig. 14.5). In this instance, we can approximate the mass above the geoid by a homogeneous, infinitely extended slab of constant thickness. The excess gravity generated by an infinite slab of thickness  $h$  and uniform density  $\rho$  can be calculated starting from a thin disk of thickness  $\delta h$  and radius  $a$ . To simplify the calculation, it is convenient to choose a reference frame with the origin coinciding with the disk centre and with a  $z$  axis coinciding with the symmetry axis of the disk. In this instance, the volume element in the Poisson integral (14.38) can be expressed in cylindrical coordinates as:  $dV' = dx' dy' dz' = r' dr' d\phi' dz'$ .

Therefore, the potential at an arbitrary location  $\mathbf{r} = z\mathbf{k}$  along the symmetry axis will be given by:

$$\begin{aligned} V(\mathbf{r}) &= G\rho \int_{\mathbf{R}} \frac{1}{|\mathbf{r} - \mathbf{r}'|} dx' dy' dz' \\ &= G\rho \int_0^a r' dr' \int_0^{2\pi} d\phi' \int_0^{\delta h} \frac{dz'}{|\mathbf{r} - \mathbf{r}'|} \end{aligned} \quad (14.57)$$

For an infinitely thin disk the quantity  $|\mathbf{r} - \mathbf{r}'|$  is independent from  $z'$  and we have:  $\mathbf{r}' \cong x'\mathbf{i} + y'\mathbf{j}$ . Then,

$$\begin{aligned}
 V(\mathbf{r}) &= G\rho\delta h \int_0^R r' dr' \int_0^{2\pi} \frac{d\phi'}{|\mathbf{r} - \mathbf{r}'|} \\
 &= G\rho\delta h \int_0^R r' dr' \int_0^{2\pi} \frac{d\phi'}{\sqrt{z^2 + r'^2}} \\
 &= 2\pi G\rho\delta h \int_0^R \frac{r'}{\sqrt{z^2 + r'^2}} dr' \\
 &= 2\pi G\rho\delta h \left[ (z^2 + R^2)^{1/2} - z \operatorname{sgn}(z) \right] \quad (14.58)
 \end{aligned}$$

The corresponding gravity field along the  $z$  axis is given by:

$$\mathbf{g}(\mathbf{r}) = \nabla V = 2\pi G\rho\delta h \left[ \frac{z}{\sqrt{z^2 + R^2}} - \operatorname{sgn}(z) \right] \mathbf{k} \quad (14.59)$$

To obtain the field generated by a thin lamina of thickness  $\delta h$ , is sufficient now to take the limit of (14.59) as  $R \rightarrow \infty$ .

The result is:

$$\begin{aligned}
 \mathbf{g}(\mathbf{r}) &= \lim_{R \rightarrow \infty} 2\pi G\rho\delta h \left[ \frac{z}{\sqrt{z^2 + R^2}} - \operatorname{sgn}(z) \right] \mathbf{k} \\
 &= -2\pi G\rho\delta h \operatorname{sgn}(z) \mathbf{k} \quad (14.60)
 \end{aligned}$$

This expression shows that the gravity of a Bouguer slab is independent from the distance of the observation point from the slab surface. Therefore, for a slab of thickness  $h$  we have that the *simple Bouguer correction*, which must be *subtracted* from observed gravity data, is given by:

$$\delta g_{SB}(\mathbf{r}) = 2\pi G\rho h \quad (14.61)$$

A *simple Bouguer anomaly* is a gravity anomaly that takes into account of both altitude and the presence of sufficiently uniform topography above the geoid. It is obtained

subtracting the correction (14.61) from free-air anomalies:

$$\Delta g_{SB}(\psi, \phi, h) = \Delta g_{FA}(\psi, \phi, h) - 2\pi G\rho h \quad (14.62)$$

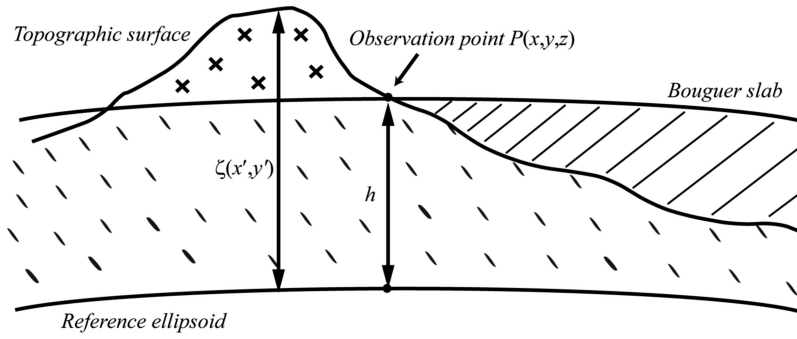
For typical crustal rocks, the value of density that is usually adopted is  $\rho = 2,670 \text{ kg/m}^3$ . In this instance, the formula for calculating simple Bouguer anomalies becomes:

$$\begin{aligned}
 \Delta g_{SB}(\psi, \phi, h) &= \Delta g_{FA}(\psi, \phi, h) \\
 &\quad - 0.1119 \times 10^{-5} h \\
 &= g(\psi, \phi, h) - \gamma_0(\psi) \\
 &\quad + 0.3086 \times 10^{-5} h \\
 &\quad - 0.1119 \times 10^{-5} h \quad (14.63)
 \end{aligned}$$

In the case of marine surveys, the simple Bouguer correction is made replacing the water, which has density  $\rho = 1,000 \text{ kg m}^{-3}$ , with a slab having density  $\rho = 2,670 \text{ kg/m}^3$  and thickness  $h$  equal to the bathymetric depth. However, now the sign of the correction (14.61) must be reversed, so that the quantity  $2\pi G\rho h$  will be added to the observed gravity. Bouguer anomalies calculated through (14.63) reflect the presence of *density contrasts* relative to the standard value  $\rho = 2,670 \text{ kg/m}^3$ , which is a reference value appropriate for most geological situations.

Simple Bouguer anomalies are mostly appropriate when the topography (or the bathymetry) is sufficiently uniform to be approximated by a slab of constant thickness. When this condition is not satisfied, further correction is necessary to take into account of the presence of masses above or below the Bouguer slab. For example, mountains give a positive contribution to the gravity measurements (although their isostatic roots give an even greater *negative* contribution). Therefore, gravity data measured in regions with moderate or rugged relief usually undergo the so-called *terrain correction*, which leads to calculate *complete Bouguer anomalies*.

The procedure of terrain correction usually requires a high-resolution *digital terrain model* (DEM), which specifies a value of altitude



**Fig. 14.6** Terrain correction for the calculation of complete Bouguer anomalies. The mountain to the *left* of the observation point is located externally to the Bouguer slab and exerts a negative (*upward*) attraction on *P*. Therefore, the corresponding correction must have positive sign.

$z' = \zeta(x', y')$  for each location with projected geographic coordinates  $(x', y')$  (Fig. 14.6). If  $P \equiv (x, y, z)$  is an observation point, the vertical component of the gravity field at *P* generated by mass elements  $dm = \rho dx' dy' dz'$  at positions  $(x', y', z')$  between the Bouguer slab and the topographic surface is given by:

$$\delta g_T(x, y, h) = -G\rho \int_{-\infty}^{+\infty} dx' \int_{-\infty}^{+\infty} dy' \int_h^{\zeta(x', y')} \frac{(h - z') dz'}{\left[ (x - x')^2 + (y - y')^2 + (h - z')^2 \right]^{3/2}} \quad (14.64)$$

This quantity is termed *terrain correction*. The corresponding complete Bouguer anomaly is obtained by adding (14.64) to the simple anomaly:

$$\Delta g_{CB}(\psi, \phi, h) = \Delta g_{SB}(\psi, \phi, h) + \delta g_T(\psi, \phi, h) \quad (14.65)$$

In the case of crustal scale studies, the terrain correction is usually negligible, thereby it is ignored. In general, Bouguer anomalies are negative in continental regions and positive over oceanic basins. This is also a consequence of the presence of low-density crustal roots below mountain regions. Free-air and Bouguer anomalies behave differently with respect to the re-

Similarly, the contribution of the *dashed region* to the *right* of *P*, which was included in the simple Bouguer correction with negative sign, should be removed through a correction having positive sign

gional isostatic equilibrium. In fact,  $\Delta g_{FA} \cong 0$  when a region is in isostatic equilibrium, whereas  $\Delta g_{SB} \cong 0$  in case of uncompensated structure.

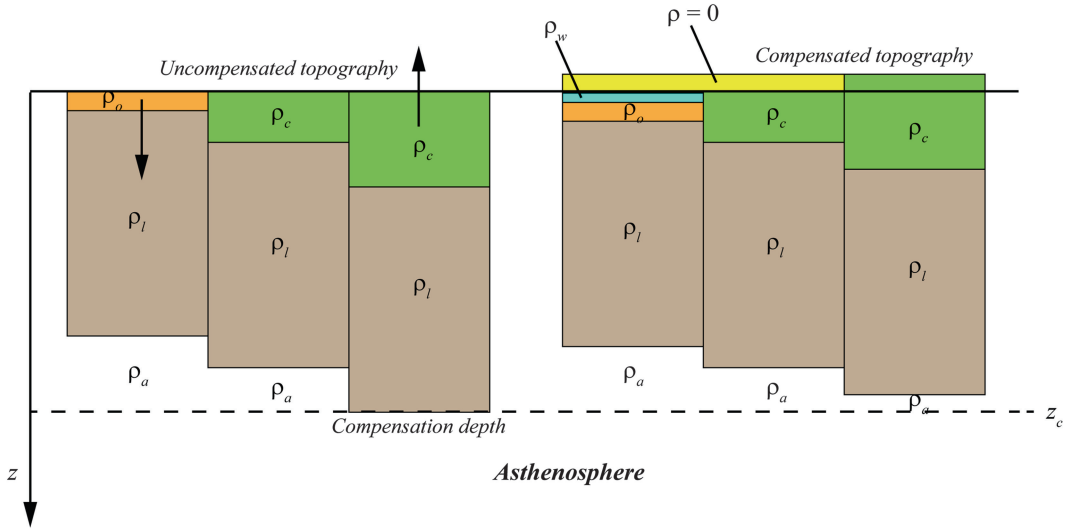
## 14.6 Isostasy and Dynamic Topography

The major part of the Earth's surface topography can be explained by *isostatic compensation* of density variations within the lithosphere. In Chap. 12, we have seen that the *isostasy principle* can be expressed as a law of invariance for the hydrostatic pressure at some compensation depth  $z_c$  in the uppermost asthenosphere:

$$P(x, y, z_c) = g \int_{-H_{\max}}^{z_c} \rho(x, y, z) dz = \text{const} \quad (14.66)$$

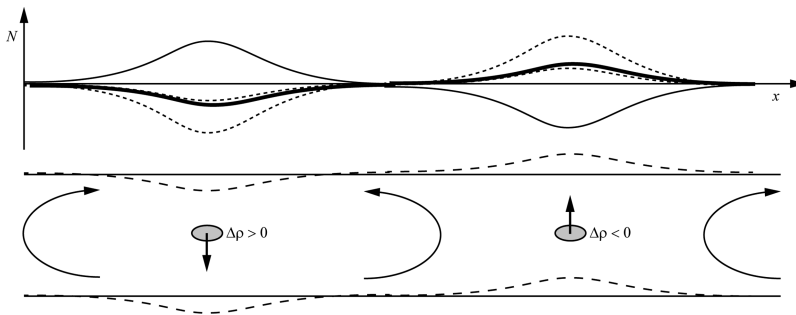
where  $H_{\max}$  is the maximum altitude. This equation implies that lateral variations of crustal or lithospheric thickness must be compensated by variations of topography, as illustrated in the example of Fig. 14.7.

The mechanism of isostatic compensation can explain the  $\sim 5\text{--}8$  km elevation of the Tibetan Plateau and the Himalayan range, which result essentially from the crustal thickening that followed the collision of India with Eurasia.



**Fig. 14.7** Mechanism of isostatic compensation. Starting from a hypothetical initial situation where the elevation is zero everywhere, a lateral increase of crustal thickness in the continental crust (green regions, density  $\rho_c$ ) will determine *upward* motion to compensate the decreased weight of the column. Similarly, the denser oceanic crust

( $\rho_o > \rho_c$ ) will subside to compensate the excess weight with water (blue regions, density  $\rho_w$ ). In this example, the lithospheric mantle has constant thickness and density  $\rho_l > \rho_c, \rho_o$ . The air (yellow regions) has negligible density. Finally, it is assumed that the asthenosphere has density  $\rho_a > \rho_l$



**Fig. 14.8** Geoid anomalies,  $N$ , and dynamic topography of a fluid mantle layer that incorporates small density anomalies. The geoid undulations result (thick solid line) from the combined effect of static density contrasts (thin

solid line) and topography variations (dashed lines) at the upper and lower boundaries of the layer induced by mantle flows (dotted lines)

However, in some regions the elevated or depressed topography cannot be explained in terms of isostasy. A classic example is given by the East African and Ethiopian plateaux (Moucha and Forte 2011). Hager (1984) and Richards and Hager (1984) were the first to propose a mechanism that today is known as *dynamic topography*.

The basic idea behind the concept of dynamic topography is illustrated in Fig. 14.8. It starts from the simple observation that mantle con-

vection should cause deformation and topography variations at the Earth's surface. In fact, sinking density anomalies pull downwards mantle material located above them. At the same time, they push the mantle downwards at their front. Similarly, a rising density contrast pulls or pushes mantle material upwards. Both these situations determine flexure of the upper mantle boundaries and surface topography changes. The same mechanism would cause variations in the topography of the CMB associated with

convective flows in the lower mantle. Therefore, both the elevation of the mid-ocean ridges and the deep bathymetry of the trench zones must be considered as consequences of the dynamic topography associated with mantle convection. However, this phenomenon can also explain the apparent paradox of negative correlation between long-wavelength geoid undulations and density anomalies in the mantle mentioned above, which was first observed by Dziewonski et al. (1977). In fact, as illustrated in Fig. 14.8, dimples and swells at the Earth's surface and along mantle discontinuities (e.g., the 670 km discontinuity or the CMB) give respectively negative or positive contributions to the local geoid undulation, and eventually can overcome the component associated with the static density distribution. Hager (1984) and Richards and Hager (1984) proved that the amplitude of dimples and swells is determined by the magnitude and depth of the density contrasts, from the depth of the layer, and from the presence of viscosity stratification in the fluid layer. They also showed that the process of formation of dynamic topography could attain a steady state, because the time interval required to reach stationary conditions has the same order of magnitude of the time employed by a density anomaly to travel a distance of a few kilometers. Therefore, a steady state is reached on a postglacial rebound time interval, essentially instantaneously on a geological time scale. In general, the effect of dynamic topography on gravity is quite large. For example, the topographic low of subduction zones may reach 10 km depth. Although dynamic topography is traditionally associated with the vertical motions of thermal (Rayleigh-Bénard) convection, horizontal mantle flows also support changes of elevation. In Chap. 13, we have seen that horizontal pressure-driven flows (Poiseuille-Couette flows) are possible and arise from horizontal pressure gradients in the asthenosphere. The existence of these flows is testified both by events of non-equilibrium plate kinematics in the geologic past (discussed in Chap. 6) and by the non-isostatic topographic slope observed in some regions (e.g., Conder and Wiens 2007). The relation between topographic slope and horizontal pressure gradient comes from the observation

that an excess of thermodynamic pressure with respect to the hydrostatic value at the compensation depth must be equivalent to uncompensated topography at the Earth's surface. Therefore, if  $h$  is the elevation exceeding the normal isostatic height, we can write the following fundamental equation (Schubert and Turcotte 1972; Schubert et al. 1978):

$$\frac{\partial p}{\partial x} = \rho_a g \frac{\partial h}{\partial x} \quad (14.67)$$

where  $\rho_a$  is the density of the asthenosphere. Therefore, assuming  $\rho_a = 3,450 \text{ kg m}^{-3}$  we have that a pressure gradient of  $100 \text{ kPa km}^{-1}$  would generate a dynamic topographic slope of  $2.96 \text{ m km}^{-1}$ .

## Problems

1. Find the gravity anomaly generated by a spherical object at 1 km depth with density contrast  $\Delta\rho = 200 \text{ kg m}^{-3}$ ;
2. Find the gravity anomaly generated by an infinitely long horizontal dike at 1 km depth, with cylindrical cross-section of radius  $R = 5 \text{ m}$  and density contrast  $\Delta\rho = 100 \text{ kg m}^{-3}$ ;
3. Find an expression for the thickness  $h_r$  of the crustal root generated by a mountain belt with average altitude  $h_m$  in excess of a normal continental crust, assuming that the lithospheric mantle has constant thickness;
4. A gravimeter based on measurements of falling body trajectory has uncertainty  $\pm 10^{-6} \text{ s}$  on reading the arrival time at distance  $d = 0.5 \pm 0.0001 \text{ m}$ . Estimate the uncertainty on gravity.

## References

- Biancale R, Balmino G, Lemoine JM, Marty JC, Moynot B, Barlier F, Exertier P, Laurain O, Gegout P, Schwintzer P, Reigber C, Bode A, König R, Massmann FH, Raimondo J-C, Schmid R, Yuan Zhu S (2000) A new global Earth's gravity field model from satellite orbit perturbations: GRIM5-S1. *Geophys Res Lett* 27(22):3611–3614

- Conder JA, Wiens DA (2007) Rapid mantle flow beneath the Tonga volcanic arc. *Earth Planet Sci Lett* 264(1):299–307
- Dziewonski AM, Hager BH, O'Connell RJ (1977) Large-scale heterogeneities in the lower mantle. *J Geophys Res* 82(2):239–255
- Hager BH (1984) Subducted slabs and the geoid: constraints on mantle rheology and flow. *J Geophys Res* 89(B7):6003–6015. doi:10.1029/JB089iB07p06003
- Hager BH, Richards MA (1989) Long-wavelength variations in Earth's geoid: physical models and dynamical implications. *Phil Trans Roy Soc Lond Series A* 328(1599):309–327
- Heiskanen WA, Moritz H (1993) *Physical geodesy*. Institute of Physical Geodesy, Technical University, Graz, Corrected reprint, 364 pp
- Lerch FJ, Nerem RS, Putney BH, Felsentreger TL, Sanchez BV, Marshall JA, Klosko SM, Patel GB, Williamson RG, Chinn DS, Chan JC, Rachlin KE, Chandler NL, McCarthy JJ, Luthcke SB, Pavlis NK, Pavlis DE, Robbins JW, Kapoor S (1994) A geopotential model from satellite tracking, altimeter, and surface gravity data: GEM-T3. *J Geophys Res* 99(B2):2815–2839. doi:10.1029/93JB02759
- Moucha R, Forte AM (2011) Changes in African topography driven by mantle convection. *Nat Geosci* 4(10):707–712
- Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2012) The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *J Geophys Res* 117:B04406. doi:10.1029/2011JB008916
- Pekeris CL (1935) Thermal convection in the interior of the Earth. *Mon Not Roy Astron Soc Geophys* 3: 343–367
- Rapp RH (1997) Use of potential coefficient models for geoid undulation determinations using a spherical harmonic representation of the height anomaly/geoid undulation difference. *J Geod* 71(5): 282–289
- Richards MA, Hager BH (1984) Geoid anomalies in a dynamic earth. *J Geophys Res* 89(B7):5987–6002. doi:10.1029/JB089iB07p05987
- Schubert G, Turcotte DL (1972) One-dimensional model of shallow-mantle convection. *J Geophys Res* 77(5):945–951. doi:10.1029/JB077i005p00945
- Schubert G, Yuen DA, Froidevaux C, Fleitout L, Souriau M (1978) Mantle circulation with partial shallow return flow: effects on stresses in oceanic plates and topography of the sea floor. *J Geophys Res* 83(B2):745–758. doi:10.1029/JB083iB02p00745
- Schwintzer P, Reigber C, Bode A, Kang Z, Zhu SY, Massmann F-H, Raimondo JC, Biancale R, Balmino G, Lemoine JM, Moynot B, Marty JC, Barlier F, Boudon Y (1997) Long-wavelength global gravity field models: GRIM4-S4, GRIM4-C4. *J Geod* 71(4): 189–208
- Tapley B, Ries J, Bettadpur S, Chambers D, Cheng M, Condi F, Gunter B, Kang Z, Nagel P, Pastor R, Pekker T, Poole S, Wang F (2005) GGM02—an improved Earth gravity field model from GRACE. *J Geod* 79(8): 467–478

---

## Appendix 1: Vector Calculus

---

### A1.1 Scalar Fields

Many physical quantities can be represented by a single real number. They are called *scalars*. Examples of scalar quantities are the temperature,  $T$ , the pressure,  $P$ , and the density,  $\rho$ . A scalar field  $\phi = \phi(\mathbf{r})$  is a continuous function of the spatial Cartesian coordinates  $(x, y, z)$ , which can be represented by a position vector  $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ . The spatial variability of a scalar field can be expressed through its *gradient*:

$$\nabla\phi \equiv \left( \frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z} \right) \equiv \frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} + \frac{\partial\phi}{\partial z}\mathbf{k} \quad (\text{A1.1})$$

The gradient of a scalar field is not a scalar quantity, because it is formed by the ordered set of three spatial derivatives. It is an example of *vector field*, which defines a vector quantity for each position  $(x, y, z)$ . In this instance, the vector components are the three derivatives  $\partial\phi/\partial x$ ,  $\partial\phi/\partial y$ , and  $\partial\phi/\partial z$ . Many scalar fields considered in geophysics are functions of only two spatial coordinates, for example latitude and longitude, thereby they can be represented as surfaces on the plane or on the sphere (Fig. A1.1).

Elevation, gravity and magnetic anomalies, and surface heat flux are examples of scalar fields in the plane. In this instance, the equation  $\phi = \phi_0$ ,  $\phi_0$  being a constant field value, defines a *contour line* for the scalar field in the  $(x, y)$  plane. This is the set of points  $(x, y)$  such that  $\phi(x, y) = \phi_0$ . A set of contour lines for equally spaced values  $\phi_0, \phi_1, \dots$  is a convenient way to illustrate the field properties, alternative to the surface representation (Fig. A1.2).

If  $\mathbf{r}$  and  $\mathbf{r} + d\mathbf{r}$  are two neighbor points on a contour line, then it results:

$$\begin{aligned} d\phi &= \phi(\mathbf{r} + d\mathbf{r}) - \phi(\mathbf{r}) = \frac{\partial\phi}{\partial x}dx + \frac{\partial\phi}{\partial y}dy \\ &= \nabla\phi \cdot d\mathbf{r} = 0 \end{aligned} \quad (\text{A1.2})$$

Therefore,  $\nabla\phi$  is always orthogonal to the contour lines of a scalar field in the plane. This implies that the direction of  $\nabla\phi$  coincides with the direction of maximum *increase* of  $\phi$  (Fig. A1.3).

Similarly, in the general case of three-dimensional fields, the set of points  $(x, y, z)$  such that  $\phi(x, y, z) = \phi_0$  is an isosurface in  $\mathbb{R}^3$ . In this instance,  $\nabla\phi$  is always orthogonal to the isosurfaces of  $\phi$ . Given a direction unit vector  $\mathbf{n}$ , the quantity:

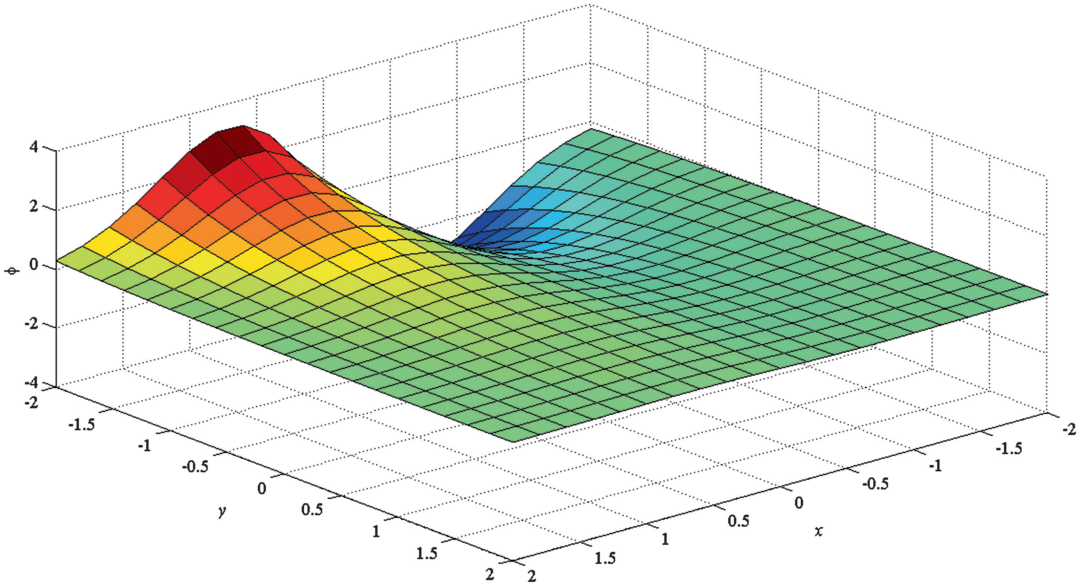
$$\nabla\phi \cdot \mathbf{n} \equiv \frac{\partial\phi}{\partial n} = \frac{\partial\phi}{\partial x}n_x + \frac{\partial\phi}{\partial y}n_y + \frac{\partial\phi}{\partial z}n_z \quad (\text{A1.3})$$

is called *directional derivative* of  $\phi$  in the direction of  $\mathbf{n}$ . It represents the variation of  $\phi$  as we move from the actual position  $\mathbf{r}$  to a neighbor position  $\mathbf{r} + d\mathbf{r}$  in the direction  $\mathbf{n}$ .

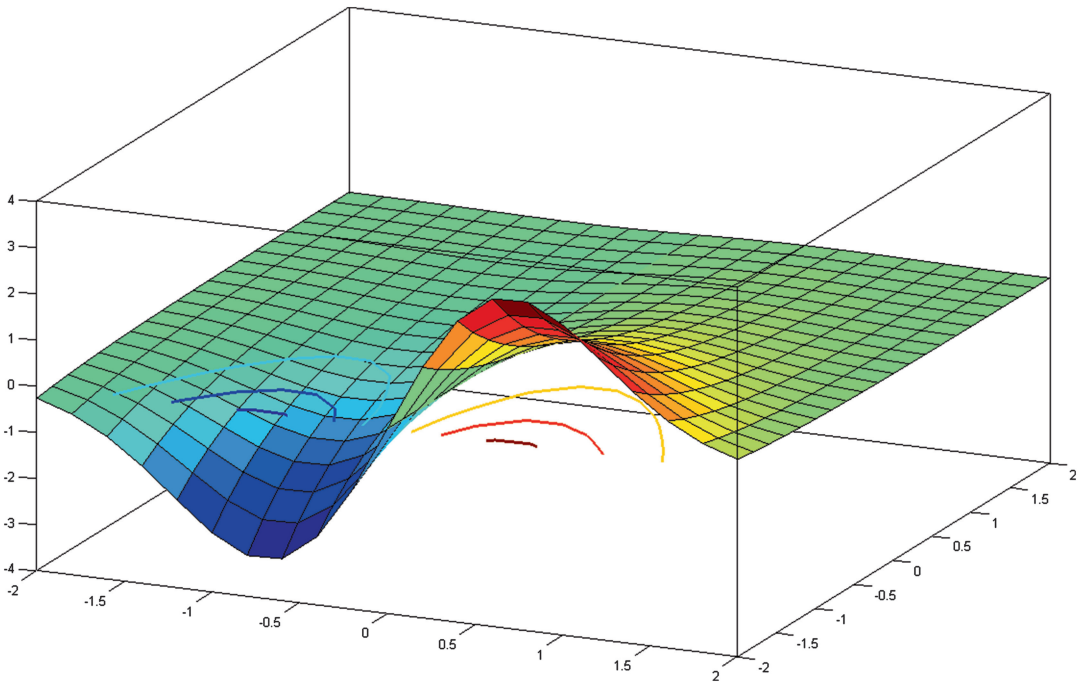
---

### A1.2 Vector Fields

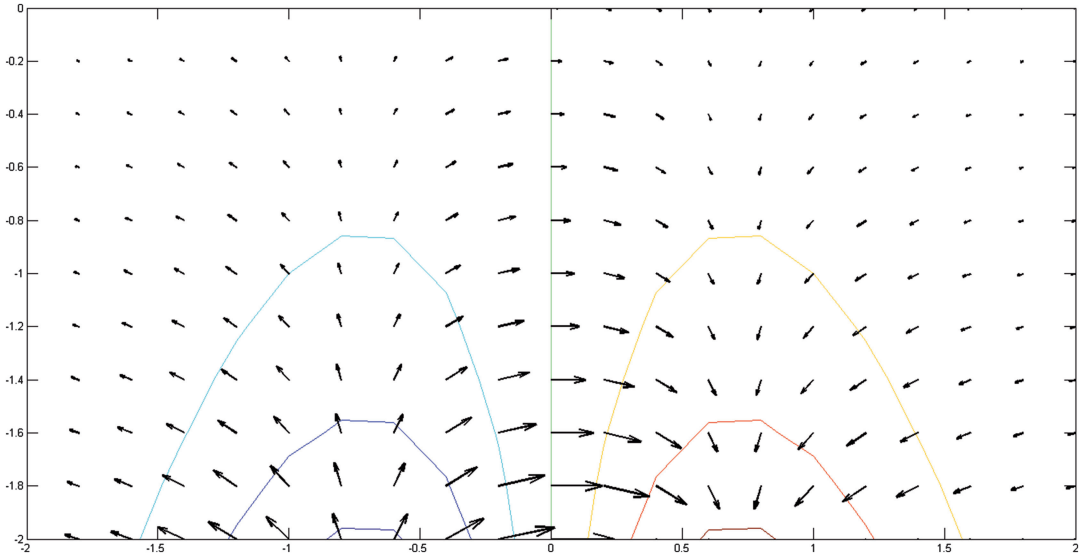
A vector field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$  associates a vector  $\mathbf{A}$  to each position  $\mathbf{r} \equiv (x, y, z)$  in the space. Classic examples are the force fields, in particular electric, gravity, and magnetic fields, but also velocity and acceleration fields. The spatial variability of vector fields is described through two differential



**Fig. A1.1** Surface representation of a scalar field  $\phi$  in the plane  $(x, y)$

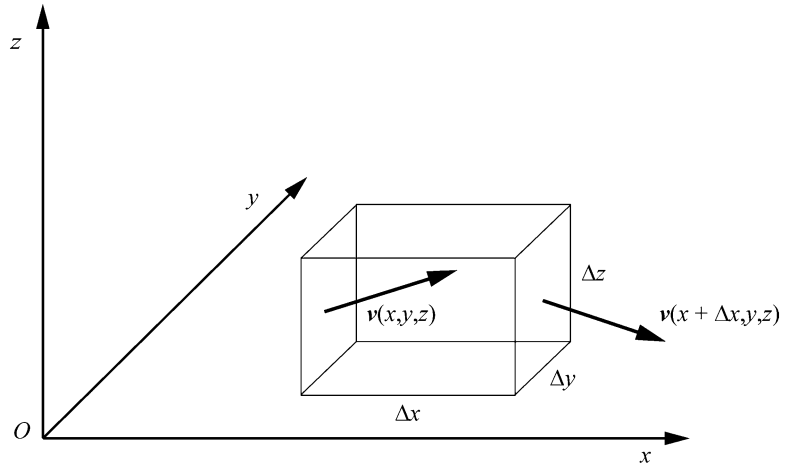


**Fig. A1.2** Contour lines for the scalar field of Fig. A1.1



**Fig. A1.3** Contour lines and gradient for the scalar field of Fig. A1.1

**Fig. A1.4** Fluid flux through a small volume



operators: the divergence and the curl. The *divergence* of a vector field is defined as follows:

$$\nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \quad (\text{A1.4})$$

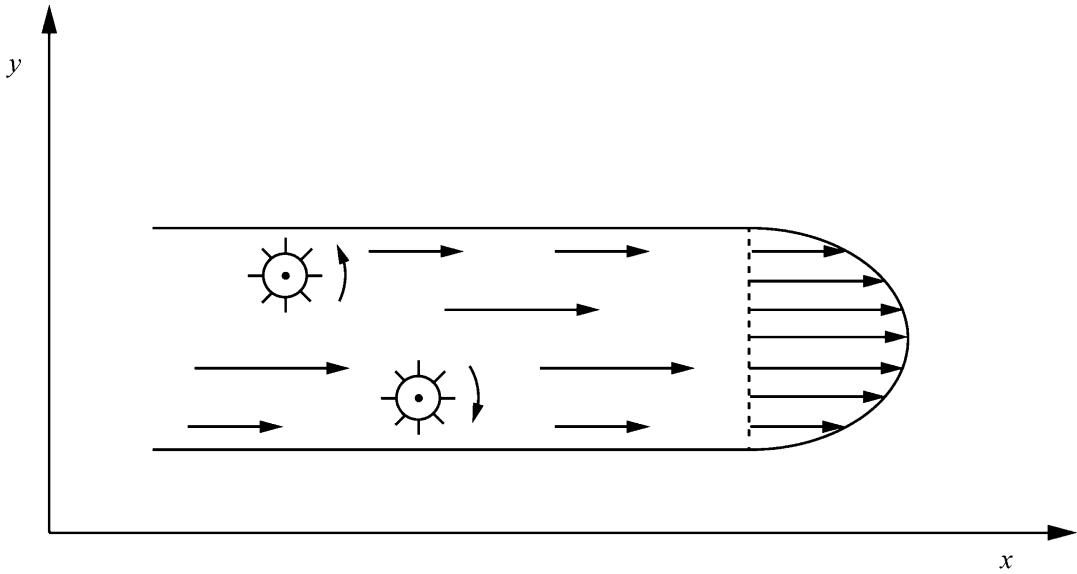
To understand the physical interpretation of this scalar field, let us consider a steady velocity field in a fluid,  $\mathbf{v} = \mathbf{v}(\mathbf{r})$ , and a small volume  $\Delta V = \Delta x \Delta y \Delta z$  at location  $\mathbf{r}$  (Fig. A1.4).

The volume of fluid which enters the volume through the left face in the  $x$  direction per unit time is given by  $v_x(x, y, z) \Delta y \Delta z$ . Similarly, the volume of fluid that leaves  $\Delta V$  in the  $x$  direction

through the right face per unit time is given by  $v_x(x + \Delta x, y, z) \Delta y \Delta z$ . Therefore, the net volume of fluid *per unit volume* and per unit time through  $\Delta V$  in the  $x$  direction is given by:

$$\begin{aligned} & \frac{v_x(x + \Delta x, y, z) \Delta y \Delta z - v_x(x, y, z) \Delta y \Delta z}{\Delta x \Delta y \Delta z} \\ &= \frac{\left[ \frac{\partial v_x}{\partial x} \Delta x \right] \Delta y \Delta z}{\Delta x \Delta y \Delta z} = \frac{\partial v_x}{\partial x} \end{aligned} \quad (\text{A1.5})$$

Consequently, the net volume of fluid per unit volume and per unit time through  $\Delta V$  is given by  $\nabla \cdot \mathbf{v}$ . The second differential operator that



**Fig. A1.5** Rotors in a pressure-driven fluid flow through a channel

describes the structure of a vector field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$  is the *curl*, which is a vector field defined as follows:

$$\nabla \times \mathbf{A} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ A_x & A_y & A_z \end{vmatrix} = \left( \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) \mathbf{i} + \left( \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) \mathbf{j} + \left( \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \mathbf{k} \quad (\text{A1.6})$$

The physical interpretation of the curl can be understood considering again the case of the velocity field within a fluid. Let us consider a steady pressure-driven fluid flow through a channel, as in Fig. A1.5. Two rotors, located close to the opposite walls would rotate in opposite directions.

The curl of the velocity field in Fig. A1.5 is given by:  $\nabla \times \mathbf{v} = -(\partial v_x / \partial y) \mathbf{k}$ . It is in the negative  $z$  direction close to the lower wall and in the positive  $z$  direction close to the upper wall. In the first case, a rotor placed in the fluid would rotate clockwise (negative angular velocity) or counterclockwise (positive rotation) according to the sign of  $\nabla \times \mathbf{v}$  and with angular velocity proportional to the magnitude of the curl. Curl and

divergence have the following important properties, which can be easily verified by the reader:

$$\nabla \times \nabla \times \mathbf{A} = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad (\text{A1.7})$$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0 \quad (\text{A1.8})$$

$$\nabla \times \nabla \phi = \mathbf{0} \quad (\text{A1.9})$$

for each vector field  $\mathbf{A}$  and for any scalar field  $\phi$ . In (A1.7), the operator  $\nabla^2$  is the *Laplacian*, which is defined by:

$$\nabla^2 \phi \equiv \nabla \cdot \nabla \phi = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} \quad (\text{A1.10})$$

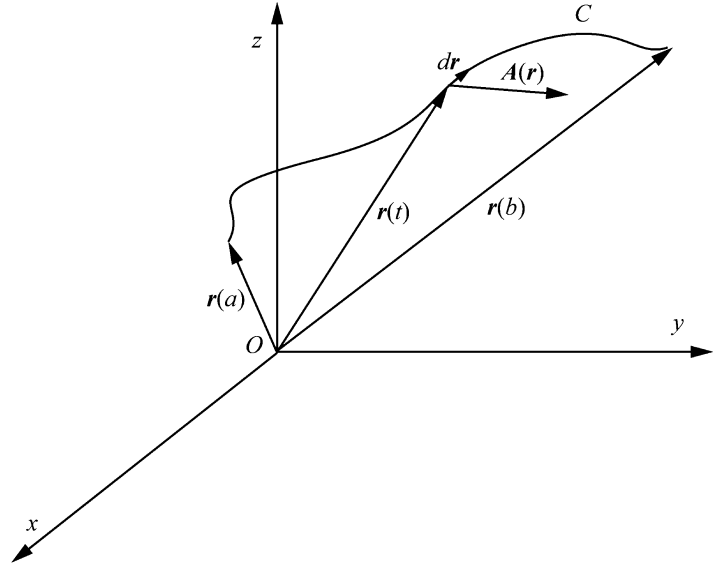
for an arbitrary scalar field  $\phi$ . In the case of a vector field, it is intended that the operator is applied to each component independently.

### A1.3 Integrals Theorems

A *path*  $C$  in  $\mathbb{R}^3$  is a line such that each point  $P \in C$  has a position vector  $\mathbf{r}$  defined by the parametric equation:

$$\mathbf{r} = \mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} ; t \in [a, b] \quad (\text{A1.11})$$

**Fig. A1.6** Geometrical relationships in the calculation of line integrals



Given a vector field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$ , a *line integral* over the path  $C$  is a scalar quantity given by:

$$I = \int_C \mathbf{A}(\mathbf{r}) \cdot d\mathbf{r} \quad (\text{A1.12})$$

Figure A1.6 illustrates the general approach followed in the computation of line integrals, such as the work associated with a force field. We simply substitute the parametric equation (A1.11) into (A1.12), which reduces the integral to a standard definite integral. To this purpose, it is necessary determining a parametric expression for the infinitesimal displacement vector  $d\mathbf{r}$ . This is accomplished differentiating (A1.11). When  $\mathbf{A}$  coincides with the gradient of a scalar field, we say that  $\mathbf{A}$  is a *potential field*. In this instance, the following analogue of the fundamental theorem of calculus holds:

$$I = \int_C \nabla\phi \cdot d\mathbf{r} = \phi(\mathbf{r}(b)) - \phi(\mathbf{r}(a)) \quad (\text{A1.13})$$

When the two end points of a path coincide, so that  $C$  forms a *closed loop*, we write:

$$I = \oint_C \mathbf{A}(\mathbf{r}) \cdot d\mathbf{r} \quad (\text{A1.14})$$

Clearly, by (A1.13) closed loop line integrals of potential fields are always zero, so that  $I$  results to be independent from the path linking  $\mathbf{r}(a)$  to  $\mathbf{r}(b)$ . For this reason potential fields are said to be *conservative*. Just as a curve in the three-dimensional space, a *surface*  $S$  in  $\mathbb{R}^3$  can be defined in parametric form. In this instance, two parameters,  $u$  and  $v$ , must be defined in a region  $D$  of  $\mathbb{R}^2$ . These parameters are used to identify position vectors on  $S$ .

Then, likewise paths, parametric surfaces are defined by equations of the form:

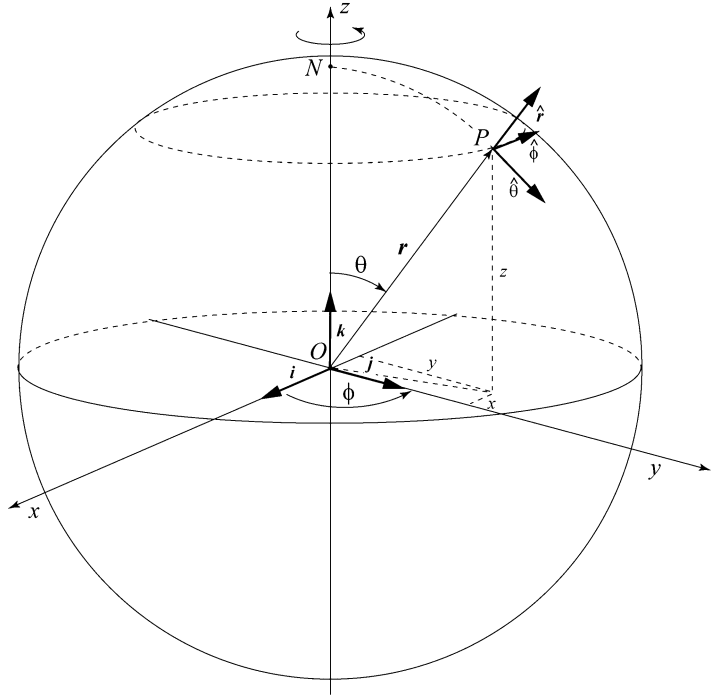
$$\mathbf{r} = \mathbf{r}(u, v) = x(u, v)\mathbf{i} + y(u, v)\mathbf{j} + z(u, v)\mathbf{k}; \\ \times (u, v) \in D \subseteq \mathbb{R}^2 \quad (\text{A1.15})$$

At any point on  $S$ , the normal versor  $\mathbf{n}$  can be used to define a vector representing the oriented surface element  $d\mathbf{S} = \mathbf{n}dS$ , which has magnitude equal to the infinitesimal area  $dS$  and normal versor  $\mathbf{n}$ . Given a vector field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$ , the *flux* through the surface  $S$  is defined as the integral:

$$\Phi = \int_S \mathbf{A}(\mathbf{r}) \cdot d\mathbf{S} \quad (\text{A1.16})$$

Once again, the integral is calculated substituting the parametric equations (A1.15) into (A1.16). The following *Stokes' theorem* is a

**Fig. A1.7** Base versors for the transformation from Cartesian to spherical derivatives



fundamental theorem of vector calculus, which allows to convert a surface integral into a closed loop line integral.

### Stokes' Theorem

Let  $S$  be a parametric surface with boundary  $C(S)$ . Then, for every vector field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$ , it results:

$$\int_S \nabla \times \mathbf{A} \cdot d\mathbf{S} = \oint_{C(S)} \mathbf{A} \cdot d\mathbf{r} \quad (\text{A1.17})$$

Another important theorem of vector analysis is the following *Gauss' theorem*, which allows to convert a volume integral into a surface integral.

### Gauss (or Divergence) Theorem

Let  $\mathbf{R} \subset \mathbb{R}^3$  be a region of  $\mathbb{R}^3$ , bounded by the closed surface  $S(\mathbf{R})$ . Then, for every vector field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$ , it results:

$$\int_{\mathbf{R}} \nabla \cdot \mathbf{A} dV = \oint_{S(\mathbf{R})} \mathbf{A} \cdot d\mathbf{S} \quad (\text{A1.18})$$

## A1.4 Spherical Coordinates

Spherical coordinates are a very common way to represent positions in Earth Sciences. Figure A1.7 shows the relation between Cartesian,  $(x, y, z)$ , and spherical,  $(r, \theta, \phi)$ , coordinates associated with a position vector  $\mathbf{r}$ .

The general transformation rule of vector fields and differential operators from Cartesian to spherical coordinates requires the introduction of the three orthogonal versors  $\hat{\mathbf{r}}$ ,  $\hat{\boldsymbol{\theta}}$ , and  $\hat{\boldsymbol{\phi}}$  (Fig. A1.7), which are directed respectively toward increasing distance from the origin, increasing colatitude (that is, southward), and increasing longitude (i.e., eastward) at point  $P \equiv (x, y, z)$ . It is easy to prove that the transformation from the base  $(\mathbf{i}, \mathbf{j}, \mathbf{k})$  to the base  $(\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$  is given by:

$$\begin{bmatrix} \hat{\mathbf{r}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\phi}} \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\sin \phi & \cos \phi & 0 \end{bmatrix} \begin{bmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{bmatrix} \quad (\text{A1.19})$$

The transformation matrix in (A1.19) is an orthogonal matrix, thereby its inverse is simply

its transpose. This allows to write immediately the inverse transformation:

$$\begin{bmatrix} i \\ j \\ k \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi & \cos \theta \cos \phi & -\sin \phi \\ \sin \theta \sin \phi & \cos \theta \sin \phi & \cos \phi \\ \cos \theta & -\sin \theta & 0 \end{bmatrix} \begin{bmatrix} \hat{r} \\ \hat{\theta} \\ \hat{\phi} \end{bmatrix} \quad (\text{A1.20})$$

## A1.5 Index Notation

Index notation is a standard way to represent the Cartesian components of vectors and tensors in continuum mechanics. In this context,  $x_1 \equiv x$ ,  $x_2 \equiv y$ ,  $x_3 \equiv z$ , the base versors  $(\mathbf{i}, \mathbf{j}, \mathbf{k})$  are substituted by  $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ , and the generic component of a vector  $\mathbf{A}$  is indicated simply by  $A_i$ . This notation combines with the *Einstein summation convention*, which requires that duplicated indices in expressions are always summed. For example,

$$\begin{aligned} A_i B_i &\equiv A_1 B_1 + A_2 B_2 + A_3 B_3 \\ &\equiv A_x B_x + A_y B_y + A_z B_z \end{aligned}$$

Some ambiguity is possible where powers of an indexed quantity occur. Therefore, an expression such as:

$$\sum_i A_i^2$$

with the summation convention is always written in the form:

$$A_i A_i$$

Two tensors are often encountered in computation based on index notation. The first one is the *Kronecker delta*,  $\delta_{ij}$ , which simply represents a component of the identity matrix:

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{otherwise} \end{cases} \quad (\text{A1.21})$$

The second tensor is the *Levi-Civita tensor*,  $\varepsilon_{ijk}$ , which is defined to be zero if any two of the indices  $ijk$  are equal, and otherwise either +1 or -1 according as  $ijk$  is an even or odd permutation of 1,2,3. Thus, in terms of the Levi-Civita tensor, the components of the cross product  $\mathbf{C} = \mathbf{A} \times \mathbf{B}$  can be written as follows:

$$C_i = \varepsilon_{ijk} A_j B_k \quad (\text{A1.22})$$

Regarding the differential operators, in index notation the components of the gradient of a scalar field  $\phi$  are simply:

$$(\nabla \phi)_i \equiv \frac{\partial \phi}{\partial x_i} ; \quad i = 1, 2, 3$$

The divergence of a vector field  $\mathbf{A} = \mathbf{A}(\mathbf{r})$  will be written as follows:

$$\nabla \cdot \mathbf{A} \equiv \frac{\partial A_i}{\partial x_i}$$

Finally, the components of the curl will be expressed as:

$$(\nabla \times \mathbf{A})_i \equiv \varepsilon_{ijk} \frac{\partial A_j}{\partial x_k} ; \quad i = 1, 2, 3$$



---

## Appendix 2: Algorithms and Data Structures

---

### A2.1 Algorithms

There are three basic components of any software system. They are: (1) a *data* collection, (2) a set of *algorithms*, and (3) a system of memory *data structures*. Data can be classified as *numerical data*, *strings*, and *compound data*. Numerical data are binary representations of natural, integer, or real numbers. Strings are sequences of symbols of *alphabets*, for example the ASCII or UNICODE alphabets. Finally, compound data are values that contain heterogeneous subsets of information (numbers, strings, etc.), typically in fixed number and sequence. The elements of compound data records are usually called *fields*. An *algorithm* is an ordered sequence of *instructions* for a virtual or real processing unit. There are three fundamental classes of instructions. The first class includes *assignments* of the form:

$$var \leftarrow expr \quad (\text{A2.1})$$

where *var* is a *variable*, that is, a memory location where we can store a value, and *expr* is an *expression* that combines a homogeneous set of variables and constants through arithmetical, logical, or string operators, depending from the expression type. Differently from physics, in computer science the fundamental distinction between variables and constants is not based on their capability to change their value. Variables are named memory locations that may contain or not a value. *Constants* are parameter values, generally stored *within* the algorithm code. The operation involved into the assignment instruction (A2.1) includes the evaluation of the expression and the assignment of the

value to the variable *var*. For example, the assignment:

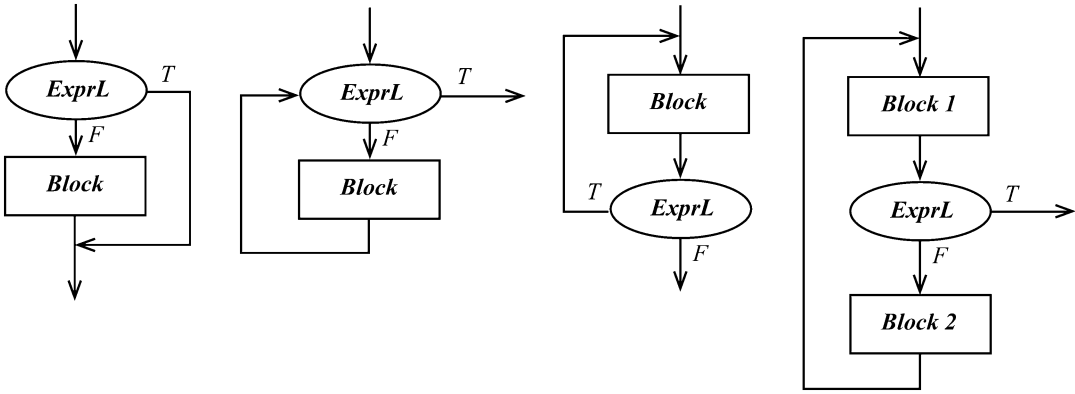
$$a \leftarrow c + x^2 - 2$$

requires reading of the value of variable *x*, computation of the second power of this value, reading of the value of variable *c*, addition between these two values, finally addition of the constant  $-2$ . The resulting value is then assigned to the variable *a*. Expressions may also contain *function calls*, which invoke execution of specialized code for the calculation of common mathematical functions or other code that must be executed several times. For example,

$$a \leftarrow y + x^2 - 2 \sin(z)$$

invokes the execution of the sine function during the evaluation of the expression. In this instance, the instruction is executed first calling the function *sin()*, to which the value of *z* is passed, then using the return value of this function in the evaluation of the expression. A second class of algorithm instructions allows to control the *program flow*, that is, the sequence of instructions that will be executed after certain steps. Program flow control instructions do not execute complex mathematical operations and do not perform assignments. They simply evaluate the state of a logical variable (a variable whose values can be *true* or *false*), or a logical expression, and decide the number of the next instruction to be executed through a *jump* instruction. For example, the instruction:

```
#5 : ...  
#6 :  $x \leq 10 \Rightarrow$  jump #20  
#7 : ...
```



**Fig. A2.1** Control flow programming structures. From left to right: Conditional branch, pre-conditional loop, post-conditional loop, and infinite iteration. *ExprL* is a

logical expression, which can assume the values *true* (*T*) or *false* (*F*). *Block* is a sequence of instructions

at step #6 is a *conditional branch*. It is executed comparing the value of  $x$  with 10. If  $x$  is less than or equal to 10, then the next instruction that will be executed is that at line #20, otherwise it will be executed the instruction at line #7. In absence of control flow instructions, the sequence of instructions that are executed by the processing unit is governed by the progression of line numbers. However, jump instructions determine a modification of the natural sequence of steps. This feature can be considered as the most fundamental source of algorithms' power. For example, backward jumps allow to execute more and more times a block of instructions, providing the basis for the construction of *cycles* (or *iterations*). Consider the following code segment:

**Algorithm A2.1**  $\text{pow}(x,n)$ :  $n$ -th power of a real number  $x$ .

Input  $x \in \mathfrak{R}$ ;  $n \in \mathbf{N}$ ; Output:  $p = x^n \in \mathfrak{R}$

```
{
0:  $p \leftarrow 1$ ;  $k \leftarrow 0$ 
1:  $k = n \Rightarrow \text{stop}$ 
2:  $p \leftarrow p \times x$ 
3:  $k \leftarrow k + 1$ 
4: jump #1
}
```

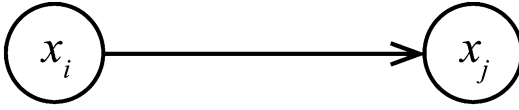
At step #0, the algorithm initializes a *counter*  $k$  and the return variable  $p$ . Line #1 contains a conditional branch: if  $k$  has attained the value  $n$ ,

the execution is stopped and the algorithm returns the actual value of  $p$ , else execution proceeds with step #2. Steps #2 and #3 form the core block of the algorithm instructions. They calculate  $x^n$  by the recurrence formula:  $x^k = x x^{k-1}$ ;  $k = 1, 2, \dots, n$ . Finally, step #4 is a backward jump that forces the program to restart at line #1. We say that the sequence of steps 1–4 represents a cycle or iteration. There are three kinds of cycles in computer programming, which are illustrated in Fig. A2.1. They can be implemented using backward or forward conditional branches or unconditioned jump statements.

The third class of algorithm instructions is represented by *input/output* directives, which allow to display results of computation, create graphics, or enter input data and parameters. These instructions are specific of the programming language used to implement the algorithms and ultimately depend from the operating system.

## A2.2 Data Structures

Let  $\mathbf{D} = \{x_1, x_2, \dots, x_n\}$  be a set of  $n$  data stored in memory variables. It is assumed that the index  $i = 1, 2, \dots, n$  associated with the individual data reflects the order of insertion into the main memory, not a logical relation between them. The set could be formed by homogeneous data, for example a collection of integers, or by heterogeneous data, for example compound data records



**Fig. A2.2** An arc linking two nodes in a data structure

of different type and length. The important thing is that in any case a logical relationship exists between the elements of the set, which must be implemented for consistent computation and data retrieval. As an example, a set of tectonic elements, faults, and site locations could be characterized by hierarchical relations between the set items, according to a kinematic model. A *data structure*  $\sigma$  on  $\mathbf{D}$  can be defined as a binary function on  $\mathbf{D}^2$ :

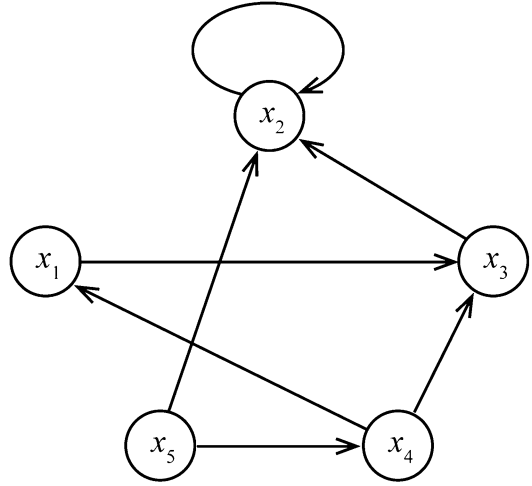
$$\sigma : (x_i, x_j) \in \mathbf{D}^2 \rightarrow \sigma_{ij} \in \{0, 1\} \quad (\text{A2.2})$$

This function simply establish the existence of a qualitative *logical* relation between two elements  $x_i$  and  $x_j$  when  $\sigma_{ij} = 1$ . When a data structure is defined on a set  $\mathbf{D}$ , the elements of  $\mathbf{D}$  are referred to as the *nodes* or the *vertices* of the structure, while the fact that  $\sigma_{ij} = 1$  is rendered graphically drawing an *arc* between nodes  $x_i$  and  $x_j$  (Fig. A2.2).

Mathematically, the pair  $(\mathbf{D}, \sigma)$  is said to be a *graph*, and the binary function  $\sigma$  is represented through a square binary matrix of order  $n$ , which is called *adjacency matrix*. For example, the function  $\sigma$  associated with the graph of Fig. A2.3 is represented by the following adjacency matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

When  $\mathbf{A}$  is a symmetric matrix, so that  $A_{ij} = A_{ji}$  for each pair of indices  $i$  and  $j$ , then the existence of an arc from a node  $x_i$  to a node  $x_j$  implies the existence of a reversed arc from node  $x_j$  to node  $x_i$ . In this instance, the lines linking



**Fig. A2.3** An example of directed graph of order 5 and size 7

nodes are drawn without arrows and represent *pairs* of arcs. These lines are called *edges*. Data structures represented by *asymmetric* adjacency matrices are called *directed graphs* or *digraphs* (Fig. A2.3). In general, the number  $n$  of vertices in a graph  $G = (\mathbf{D}, \sigma)$  is called *order* of  $G$ , while the number  $m$  of arcs or edges is referred to as the *size* of  $G$ . Two nodes are *adjacent* when they are joined by an edge or an arc. The *neighborhood*  $I(x)$  of a node  $x$  is the set of nodes adjacent to  $x$ , and the number of elements in  $I(x)$  is called the *degree* of  $x$ . A *path* between two vertices  $x$  and  $y$  is a sequence of nodes  $(x, x', x'', \dots, y)$ , such that  $x'$  is adjacent to  $x$ ,  $x''$  is adjacent to  $x'$ , etc. If two nodes are joined by an arc directed from  $x$  to  $y$ , we say that  $x$  *dominates*  $y$ , or  $x \rightarrow y$ . Therefore, the existence of a path between  $x$  and  $y$  can be written as:  $x \rightarrow x' \rightarrow x'' \rightarrow \dots \rightarrow y$ . Now let us consider the matrix  $\mathbf{A}^2$ , whose elements are given by:

$$A_{ij}^2 = \sum_{k=1}^n A_{ik} A_{kj} \quad (\text{A2.3})$$

Clearly, for each value of the dummy index  $k$ , the term  $A_{ik} A_{kj}$  contributes to the sum if and only if  $x_i \rightarrow x_k$  and  $x_k \rightarrow x_j$ , thereby there exists the path  $(x_i, x_k, x_j)$ . Therefore, the element  $A_{ij}^2$  represents the number of paths of length two joining nodes  $x_i$  and  $x_j$ . We can extend this concept to

higher powers of  $\mathbf{A}$ , saying that an element  $A_{ij}^k$  of the matrix  $\mathbf{A}^k$  represents the number of paths of length  $k$  joining node  $x_i$  to  $x_j$ . In general, given any two vertices  $x$  and  $y$ , several distinct paths of different length may exist that join these nodes. For example, in Fig. A2.3 node  $x_2$  is the end vertex of three walks that start at  $x_5$ , namely:  $(x_5, x_2)$ ,  $(x_5, x_4, x_3, x_2)$ , and  $(x_5, x_4, x_1, x_3, x_2)$ , which have length one, three, and four respectively. The distance  $d(x, y)$  between two nodes is defined as the length of the smallest path joining  $x$  and  $y$ . It is easy to verify that for any undirected graph  $G = (\mathbf{D}, \sigma)$ , the function  $d$  is a metric on  $\mathbf{D}$ :

$$1. \ d(x, y) \geq 0; \ d(x, y) = 0 \iff x = y \quad (\text{A2.4})$$

$$2. \ d(x, y) = d(y, x) \quad (\text{A2.5})$$

$$3. \ d(x, y) + d(y, z) \geq d(x, z) \quad (\text{A2.6})$$

A major task in computer science is to find algorithms that compute the function  $d$  for any pair of nodes in a graph  $G$ . The following *breadth-first search algorithm* (BFS) is a classic solution to this problem. Given  $x \in \mathbf{D}$ , the BFS algorithm determines  $d(x, y)$  for each  $y \in \mathbf{D}$ . When the algorithm stops, if  $d(x, y) = \infty$  for some node  $y$ , then vertex  $y$  cannot be reached starting from  $x$ :

**Algorithm A2.2** *bfs(x)*: Distances from a node  $x$

Input:  $x \in \mathbf{D}$

Output:  $d(x, y) \ \forall y \in \mathbf{D}$

```
{
0:  $i \leftarrow 0$ ;  $d(x, x) \leftarrow 0$ ;  $\forall y \in \mathbf{D} \mid y \neq x, d(x, y) \leftarrow \infty$ ;
    $f \leftarrow \text{false}$ 
1:  $\forall y \in \mathbf{D} \mid d(x, y) = i, \forall z \in I(y) \mid d(x, z) = \infty$ ,
    $d(x, z) \leftarrow i + 1, f \leftarrow \text{true}$ 
2:  $f = \text{false} \Rightarrow \text{stop}$ 
3:  $i \leftarrow i + 1$ ;  $f \leftarrow \text{false}$ 
4: jump #1
}
```

The variable  $i$ , which at step #0 is initialized to zero, represents the current estimate of the distance of a node from the starting node. All

distances, except  $d(x, x)$ , are initialized to infinite. When the algorithm is running, a value of distance  $d(x, y) = \infty$  indicates that the corresponding node  $y$  has not yet been visited. Finally, a flag  $f$  is used to test if at least one node with the distance  $i + 1$  has been found in the current iteration. This algorithm is structured on three nested loops. The two innermost loops are located at line #1. They check if the nodes  $y$  with actual distance  $i$  from  $x$  have a node  $z$  in their neighborhoods  $I(y)$  that has not yet been visited. If a node  $z$  of this kind is found, its distance from  $x$  is set to  $i + 1$  and the flag  $f$  is set to *true*, indicating that algorithm must execute a new iteration. The loop between lines #1 and #4 is responsible for increasing the variable  $i$  at each iteration and governs the mechanism of fan-shaped expansion that is characteristic of this algorithm. The procedure ends, at line #2, when all reachable nodes have been visited. Such a condition is clearly indicated by a value  $f = \text{false}$  after the execution of the inner loops at line #1.

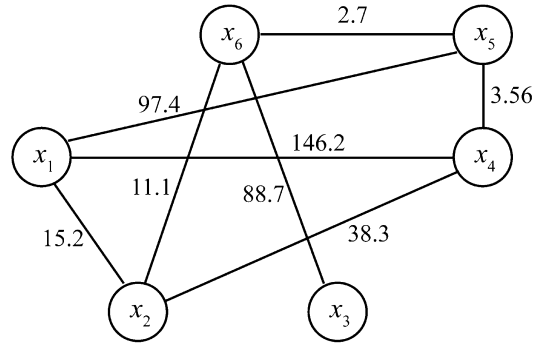
A more general class of data structures is represented by the *weighted graphs*. In many situations, the arcs or the edges of a graph carry information. For example, in a graph representing routes between cities, the nodes would contain information about the cities, while the edges could store the distance in km between any pair of neighbor cities. Therefore, the binary function (A2.2) can be considered as a special case of data structure, such that each arc has unit length.

In the most general case,  $\sigma$  can be defined as a non-negative real function on  $\mathbf{D}^2$ :

$$\sigma : (x_i, x_j) \in \mathbf{D}^2 \rightarrow \sigma_{ij} \in [0, \infty] \quad (\text{A2.7})$$

An example of weighted undirected graph is illustrated in Fig. A2.4. Just like normal graphs, in weighted graphs  $\sigma(x, y) = 0$  is interpreted as absence of direct link between nodes  $x$  and  $y$ . A substantial difference of this class of graphs with respect to unweighted data structures is represented by the possibility that the shortest path between two nodes  $x$  and  $y$  does not coincide with the path having the smallest number

**Fig. A2.4** An example of undirected weighted graph of order 6 and size 8



of intermediate nodes. This is a consequence of the fact that in the case of weighted graphs the distance between two adjacent nodes is not anymore unity but it is given by the real-valued measure  $\sigma(x,y)$ . Consequently, the BFS algorithm is not applicable to weighted graphs.

The following *Dijkstra algorithm* computes the minimum distance between two nodes  $x$  and  $y$  in a weighted graph. Let  $x$  be the starting node and  $d(x,z)$  an ansatz about the distance of a vertex  $z$  from  $x$ . The quantity  $d(x,z)$  will be updated several times during the algorithm execution. At any iteration, the algorithm of Dijkstra considers a subset  $\mathbf{T}$  of  $\mathbf{D}$ , from which a node  $z$  having minimum distance from  $x$  is selected. At this point, for each element  $w$  in the neighborhood of  $z$  and belonging to  $\mathbf{T}$ , such that the distance  $d(x,w) > d(x,z) + \sigma(z,w)$ , the distance  $d(x,w)$  is set to  $d(x,z) + \sigma(z,w)$ . Then, the algorithm updates  $\mathbf{T}$  by removing  $\{z\}$  and the operation is repeated. The algorithm terminates when the selected element of  $\mathbf{T}$  coincides with  $y$ . If  $y$  cannot be reached starting from  $x$ , the algorithm terminates with  $d(x,y) = \infty$ . The function *select()* called at step #1 is a simple linear search in  $\mathbf{T}$ , which returns the first node in  $\mathbf{T}$  having minimum distance from  $x$ . The algorithm of Dijkstra is used in seismic tomography and refraction seismology to determine the best seismic ray path joining a source to a receiver on the basis of a velocity model. In fact, according to Fermat's principle, the path taken between two points by a seismic ray is the one that can be traversed in the least time.

**Algorithm A2.3**  $sp(x,y)$ : Shortest path between two nodes  $x$  and  $y$  (Dijkstra's algorithm).

Input:  $x, y \in \mathbf{D}$

Output:  $d(x,y) \in \mathbb{R}$

```

{
0:  $d(x,x) \leftarrow 0$ ;  $\forall y \in \mathbf{D} \mid y \neq x, d(x,y) \leftarrow \infty$ ;
    $\mathbf{T} \leftarrow \mathbf{D}$ 
1:  $z \leftarrow \text{select}(\mathbf{T})$ 
2:  $z = y \Rightarrow \text{stop}$ 
3:  $\forall w \in I(z) \mid w \in \mathbf{T}, d(x,w) > d(x,z) + \sigma(z,w)$ 
    $\Rightarrow d(x,w) \leftarrow d(x,z) + \sigma(z,w)$ 
4:  $\mathbf{T} \leftarrow \mathbf{T} - \{z\}$ 
5: jump #1
}
```

Both the BFS and Dijkstra algorithms can be used to test the *connectivity* of a data structure. A graph  $G$  is *connected* if for each pair of nodes  $x$  and  $y$  there exists a path joining them. An alternative and widely used technique for determining graph connectivity is known as *depth-first search* (DFS). In this approach, we start from a node  $x_0$  and visit an arbitrary element  $x_1$  in the neighborhood of  $x_0$ . At this point, a node  $x_2$  in the neighborhood of  $x_1$  and not yet visited is considered. Then, the algorithm proceeds through a depth search, until a vertex  $x_k$  is found, such that the neighborhood  $I(x_k)$  is empty or any node in  $I(x_k)$  has been already visited. Then, the depth search restarts from  $x_{k-1}$ . When all the paths rooted at  $x_{k-1}$  have been explored, the algorithm restarts from  $x_{k-2}$ , and so on. Therefore, the procedure alternates depth search with a backward step, until the whole neighborhood of  $x_0$

has been visited, in which case the algorithm terminates. The procedure described above can be easily implemented in recursive form, because the backward steps imply the use of a *stack* where the address of starting depth search nodes is stored. Below we present a variant of the classic algorithm, which can be used to generate a map of the distances of all the nodes of  $G$  from a starting node  $x$ . This algorithm represents the basis for building more complex procedures that analyze properties of graphs.

**Algorithm A2.4**  $dfs(x)$ : Depth-first search from a node  $x$  (shell).

Input:  $x \in \mathbf{D}$   
 Output:  $d(x,y) \forall y \in \mathbf{D}$   
 {  
 0:  $\forall y \in \mathbf{D}, d(x,y) \leftarrow \infty$   
 1:  $dfs(x,0)$   
 }

Algorithm A2.4 represents a shell from which we call the recursive procedure of DFS. It initializes the distance of all the nodes of  $G$  (including  $x$ ) to infinity, which means “unreachable”. At the next step, the algorithm calls the recursive procedure  $dfs()$  listed below, thereby triggering the depth search sequence within  $G$ .

**Algorithm A2.5**  $dfs(y,i)$ : Depth-first search from a node  $y$  (recursive version).

Input:  $y \in \mathbf{D}, i \in \mathbf{N}$   
 Output: None  
 {  
 0:  $d(x,y) \leq i \Rightarrow \text{stop}$

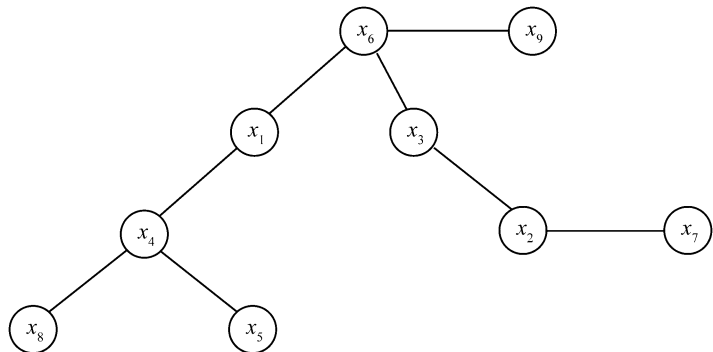
```
1:  $d(x,y) \leftarrow i$ 
2:  $\forall z \in I(y), dfs(z,i+1)$ 
}
```

## A2.3 Trees

*Trees* represent a fundamental class of data structures, characterized by the absence of cyclic paths and by connectivity. Therefore, a tree can be defined as an acyclic connected graph (Fig. A2.5). The nodes having unit degree are called *terminal nodes* or *leaves*, while the remaining vertices are called *internal nodes*. An interesting property of the trees is that removal of a single edge is sufficient to separate the data structure into two disjoint parts. This feature implies the lowest degree of connectivity for a graph. The following theorems determine the fundamental properties of this important class of graphs.

**Theorem A2.1** *A graph  $G$  is a tree iff any pair of nodes is joined by a unique path.*

*Proof* By definition  $G$  is connected. Consequently, any pair of nodes is joined by at least one path. However, if two nodes  $x$  and  $y$  were joined by two or more distinct walks,  $G$  would have cyclic paths, which would contradict the hypothesis that  $G$  is a tree. Let us assume now that any pair of nodes in  $G$  is joined by a unique path. In this instance,  $G$  is clearly a connected graph. Furthermore, if we could find a cyclic path passing through two nodes  $x$  and  $y$ , then these two elements would be joined by distinct paths,



**Fig. A2.5** An example of tree of order 9

which would contradict the starting hypothesis. As a consequence, such a path cannot exist and  $G$  is a tree. ■

**Theorem A2.2** *A graph  $G$  is a tree iff  $G$  is connected and  $m = n - 1$ .*

*Proof* Given a tree  $G$  of order  $n$ , we can prove that  $m = n - 1$  by induction. In fact, it is trivially true that if  $n = 1$  then  $m = 0$ . Let us assume now that the equality is proved up to some order  $n - 1$ , and consider a tree of order  $n$ . If we remove the edge joining two adjacent nodes  $x$  and  $y$ , we obtain two disjoint subtrees  $G_1$  and  $G_2$ , each having order  $n_i < n$  ( $i = 1, 2$ ) and size  $m_i = n_i - 1$ . Therefore:

$$m = m_1 + m_2 + 1 = n_1 + n_2 - 1 = n - 1$$

Conversely, let  $G$  be connected and  $m = n - 1$ . We must prove now that  $G$  is acyclic. To this purpose, let us assume that a cyclic path  $c$  exists in  $G$ . If we remove an edge from  $c$ , we obtain a graph that is still connected and of the same order of  $G$ , but the size will be decreased to  $n - 2$ . This result is clearly impossible, because a connected graph must have at least  $n - 1$  edges. Therefore,  $G$  must be acyclic, thereby it is a tree. ■

Let us consider now a subclass of tree structures that has great practical interest: the *rooted trees*. These graphs are directed acyclic structures characterized by the existence of a special node  $r$ , called the *root*, from which it is possible to reach any other vertex in the graph through unique paths (Fig. A2.6). The following theorem determines all the fundamental properties of these graphs:

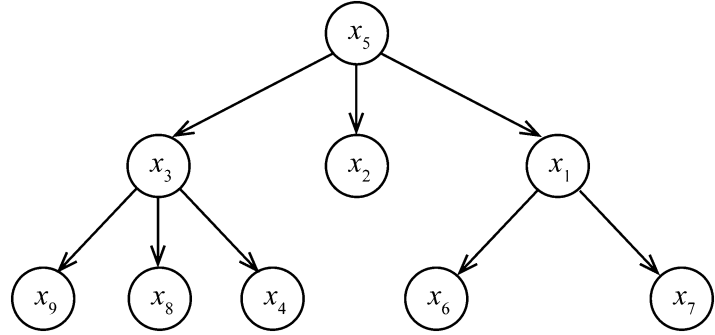
**Theorem A2.3** *A directed connected graph  $G$  is a rooted tree iff: (a) there exists a unique unreachable node  $r$  in  $G$  (that is, without incoming arcs); (b) any other node in the graph has one input arc. Furthermore, removal of a single arc separates  $G$  in two disjoint subtrees.*

*Proof* Let  $G$  be a rooted tree. In this instance, there exists a unique path joining the root  $r$  with

any other node  $x$ . Let us assume now that a node  $x$  can be found such that  $x \rightarrow r$ . In this instance, the arc from  $x$  to  $r$  would form, together with the path from  $r$  to  $x$ , a cycle. This would contradict the hypothesis that  $G$  is a rooted tree. Therefore,  $r$  cannot have incoming arcs. Let us consider now a node  $x \neq r$ . Clearly,  $x$  has at least one incoming arc, because there exists a path joining  $r$  to this node. Let us assume that  $x$  has more than one input arcs, and let  $z$  and  $y$  be two nodes such that  $z \rightarrow x$  and  $y \rightarrow x$ . Two distinct paths exist that join  $r$  with  $z$  and  $y$ . Therefore, extending these paths by the arcs  $z \rightarrow x$  and  $y \rightarrow x$ , we would obtain two alternative paths that reach  $x$  starting from the root. This would be in contrast with the definition of rooted tree, thereby a node  $x \neq r$  must have a unique incoming arc. We also observe that removal of any arc  $x \rightarrow y$  would cause  $y$  to be unreachable, thereby the tree would be decomposed in two disjoint parts. As a consequence, the root  $r$  is the only node devoid of incoming arcs. Now, let  $G$  be a directed graph containing a unique node  $r$  without incoming arcs and let us assume that any other node has just one incoming arc. We want to prove that in this case  $G$  is a rooted tree. To this purpose, we first observe that at most one path can exist joining  $r$  to any other node  $x$ , because the presence of two alternative paths from  $r$  to a node  $x$  implies the existence of a convergence node that could coincide with  $x$  or an “upstream” node. This node would have two input arcs, in contrast with the starting hypothesis. Regarding the acyclicity of  $G$ , it is sufficient to observe that the root  $r$  cannot be included in any cycle, while any cycle not including  $r$  should have at least an input arc at some node  $z$ . This node would be joined to the root through a path that external to the cycle, and to itself through the cyclic path. Therefore,  $z$  would have two incoming arcs, which contrasts the starting hypothesis. Consequently,  $G$  is a rooted tree. ■

The previous theorem has the immediate corollary that the adjacency matrix  $A$  will have a column, corresponding to the root, whose elements are all zero. Conversely, the remaining columns will have one element set to unity and  $n - 1$  elements set to zero. Therefore, the

**Fig. A2.6** An example of rooted tree of order 9



adjacency matrix of a rooted tree is easily recognizable. We also note that the nodes of these structures can be classified according to their *level*. This quantity simply represents the distance from the root node. The *height* of a rooted tree is defined as the maximum number of levels in the tree (including level zero). For example, the rooted tree of Fig. A2.6 has height three. A *traversal* of a tree structure is an algorithm that visits each node once, independently from the type of access that is performed to the nodes. These algorithms have great practical importance and derive from the DFS algorithm discussed above. However, in this application the starting node always coincides with the root and there are two different traversal modes. In a *preorder* traversal, access to each node is performed *before* calling the depth search, whereas in a *postorder* algorithm, data access is accomplished *after* completion of depth search. Let  $\mathbf{T}(r)$  be a rooted tree with root  $r$ . If  $x \in \mathbf{D}$  is a node, we indicate by  $\mathbf{T}(x)$  the subtree obtained from  $\mathbf{T}(r)$  after removal of the incoming arc to  $x$ . Clearly,  $\mathbf{T}(x)$  is also a rooted tree and its root is  $x$ . If  $J(x)$  is the set of nodes  $y$  that can be reached in one step from  $x$ , it is easy to modify the recursive DFS algorithm to the specific case of rooted trees.

**Algorithm A26** *preorder(x)*: Preorder traversal from a node  $x$  (recursive version).

Input:  $\mathbf{T}(x)$   
 Output: Nessuno  
 {  
 0: *access(x)*

1:  $\forall y \in J(x), \text{preorder}(\mathbf{T}(y))$   
 }

**Algorithm A27** *postorder(x)*: Postorder traversal from a node  $x$  (recursive version).

Input:  $\mathbf{T}(x)$   
 Output: Nessuno  
 {  
 0:  $\forall y \in J(x), \text{postorder}(\mathbf{T}(y))$   
 1: *access(x)*  
 }

For both these algorithms, the function *access(x)* represents any access procedure to the data associated with node  $x$ . In the case of the tree in Fig. A2.6, the two algorithms would generate the sequences:

*Preorder* :  $x_5 x_3 x_9 x_8 x_4 x_2 x_1 x_6 x_7$

*Postorder* :  $x_9 x_8 x_4 x_3 x_2 x_6 x_7 x_1 x_5$

These sequences imply an ordering criterion within the sets  $J(x)$  for each  $x \in \mathbf{D}$ . For example, in the case of the rooted tree of Fig. A2.6, we have:  $J(x_5) = \{x_3, x_2, x_1\}$ , thereby the subtrees  $\mathbf{T}(x_3)$ ,  $\mathbf{T}(x_2)$ , and  $\mathbf{T}(x_1)$  will be traversed in this order. Clearly, any other ordering of the sets  $J(x)$  will produce a different sequence. In general, the rooted trees are useful when the data must be arranged according to a *hierarchical order*. In plate kinematics, for example, this is the case of the rotation models that describe the set of kinematic relations between conjugate pairs of tectonic plates.

---

# Index

## A

Absolute magnetic permeability, 91  
Absolute velocity field, 62, 63  
Active asthenospheric drag, 347  
Active upwelling, 13, 23  
Adiabatic, 5, 8, 10, 13, 15, 246, 326, 339, 341, 358  
Adiabatic temperature gradient, 341  
Aftershock, 293  
Age-distance plot, 159, 163, 164  
Alkali atoms vapor sensors, 146  
Alpine style of orogeny, 45  
Alternate field demagnetization, 183  
Alternate path, 64–66  
Amonton's law, 243  
Ampere's law, 86, 88, 105, 109  
Amphibolites, 6  
Amplitude transport equation, 258, 259  
Andesitic, 6  
Anelastic behaviour, 241  
Angular acceleration vector, 207–212  
Angular momentum, 30, 34, 72, 75, 84, 89, 96–98, 206, 302  
Anisotropy, 11, 15, 16, 91, 98, 178, 179, 347  
Antiferromagnetic, 99, 100  
Antipole, 33, 211, 212, 214  
Apparent polar wander (APW) paths, 159, 160, 192–200, 202, 204, 206, 219, 221  
Areal flux, 345, 346, 349  
Aspect ratio, 361  
Asthenosphere, 3, 7–19, 22, 23, 30, 31, 40, 42, 44, 72, 201, 206, 239, 243, 250, 270, 306, 307, 321, 326, 327, 329, 331, 333, 335, 337, 341, 344–348, 359, 375–377  
Asthenospheric counterflow, 346  
Asymmetric deformation, 67  
Asymmetric rifting, 44  
Asymmetric strain tensor, 303–304  
Auxiliary fault plane, 293

## B

Base-station magnetometer, 146, 147  
Beach ball, 294  
Biharmonic equation, 350, 353  
Biot-Savart law, 81  
Blocking energy, 182

Body force density, 30, 236, 339

Body forces, 30, 227, 236, 244, 253, 286, 290, 293, 333, 334, 339

Bohr magneton, 97

Boundary conditions, 86, 87, 321, 322, 324, 325, 329, 342, 344, 345, 347, 353, 361, 364

Boussinesq approximation, 354, 356–358

Bow shock, 120

Bow wave. *See* Head wave

Brittle deformation, 243

Brittle-ductile transition zone, 243

Brune's formula, 301, 302, 305, 306, 313, 315

Bulk Earth Urey ratio, 320

Bulk modulus, 5, 246

Bulk properties, 29

Bulk viscosity, 340

Buoyancy force, 357, 359, 360

Burgers model, 241

*b*-value, 298

Byerlee's law, 243, 280

## C

Calibration points, 117

Canonical orthogonality relations, 365

Canted antiferromagnetism, 99

Carbonatite liquids, 10

Cauchy

momentum equation, 235–236, 246, 286

stress surface, 231

theorem, 229–232, 253, 343

Caustics, 268, 269

Center of mass, 30

Central Atlantic Magmatic Province (CAMP), 13

Centrifugal force, 366

CFD. *See* Computational fluid dynamics (CFD)

Chemical composition, 3, 4, 6, 14, 15, 20, 94, 354

Chemical remanent magnetization (CRM), 114–115, 197

Chron, 126, 139, 171, 173–178, 228

CIG. *See* Computational Infrastructure for Geodynamics (CIG)

Clapeyron slope, 34–36

Classic gyromagnetic ratio of the electron, 98

Clausius-Clapeyron relation, 35

Cleaning, 196–199, 209

Closure rule, 62, 84

- CMB. *See* Core-mantle boundary (CMB)  
COBs. *See* Continent-ocean boundaries (COBs)  
Coda, 271, 274  
Coefficient of thermal conductivity, 332, 358  
Coefficient of thermal expansion, 22, 344, 369  
Coercive field, 107, 110  
Compensation depth, 345, 347, 389  
Complete Bouguer anomalies, 388, 389  
Components, 10, 17, 20, 21, 28–30, 34, 36, 45, 47–50, 71, 76, 77, 84, 85, 87–89, 96, 97, 99, 100, 103, 107, 110–112, 115, 121, 122, 124, 127, 130–132, 139, 143–145, 150, 153, 155, 156, 160, 161, 172, 182–184, 204, 218–221, 228–236, 239, 241, 245, 246, 250, 252–254, 257, 266, 272, 273, 277, 285, 287, 288, 290–294, 296, 302, 303, 310, 312, 313, 319, 327, 338, 342, 344, 346, 349, 350, 352, 354, 356, 357, 375, 377, 379, 382, 385  
Composition, 3–25, 63, 94, 95, 124, 244, 250, 271, 319, 323, 354  
Compressible fluids, 370  
Computational fluid dynamics (CFD), 340, 354  
Computational Infrastructure for Geodynamics (CIG), 354  
Conductors, 81, 87, 104, 105, 108, 110  
Conjugate, 13, 14, 38–40, 44, 48, 49, 51, 52, 54–66, 159–161, 164, 166–168, 170, 199, 206, 293, 393  
    plate boundaries, 58, 65  
    plates, 38–40, 48, 49, 54–58, 62, 63, 65, 159–161, 199  
Conservation of energy, 269, 321, 341, 342, 344, 346, 358  
Constitutive equation, 237–245, 253, 337, 339  
Continental crust  
    deep, 4, 6  
    middle, 6  
    upper, 4, 6  
Continental geotherm, 320–323  
Continental tectonics, 66, 68  
Continent-ocean boundaries (COBs), 20, 26, 65–67, 70, 71, 73–75, 126, 362  
Continuity equation, 337–339, 361  
Continuum mechanics, 4, 29–31, 33, 236, 257, 337, 385  
Convection, 17, 21–23, 25, 104, 239, 240, 317, 319, 320, 328, 339, 341, 348, 354–360, 370, 376, 377  
Convective current density, 104  
Converted phases, 274  
Convolution, 171, 174  
Cooling of the oceanic lithosphere, 323, 326–332  
Core complex, 45  
Core-mantle boundary (CMB), 8, 25, 39, 104, 118, 247, 250, 251, 261, 264, 265, 275, 276, 289, 290, 319, 333, 376, 377, 390, 391  
Corner flows, 335, 353, 361  
Coronal mass ejections (CME), 121, 122  
Coseismic phase, 294, 295  
Couette flow, 345–348, 377  
Coulomb energy, 112  
Coulomb's law, 104  
Counterflows, 18, 19, 346  
CQZ. *See* Cretaceous Quiet Zone (CQZ)  
Creep, 232, 239–241, 243, 244, 301, 306  
Cretaceous Quiet Zone (CQZ), 118, 214  
Critical blocking temperature, 196  
Critical incidence angle, 278, 279  
Critical Rayleigh number, 374  
CRM. *See* Chemical remanent magnetization (CRM)  
Crossing points, 165–167  
Crossover point, 161, 162, 288  
Crustal magnetic field, 135, 138  
Crustal phases, 273  
Cryogenic magnetometer, 197  
C-series, 118  
Curie temperature, 107–109, 111–114, 196, 198, 199  
Current plate motions, 83–90  
Cyclotron frequency, 96
- D**  
Dashpot, 238, 240  
Declination, 36, 112–115, 139, 140, 156, 159, 160, 182, 185, 186, 188, 189, 191, 192, 196–200  
Declination anomaly, 114  
Deformation, 11, 14–16, 29, 31, 41, 42, 47, 50, 54, 59, 66–68, 70, 227, 232–235, 237–239, 241–245, 249, 250, 253, 279, 287, 291, 293, 294, 298, 301–315, 341, 342, 376  
Deformation pattern, 307  
Degree of melting, 10, 11  
Dehydration, 16, 19  
Delay time, 269–271  
Demagnetizing factors, 178  
Demagnetizing field, 179, 180  
Density, 3–8, 11, 16, 17, 21–24, 30, 81, 86, 89, 90, 104, 105, 120, 122, 126, 130, 136, 137, 150, 152–155, 186–188, 236, 245, 250, 251, 253–255, 268, 269, 274, 304, 319, 320, 323, 330–334, 337–339, 341, 354, 355, 357–359, 363, 365, 366, 369, 370, 372–377, 379  
    of charge, 81  
    contrasts, 21, 333, 374, 376, 377  
    of current, 81, 104, 105  
    of mass, 30  
Depth phases, 275  
Detrital remanent magnetization (DRM), 101, 183  
Deviator, 232  
DGRF, 148  
Diamagnetism, 88, 89  
Dieterich-Ruina law. *See* Rate-and-state friction law  
Differentiation, 4, 6, 20, 171, 326  
Diffusion  
    equation, 106, 320, 321, 325, 326  
    limit, 106  
    time, 106  
Digital terrain model, 374  
Dilatation, 234, 235, 245–249, 252  
Dip, 12, 13, 126, 285, 293, 296, 302, 304–309, 311, 312  
Dipole-dipole interaction, 97, 98, 178, 179  
Dipole equation, 114, 189, 191, 217–219  
Dipole field, 84, 85, 112, 115, 186, 188, 189, 217, 218, 220

Direct velocity effect, 281  
 Dirichlet boundary-value problem, 127, 130  
 Displacement field, 232–236, 246, 249, 279, 280, 286, 287, 289–291, 302  
 Dissipation function, 342, 343  
 Diurnal correction, 144, 147, 148  
*D'' layer*, 25  
 Double couples, 289–293  
 Doublet flow, 351, 352  
 Doublet strength, 352  
 Downdip extension, 306, 314  
 Downdip shortening, 306  
 Drift curve, 148, 149  
 DRM. *See* Detrital remanent magnetization (DRM)  
 Dry peridotite solidus, 10, 322  
 Ductile deformation, 243, 244  
 Dynamic form factor, 368  
 Dynamic friction coefficient, 280–284  
 Dynamic pressure, 120, 121, 335  
 Dynamic topography, 346, 370, 375–377

## E

Earthquake focus, 285, 291, 294  
 East Coast Magnetic Anomaly (ECMA), 52  
 Eccentric dipole field, 112, 218  
 Eckert number, 347  
 ECMA. *See* East Coast Magnetic Anomaly (ECMA)  
 Edge-driven convection, 348  
 EGM2008, 365, 369  
 Eigenvalue, 32, 136, 230, 246, 293, 305  
 Eigenvector, 32, 136, 230, 235, 309  
 Eikonal equation, 257–259  
 Eikonal surface, 260–261  
 Elastic behavior, 237, 240, 241, 245  
 Elastic-ductile lithosphere, 14  
 Elastic modulus, 237, 239, 240  
 Elastic rebound theory, 279, 280  
 Elastic tensor, 253, 254  
 Elasto-plastic rheology, 242, 243  
 Electric conductivity, 104  
 Electric current, 81–83, 86, 103, 104, 108  
 Electric fields, 86, 104–106, 108, 178  
 Electromagnetic induction, 106  
 Electromotive force, 108, 111  
 Electrostatic field, 105, 107  
 Electrostatic potential, 107  
 Ellipticity coefficient. *See* Dynamic form factor  
 Endothermic, 21, 22, 333  
 Energy balance, 341–344  
 Energy density, 136, 137, 253–255, 268, 269, 274, 323, 343  
 Epochs, 74, 115, 139, 143, 148, 190  
 Equation of conductive heat transfer, 320, 321  
 Equation of state, 339, 341, 354–356, 358  
*E* region, 123  
 Eulerian formulation, 337  
 Euler pole, 32, 33, 39, 55–57, 59–66, 69, 74, 75, 77, 143, 166, 168–170, 202, 204, 206, 221  
 Euler's theorem, 30–34, 57, 62, 308

Euler vector, 32, 34, 48, 56, 62, 69, 70, 72–75, 164, 204, 207  
 Evolution of triple junctions, 64  
 Exchange energy, 98  
 Exothermic, 21, 22, 333  
 Extensive variables, 30, 3330

## F

Faraday  
   dynamo, 108  
   experiment, 108  
   law, 106–108, 110  
 Far-field, 91, 290, 293, 294  
 Far-sided effect, 217  
 Ferrers normalization, 133, 134  
 Ferrimagnetic solids, 99  
 Ferromagnetism, 88, 95, 96, 99, 177  
 Finite reconstruction, 57, 58, 62, 66, 68, 69, 143, 165–167, 199  
 Finite rotations, 33, 55, 58, 66, 67, 168–170, 199, 202  
 First-motion displacement, 291  
 Fisher's statistics, 186, 192  
 Fit of Pangaea, 59, 60, 220  
 Flattening, 19, 221, 367–369, 372  
 Flattening factor, 221, 367, 368  
 Flower structures, 43  
 Flow lines, 10, 54, 56, 66, 67, 109, 119, 309  
 Fluctuation wavelength, 360  
 Focal mechanism, 36, 257, 285, 293–296, 302, 306, 307, 314, 315  
 Focal sphere, 294, 295  
 Force couple, 83, 288–290, 292, 293  
 Formula of Somigliana, 368, 369, 372  
 Forward modelling, 151, 157–164, 172  
 Fourier's law, 317–320, 329  
 Fourier transform, 158, 170–172, 174, 175, 290  
 Fracture zones, 19, 37, 38, 43, 49, 55–58, 64, 66, 143, 164–170, 207, 335, 363, 373  
 Frank's model, 308, 309  
 Free-air anomaly, 373  
 Free-air correction, 373  
 Free boundary, 64–66, 70  
 Free space permittivity, 98  
 Full normalization, 365  
 Full spreading rate, 13, 37, 38, 159, 162, 164, 332  
 Fully normalized harmonics, 365

## G

GAD. *See* Geocentric axial dipole (GAD)  
 Gaussian curvature, 307–309  
 Gauss's law  
   coefficients, 134, 139, 218  
   law, 106, 126, 127  
   theorem of the arithmetic mean, 130  
 Geocentric axial dipole (GAD), 114, 115, 121, 159, 160, 188, 190, 199, 217–220  
 Geocentric reference frames, 34, 35, 74, 75, 77, 190  
 Geodetic latitude, 368

Geodetic methods, 70  
 Geodetic-seismic strain rate, 305  
 Geodynamo, 107–112, 340  
 Geographic coordinate system, 34  
 Geoid, 364–370, 374, 376, 377  
 Geoid undulations, 369, 370, 376, 377  
 Geomagnetic coordinate system, 35, 156  
 Geomagnetic polarity time scales, 70, 115–119, 149, 157, 159, 161  
 Geomagnetic pole, 112, 113, 115, 191  
 Geomagnetic storms, 121, 122  
 Geometrical fit, 60, 61  
 Geometrical spreading, 259–262, 268, 274, 297  
 Geometric altitude, 372  
 Geopotential models, 365  
 Geotherm, 8–10, 17, 124, 320–324, 329, 335, 341, 347  
 Global heat flow, 319  
 Global Paleomagnetic Database (GPMDB), 192  
 Grain size, 94–96, 237–319  
 Granulites, 6, 15, 94  
 Gravimeters, 372, 377  
 Gravitational constant, 150, 363, 369  
 Gravitational geopotential, 363  
 Gravity anomalies, 37, 38, 52, 151, 158, 346, 363, 372, 373  
 Gravity field, 86, 150, 155, 331, 342, 363–377  
 Green identities, 126  
 Green's function, 150  
 Gutenberg-Richter law, 297–298, 306

## H

Half-space cooling (HSC) model, 328  
 Harmonic function, 86, 87, 127, 130, 145, 173, 174, 350  
 Head wave, 264, 265  
 Healing, 282, 283  
 Heat  
   conduction, 01, 318, 323–327, 342  
   flux, 317–320, 322, 323, 326, 328, 329, 379  
   power, 343  
   radiation, 317  
 Hellinger's algorithm, 168–170  
 Helmholtz's decomposition theorem, 287  
 Helmholtz's equation, 136, 251  
 HEP. *See* Hot spot Euler pole (HEP)  
 Homogeneous equation of motion, 236  
 Hooke's law, 237, 245–246, 253, 254  
 Horizontal slowness, 263  
 Hot spot Euler pole (HEP), 202, 204  
 Hot-spots, 23, 201, 204  
 Hot spot tracks, 201–203  
 Hydrodynamic lift, 335  
 Hydrostatic  
   pressure, 8, 120, 228, 232, 238, 243, 244, 246, 322, 331, 341, 357, 358, 375  
   strain, 234  
   stress, 230, 232  
 Hydrous melting, 19

Hypocenter, 285, 313  
 Hysteresis loop, 93

## I

IAB. *See* Island arc basalts (IAB)  
 IAGA. *See* International Association for Geomagnetism and Aeronomy (IAGA)  
 ICB. *See* Inner-core boundary (ICB)  
 IGRF. *See* International Geomagnetic Reference Field (IGRF)  
 IMF. *See* Interplanetary magnetic field (IMF)  
 Incidence angle, 262–265, 269, 273–275  
 Inclination, 36, 101, 112–115, 139, 140, 155, 156, 159–162, 182, 185, 186, 188, 189, 191, 197, 199, 217–221  
   anomaly, 115, 219–221  
   shallowing, 199  
 Incompatible elements, 11, 15, 18  
 Incompressible fluids, 339, 340, 344, 345, 349, 353, 356, 361  
 Incompressible Newtonian fluid, 339  
 Induced magnetization, 89–91, 93, 94, 111, 123, 144  
 Inertial tensor, 34, 73, 372  
 Infinitesimal deformation, 233, 234  
 Initial oceanic crust, 13  
 Initiation of subduction, 333  
 Inner-core boundary (ICB), 104, 250, 251, 275  
 Intensive variables, 29, 330, 337  
 Intermagnet network, 147  
 Internal energy, 92, 93, 342–344  
 Internal energy density, 343  
 Internal heating ratio, 320  
 International Association for Geomagnetism and Aeronomy (IAGA), 124  
 International Geomagnetic Reference Field (IGRF), 130–140, 143, 144, 148, 156, 159  
 International Terrestrial Reference Frame (ITRF), 74  
 Interplanetary magnetic field (IMF), 119–121  
 Interseismic phase, 279, 280  
 Inverse modelling, 158, 161  
 Ionosphere, 103, 118–123, 132, 143, 144, 147  
 Isentropic adiabat geotherms, 8  
 Island arc basalts (IAB), 4  
 Isochron map, 49, 149, 163–168  
 Isochrons, 49, 149, 163–168, 170, 175  
 Isostasy principle, 331, 375  
 Isostatic compensation, 375, 376  
 Isostatic equilibrium, 239, 331–333, 375  
 Isothermal compressibility, 330, 355  
 Isotropic material, 104, 253, 257  
 Isotropic stress, 232  
 ITRF. *See* International Terrestrial Reference Frame (ITRF)

## J

Jeans relation, 136  
 Jurassic Quiet Zone (JQZ), 118

**K**

- Kelvin rheology, 240, 241
- K index, 122
- Kinetic energy, 33, 120, 253, 254, 342
- Kinetic energy density, 253, 254
- Koenigsberger ratio, 94, 123
- Kostrov's formula, 301–302, 304–306, 313, 315

**L**

- Lagrangian representation, 337
- Lam parameters, 245, 247, 249, 255
- Langevin function, 90, 91
- Langevin's equation, 90
- Laplace's equation, 86, 87, 126–128, 130–132, 134, 136, 321, 364
- Large igneous provinces (LIP), 13
- Large low-shear-velocity provinces (LLSVP), 23, 25
- Larmor angular frequency, 84
- Lateral deformation of slabs, 314
- Lateral strain, 308, 311–314
- Lattice-preferred orientations (LPO), 15
- Layer
  - A, 13, 99
  - B, 13, 99, 124
  - C, 124
- Legendre associate polynomials, 132
- Legendre equation, 132
- Levelling, 148–149
- Linear velocity, 32, 48, 62, 69, 72, 74, 175, 205
- LIP. *See* Large igneous provinces (LIP)
- Lithosphere, 3, 7, 14–19, 39–41, 44, 72–75, 201, 203, 204, 206, 210, 212, 227, 237, 239, 243, 244, 270, 271, 276, 298, 301–315, 317–336, 344, 361, 369, 375
- Lithosphere-asthenosphere boundary (LAB), 15–17, 73, 74, 183, 319, 327, 341, 344, 347, 348, 361
- Lithospheric mantle, 13–16, 19, 321, 323, 341, 376, 377
- Local coordinate systems, 36, 85, 160, 195, 230
- Local magnitude, 38, 297
- Long-wavelength anomalies, 172
- Lorentz force, 105, 107–109, 120, 122
- Love waves, 274, 275
- Lower crust, 6, 44, 45, 323, 335
- Lower mantle, 3, 18, 21–25, 42, 202, 237, 239, 250, 333, 370, 377
- Low-velocity zone (LVZ), 270–272
- LPO. *See* Lattice-preferred orientations (LPO)

**M**

- MacCullagh formula, 370–372
- Magan, 158
- Magnesiowüstite, 5, 20–22, 333
- Magnetic anisotropy, 91
- Magnetic anomalies, 14, 38, 39, 50, 52, 55, 57, 66, 70, 115–118, 121, 124, 125, 140, 143–175, 379
  - interpretation, 148, 157, 161, 162, 168
- Magnetic basement, 117, 123

- Magnetic diffusivity, 106
- Magnetic dipole, 35, 84, 85, 87, 89, 90, 97, 98, 112, 114, 154, 189, 217
- Magnetic domains, 92, 94, 95
- Magnetic field intensity, 91, 124, 148
- Magnetic fields, 30, 35, 36, 81–91, 94, 97, 99, 103–112, 114, 119–124, 126, 136, 143–146, 148, 149, 154, 156, 218, 227, 379
- Magnetic flux, 108, 109
- Magnetic induction equation, 106
- Magnetic moment, 35, 83–90, 92, 93, 96–100, 112–115, 122, 145, 146, 154, 178, 183, 189
- Magnetic permeability, 82, 91
- Magnetic poles, 112
- Magnetic pressure, 120, 136
- Magnetic susceptibility, 90, 91, 93, 99, 100, 123
- Magnetization, 30, 39, 81–101, 103, 116, 117, 121, 123–126, 143–145, 149, 150, 154, 155, 157–165, 177–185, 188, 189, 194, 220
- Magnetization current, 88, 91, 130
- Magnetohydrodynamics (MHD), 103, 112
- Magnetopause, 120, 121
- Magnetosheath, 120
- Magnetosphere, 103, 118–123, 132
- Magnetotail, 119, 120, 122
- Main geomagnetic field, 103–107, 112
- Mantle minerals, 3, 5, 18, 20, 245, 246
- Mantle phases, 275
- Mantle plume, 12, 23, 63, 210, 348, 349, 352
- Material derivative, 338, 339, 342, 344, 358
- Maxwell rheology, 238, 239, 241
- Maxwell's equations, 85–87
- Maxwell's relations, 8
- Mean paleopoles, 192, 193, 201
- Mechanical parameters, 5, 247, 257
- Mechanical power, 342, 343
- Median filtering, 148, 149
- Median filtering algorithm, 149
- Melting regime, 10, 11, 15, 326, 327
- MHD. *See* Magnetohydrodynamics (MHD)
- Micro-coercivity, 180, 183, 184
- Microlevelling, 149
- Mid-ocean ridge basaltic (MORB), 7, 9, 11, 13, 14, 16, 18–20, 116, 123, 326, 331
- Mid-ocean ridges, 7, 8, 10, 18, 19, 36–39, 42, 49, 55, 70, 72, 124, 326, 377
- Mid-ocean ridge velocity (MORVEL), 70, 71, 74–76
- Minerals, 3–6, 11, 14, 15, 18–20, 22, 29, 81–101, 123, 124, 184, 243, 245, 246, 272, 319, 323, 333
- Mixed boundary-value problem, 128
- Moho, 15, 52, 247, 272–274, 319, 322, 323
- Mohorovičić discontinuity, 15
- Molnar's formula, 314
- Moment magnitude, 112, 297
- Moment rate function, 291
- Moments of inertia, 34, 371, 372
- Moment tensor, 292–296, 302, 304, 305
- Moment tensor components, 292, 296, 302
- Momentum density, 338

Momentum of inertia, 33, 34  
 Monochromatic plane wave, 252, 274  
 Monochromatic waves, 251  
 MORB. *See* Mid-ocean ridge basaltic (MORB)  
 MORVEL. *See* Mid-ocean ridge velocity (MORVEL)  
 M-series, 118, 119  
 Multi-domain (MD) grains, 95, 184

## N

Narrow rift, 44  
 Natural remnant magnetization (NRM), 100, 155, 159, 177, 182–184, 189, 199  
 Natural smoothing spline, 196  
 Navier-Stokes equations, 339–341, 344, 345, 349, 353, 356, 357  
 Near field, 272–274, 290, 291  
 Néel's theory, 177–182  
 Néel temperature, 99  
 Neumann boundary-value problem, 128  
 Newtonian fluids, 238–241, 354  
 NNR-NUVEL-1A, 74  
 Non-conservative electric field, 108  
 Non-equilibrium states, 182, 207, 208, 347  
 No-net-rotation (NRR), 73  
 Non-volcanic passive margins, 13  
 Normal gravity, 154, 367–369, 372  
 Normal stress, 228, 230, 231, 243–245, 280, 281, 284  
 No-slip boundary condition, 342, 344, 345, 361  
 NRM. *See* Natural remnant magnetization (NRM)  
 NUVEL-1, 70  
 NUVEL-1A, 69, 70

## O

Oblate spheroid, 367, 372  
 Ocean-continent transition (OCT), 12–14, 52  
 Ocean-continent transition zone, 12  
 Ocean floor depth, 327, 331, 332  
 Oceanic crust, 3, 4, 7–15, 19, 37, 40, 41, 50–52, 116, 117, 124, 126, 143, 149, 157, 160, 164, 227, 247, 319, 320, 326, 328, 333, 376  
 Oceanic island arc basalts, 4  
 Oceanic trenches, 7  
 Ocean island basalts (OIB), 4  
 OCT. *See* Ocean-continent transition (OCT)  
 Octupole components, 218–220  
 Ohm's law, 104, 105, 111  
 OIB. *See* Ocean island basalts (OIB)  
 Ophiolite sequences, 11  
 Order of a tectonic plate, 48  
 Orogenic belts, 43, 45  
 Orthogonality conditions, 133, 134, 137  
 Orthogonality relations, 364, 365  
 Orthometric height, 365, 366  
 Outer core, 104, 106, 107, 112, 237, 275, 276  
 Oval of 95% confidence, 191  
 Overhauser effect, 145, 146  
 Over-relaxation factor, 87

## P

Paleolatitude, 114, 160, 161, 192, 195–200, 202, 219, 221  
 Paleolongitude, 190, 201, 202  
 Paleomagnetic  
   directions, 159, 183, 185–188, 190, 194, 199  
   Euler poles, 202, 203, 206  
   field, 84, 115, 155, 159, 186, 190, 217–221  
   frame, 62, 200, 201  
   reference frame, 36, 199–203, 205  
 Paleopole, 159, 161, 188–199, 201–203, 217, 219  
   transfer, 196, 197  
 Paleotectonic reference frame, 62, 63, 200  
 Palinspastic restoration, 54  
 Parallax correction, 147  
 Paramagnetism, 88  
 Pareto distribution, 304  
 Parseval identity, 172  
 Partial demagnetization, 183  
 Partial melting, 9, 10, 16, 18, 19, 272, 321  
 Partial uncertainty rotation (PUR), 169, 170  
 Pauli exclusion principle, 98, 99  
 PCM. *See* Plate cooling model (PCM)  
 Peridotite ridge, 14  
 Peridotite solidus, 8, 10, 322  
 Perovskite, 5, 20–25, 333  
 Phase(s), 3, 4, 6, 7, 10, 12, 16, 20–23, 25, 44, 63, 67, 68, 83, 94, 115, 122, 145, 146, 148, 171, 172, 207, 209–211, 215, 216, 251, 257, 258, 271–281, 284, 294, 295, 319, 333, 354  
 Phase transformations, 20, 271  
 Picking, 272, 276  
 Piclogite model, 20  
 Plane waves equation, 248  
 Plasma, 106, 118–120, 122, 237, 245  
 Plastic rheology, 242, 243  
 Plate boundaries, 12, 36–49, 58, 59, 63–65, 70, 72, 77, 78, 146, 165, 201, 206, 207, 209, 245, 279, 296, 302, 305, 308, 335, 370  
 Plate circuit, 54–59, 61–64, 70, 72  
 Plate cooling model (PCM), 328–333  
 Plate reconstructions, 40, 45, 50, 51, 54, 62, 67, 166, 202, 214  
 Plate tectonic configuration, 45–47, 63–69  
 Plate traversal algorithm, 46–49  
 Plumb lines, 36, 365, 366  
 Plume generation zones (PGZ), 25  
 Poiseuille-Couette flows, 345, 346, 377  
 Poiseuille flow, 345  
 Poisson's equation, 287, 363  
 Poisson solid, 246, 250, 274  
 Poisson's ratio, 246  
 Poisson's relation, 154, 155  
 Polar cusps, 119, 120  
 Polarity inversions, 115–118  
 Pole of acceleration, 212, 214–216  
 Polynomial fitting, 148, 149  
 Postglacial rebound, 239, 370, 377  
 Post-perovskite, 25  
 Postseismic phase, 279  
 Potential energy, 97, 178, 180, 253, 254

Potential field, 52, 86, 126, 130, 170, 172, 320, 350, 363, 383  
 Potential temperature, 9, 10, 13, 15, 326  
 Power, 146, 342, 343, 370, 385, 386  
 Power-law creep, 243  
 Power spectrum, 138  
 Prandtl body, 242  
 Prandtl number, 347  
 Precision parameter, 186–188, 191  
 Preliminary Reference Earth Model (PREM), 250, 251, 270, 276, 322  
 Pre-rift reconstruction, 52–54  
 Pressure axis, 291, 294  
 Pressure driven flow, 17, 345–348, 377, 382  
 Pressure force, 339  
 Primary fault plane, 293  
 Primary NRM, 100, 183  
 Principal axes of stress, 230, 246  
 Principal strain axes, 235  
 Profile obliquity angle, 158, 159, 161  
 Prograde travel-time curves, 268, 270  
 Proton precession magnetometer, 84, 145, 146  
 Pseudo-single-domain (PSD) grains, 96  
 Pure shear, 44, 234, 235, 238, 245, 302  
 P-waves, 6, 11–13, 15, 249–252, 258, 271, 273–276, 290, 291, 294  
 Pyrolite model, 20, 21

## Q

Quadrupole components, 218–220  
 Quality factor, 192  
 Quantum numbers, 98  
 Quiet daily variations, 122

## R

Radiation pattern, 286, 290–294  
 Radiogenic decay, 319  
 Rake, 285, 293, 294, 296, 302, 304  
 Range, 12, 17, 44, 50, 54, 106, 115, 122, 124, 134, 144, 158, 172, 178, 184, 185, 189, 201, 243, 252, 257, 265–268, 270–272, 276, 277, 282, 295, 301, 306, 312, 313, 330, 369, 375  
 Rankine half-body, 352, 353  
 Rare-earth elements (REE), 11  
 Rate-and-state friction law, 280  
 Rayleigh–Bénard convection, 17, 354–361, 377  
 Rayleigh–Bénard rolls, 354  
 Rayleigh number, 22, 360  
 Rayleigh waves, 274, 275  
 Ray parameter, 261, 262, 264–268, 270  
 Reconstruction algorithm, 58, 61–64  
 Reduction of raw gravity, 372  
 REE. *See* Rare-earth elements (REE)  
 Reid's theory. *See* Elastic rebound theory  
 Remnant magnetization, 93, 94, 96, 100, 123, 143, 144, 155, 157, 184, 189, 220  
 Representation formula, 130, 173  
 Residual asthenospheric column, 11, 15

Resistive forces, 206, 334, 335  
 Retrograde travel time curve, 268  
 Rheological function, 237  
 Rheological profiles, 243, 244  
 Rheology, 237–244, 317, 339  
 Ridge jump(s), 38–40, 161–163, 165, 175  
 Ridge propagation, 39  
 Ridge push, 206, 332–335  
 Ridge segment reorientations, 38  
 Ridge segments, 37–39, 160, 163–166, 168  
 Rift zones, 12, 44, 52  
 Rigidity modulus. *See* Shear modulus  
 Rigid rotation, 30–34, 233, 234  
 Ring current, 122  
 Ringwoodite, 19–22, 333  
 Rock units, 183, 185–190, 192  
 Rodrigues' formula, 132  
 Rotation model, 54–59, 66, 68, 195, 199, 202, 393  
 Rotation parameter, 168, 368  
 Rotation tree, 58, 206

## S

Saint–Venant body, 242, 280  
 Sampling, 74, 146, 147, 168, 182–185, 196, 199, 217  
 Saturation magnetization, 90, 93–95, 177, 179  
 Scalar magnetic potential, 86, 130  
 Scalar magnetometer, 84, 144–146  
 Scalar potential, 287  
 Scalar seismic moment, 294, 296, 302, 304  
 Scale heights, 355, 358  
 Scattering, 23, 25, 185, 271, 328  
 Schmidt quasi-normalization, 134  
 SD grains. *See* Single-domain (SD) grains  
 Seaward dipping reflectors (SDR), 13  
 Second invariant of the strain rate, 305  
 Second viscosity, 339, 342  
 Sectoral harmonics, 135, 137  
 Secular cooling, 317, 319, 320  
 Secular strain rate, 280  
 Secular variation, 35, 112–115, 122, 139, 143, 147, 182, 185, 186  
 Sedimentary inclination error, 220, 221  
 Seismic cycle, 279, 280, 306  
 Seismic phases, 272–278  
 Seismic ray, 247, 257–278, 294, 390  
   equation, 258, 259  
   method, 247  
 Seismic refraction, 13, 52, 53, 264  
 Seismic tomography, 18, 24, 25, 257, 273, 278, 390  
 Seismic wave equation, 247  
 Seismogram filtering, 272  
 Self-exciting dynamo, 110  
 Self-inductance, 111  
 Serpentinization, 19, 126  
 Serpentinized mantle, 14  
 Shadow zone, 270, 271, 276  
 Shallower inclinations, 219, 221  
 Shape anisotropy, 178, 179  
 Shear modulus, 245, 305

- Shear stress, 228, 230, 237, 242, 243, 282, 295, 296, 331, 335, 336, 342  
 Short asynchronous pulses, 207  
 Similarity variable, 325, 326  
 Simple Bouguer anomaly, 373, 374  
 Simple Bouguer correction, 373–375  
 Simple shear, 44, 45, 234, 235, 302, 306  
 Single-domain (SD) grains, 95, 177, 178, 180, 181, 184  
 Sink flow, 350, 351  
 Slab, 16, 19, 21–23, 25, 29, 30, 42, 59, 72–75, 206, 216, 242, 278, 293, 301, 302, 306–314, 321, 332–335, 359, 370, 373–375  
     buoyancy, 333  
     pull, 72–74, 332–335  
     underthrusting, 307  
 Slide–hold–slide experiment, 282, 283  
 Sliding window method, 194  
 Slip weakening, 280  
 Slowness vector, 259  
 Small circle arc, 32, 40, 41, 54, 55, 57, 64, 68, 169, 202–205, 307–309  
 Smoothed APW paths, 195, 197, 199, 206  
 Smoothing parameter, 195–198  
 Snell's law, 262–263, 265–267, 269, 270, 294  
 Solar cycle, 122  
 Solar flare, 121, 122  
 Solar quiet (Sq) variations, 122, 143  
 Solar wind, 118–121  
 Solenoid, 109–112, 120, 126, 145, 146, 287, 338, 356  
 Somigliana tensor, 287  
 Source mechanism, 285, 293  
 South Atlantic anomaly, 140  
 Specific entropy, 344  
 Specific heat, 8, 320, 344  
 Specific volume, 21, 330, 343, 344  
 Spherical harmonic expansion, 130–140, 218, 363–365, 367, 372  
 Spherical polygons, 50, 67, 74  
 Spontaneous magnetization, 92, 94, 99, 100  
 Spreading asymmetry, 38, 39, 163  
 Spreading centers, 7, 10, 11, 14, 37–39, 41, 42, 124, 126, 159, 164–166, 334  
 Spreading obliquity, 38, 158  
 Spring–slider analogy, 284  
 Stage poles, 55, 56, 59, 62, 64, 67, 68, 160, 165, 206–208, 214, 215  
 Stage rotation, 55–58, 62, 63, 67, 205, 206  
 Stagnation points, 349, 352–353  
 Stagnation streamline, 353  
 Standard solid, 241  
 States of matter, 237  
 Statistical tests, 163  
 Stick–slip model, 280  
 Stokes' coefficients, 364–365, 367  
 Stokes' condition, 340  
 Stokes' law, 359  
 Stokes solution, 290  
 Stokes' theorem, 108, 127, 334  
 Strain–energy function, 253, 254  
 Strain localization, 242, 243  
 Strain rate, 238, 240–243, 280, 301–305, 311–314, 337, 339, 340, 342  
 Strain tensor, 233–235, 246, 254, 302–304  
 Stream function, 348–354  
 Streamline, 42, 119, 120, 348–351, 353, 354, 359  
 Stress  
     drop, 294–296  
     invariants, 232, 234, 305  
     relaxation, 239  
     tensor, 227–232, 246, 253, 254, 339, 342–344  
 Stretching factor, 44, 52, 53  
 Strike, 37, 68, 70, 119, 155, 156, 158, 160–162, 166, 182, 231, 263, 269, 285, 293, 296, 302–304, 306  
 Strike–slip faults, 36, 42, 45, 49, 64, 293  
 Sub–chron, 118, 161  
 Subduction roll–back, 42  
 Subduction zones, 6, 7, 11, 17, 19, 40–42, 46, 70, 75, 305, 307–311, 335, 353, 370, 377  
 Sublattices, 99  
 Suction, 334, 335  
 Superparamagnetic condition, 184  
 Superparamagnetic grains, 184  
 Superplume, 23–25  
 Surface forces, 227–229, 235, 236, 246, 253, 296, 339, 342, 343  
 Surface spherical harmonic, 134, 364  
 Surface waves, 252, 272–274, 347  
 Survey lines, 147–149, 158  
 S waves, 5, 15, 18, 25, 249–251, 255, 258, 270–276, 290, 291, 297  
 Syn–rift stage, 53  
 Synthetic seismogram, 247  
  
**T**  
 Take–off angle, 265, 268, 269, 278, 294  
 TBL. *See* Thermal boundary layer (TBL)  
 Tectonic elements, 49–54, 57–61, 64, 67, 68, 165, 199, 387  
 Tectonic plates, 4, 12, 14, 17, 23, 30–34, 37, 41, 46–48, 54, 58, 59, 62–66, 70, 72, 73, 159, 199, 202, 206, 207, 227, 280, 308, 311, 333, 334, 346, 347  
 Tectonic stages, 9, 45  
 Teleseismic distance, 275, 276  
 Temperature distribution, 17, 320, 323, 325, 327, 347  
 Tension axis, 291, 294  
 Terminal velocity, 359  
 Terrain correction, 374, 375  
 Tesseral harmonics, 135, 137  
 Thermal  
     activation energy, 181–182, 243  
     anomaly, 12, 17, 21, 22, 333, 359, 360  
     barrier, 22  
     boundary layer, 15, 16, 25, 323, 326, 327  
     buoyancy, 22, 23  
     convection (*see* Rayleigh–Bénard convection)  
     demagnetization, 184

diffusion distance, 325  
 diffusivity, 8, 320, 326, 347  
 isostasy, 331  
 structure, 317, 341  
 Thermal boundary layer (TBL), 15, 16, 25, 323, 326, 327  
 Thermoremanent magnetization (TRM), 100, 116, 123, 124, 177–185  
 Thinning of passive margins, 53  
 Tie lines, 147, 148, 158  
 Time-averaged geomagnetic field, 114, 115, 135, 159, 190, 217, 218  
 Time-dependent method, 340  
 Titanohematites, 94, 95, 99, 100  
 Titanomagnetites, 124, 184  
 Topological theorem, 63–69  
 Torque, 30, 42, 72–75, 83, 84, 99, 146, 206, 227, 229, 292, 333–335  
 Total energy density, 253, 255  
 Total field, 91, 123, 124, 138, 139, 144, 146, 147, 149, 156, 157, 161  
 Total field magnetic anomaly, 144, 149  
 Total reconstruction matrix, 61–63, 200  
 Total reconstruction pole, 62  
 Traction, 72, 207–212, 214, 215, 228–230, 235, 253, 334, 342  
 Transcurrent boundaries, 43, 44  
 Transform faults, 19, 37, 38, 42, 55, 63, 64, 69, 70, 168  
 Transition zone, 3, 4, 7, 12, 18–23, 41, 72, 73, 120, 239, 243, 250, 306, 307, 314, 333, 334, 344, 346, 353  
 Travel time, 258–260, 266, 268–271, 276, 277  
   curve, 266, 268–271, 276, 277  
   tables, 276  
 Trench retreat, 42  
 Trench zones, 40–42, 308, 311, 377  
 Triple junctions, 45–49, 56, 64, 65, 70, 72  
 Triplication, 268–270, 272  
 TRM. *See* Thermoremanent magnetization (TRM)  
 True polar wander, 203–204  
 Turning point, 264–268, 270, 273, 275  
 Two-dimensional flow, 348–354

## U

Ultramafic rocks, 15  
 Uniaxial anisotropy, 178  
 Uniform flow, 350, 351  
 Upper continental crust, 4, 6  
 Upper mantle  
   peridotites, 15  
   viscosity, 17, 347  
 Upward continuation, 170, 172–175  
 Urey ratio, 319, 320

## V

Vacuum permittivity, 104  
 Vector potential, 85, 287, 290  
 Velocity of upwelling, 10  
 Virtual geomagnetic pole (VGP), 191, 192  
 Viscoelastic materials, 238, 239, 241  
 Viscous behavior, 238, 241, 243  
 Viscous dissipation, 342, 343, 346, 347, 358  
 Viscous force, 339  
 Volcanic OCT, 13, 52  
 Volcanic passive margins, 12, 13  
 Volume elements, 8, 9, 29–31, 81, 88, 120, 227, 228, 234–236, 253, 261, 339, 342, 373  
 Vortex flow, 351  
 Vorticity, 350

## W

Wadati–Benioff seismicity, 312  
 Wadati–Benioff zone, 306–309, 311–314  
 Wadsleyite, 19–22, 333  
 Water, 10, 14, 18–20, 90, 124, 145, 243, 272, 316, 326, 327, 331, 374, 376  
 Water flow, 18, 19  
 Wave equation, 247–249, 252, 257, 258, 264, 290, 325  
 Wavefront, 15, 248, 257–264, 268, 269, 275  
 Wavelength, 124, 135, 136, 138, 146, 148, 171, 172, 252, 254, 255, 360, 370, 372, 377  
 Wavenumber, 171, 172, 175, 252, 275  
 Wernicke's model, 44  
 Westward drift, 112  
 WGS84. *See* World Geodetic System (WGS84)  
 Whole Earth phases, 275  
 Whole-mantle convection, 22, 23, 25  
 Wide rift, 44  
 Wilson cycle, 7, 19, 45  
 Wood–Anderson seismograph, 297  
 World Digital Magnetic Anomaly Map, 124, 125  
 World Geodetic System (WGS84), 368, 369

## X

Xenoliths, 16

## Y

Yield-strength envelope, 243  
 Yield stress, 242–244  
 Young modulus, 246

## Z

Zero-order levelling, 148  
 Zonal harmonics, 135, 137, 367





