Introduction

In this introduction we try to give some idea of the motivation and content of the course of lectures on which this book was based. Most of the points mentioned will be discussed in the text, but some of them are referred to merely in order to indicate possible extensions.

1. At the origin of the word geometry is the word *measure.* However, even when we are aiming for an explicit numerical description, the study of *form* necessarily has to precede the detailed construction and hence the measurement.

As an example, take the simple problem of drawing the *apparent outline* of a surface S (given by an equation $f(x, y, z) = 0$) with respect to projection into the xy plane. This is the curve C which is the projection into the xy plane of the curve D consisting of points (x, y, z) of S where the tangent plane is vertical, and therefore given by the equations

$$f(x, y, z) = 0, \quad f'_z(x, y, z) = 0.$$

The naïve approach to this question is to trace out the curve point by point after dividing up the xy plane into small squares. In doing this we may obtain a very inaccurate result for the following reason: tracing the curve C point by point assumes that it is regular, whereas – and this is a fundamental observation – this curve almost certainly has cusp points (we shall see why later). Therefore we need to start differently, by studying the form of this curve and specifically determining its cusp points and their tangents, before we complete the picture by drawing (point by point) the regular branches that connect them.

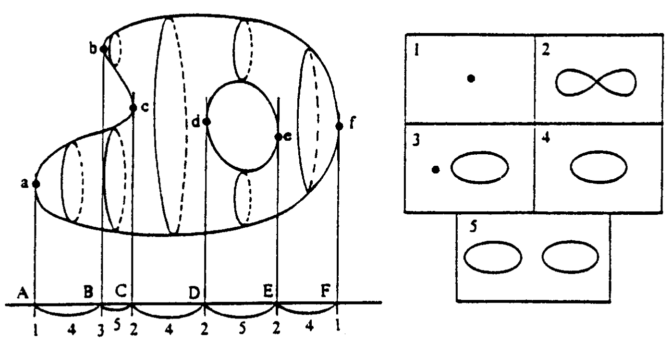
If we now want to find out what happens when we deform the surface S (by simply changing the chosen projection, for example) we must again start by investigating how the cusp points move. In fact in the general case they move regularly as long as two of them do not meet; after such a collision there is a 'change of form', a 'change of state' as a physicist would say, or, as Thom[1] says, a *catastrophe.*

[1] René Thom, French mathematician, born 1923, Fields Medal 1958, gave the name *catastrophe theory* to a collection of definitions and theorems that are part of the theory of singularities of differentiable maps. This term has had an astonishing treatment in the media (see Sect. 5.12).

2. We meet phenomena like this in the related context of *Morse*[2] *theory*. Here is a simple example.

Again in three-dimensional space, consider a (compact) surface S, given as above by an equation $f(x, y, z) = 0$, and imagine trying to discover the form of S by studying its sections by planes $z = a$. In this way we would obtain S as a union of 'curves' $C(a)$, where each $C(a)$ is defined in the plane $z = a$ by the equation $f(x, y, a) = 0$. The curves $C(a)$ are regular at all points (x, y, a) at which $f'_x(x, y, a)$ and $f'_y(x, y, a)$ do not vanish simultaneously: in fact this is a consequence of the *Implicit Function Theorem*.

In general (here, as previously, the justification for this 'in general' appeals to what is called the Transversality Theorem) the 'bad' points are finite in number and are the *critical points*; there must be some of these, if only the maxima and minima of the function z on S. The corresponding values of a are the *critical values*. Among the essential points of Morse theory is that, roughly speaking, on the one hand "nothing happens between the critical values" as significant changes in form of the curves $C(a)$ take place only for critical values (and, for those values, at the critical points), while on the other hand the global form of the surface S can be reconstituted starting from knowledge of the critical points alone (see Figs. 0.1a,b). Here also, if we deform the surface S then the critical points and critical values move continuously as long as two critical points do not collide.



(a) (b)

Fig. 0.1. (a) Critical points and critical values, (b) the five forms of level curves

3. In these two examples we see a general philosophy emerging. To explain this we use the first example, which is somewhat richer. The object is to study the projection map $(x, y, z) \mapsto (x, y)$ from the surface S to the plane P with equation $z = 0$, going from the general to the particular.

[2] Marston MORSE (1892-1977), American mathematician, originator (together with his compatriot Hassler WHITNEY who will feature significantly in this text) of a large number of the ideas that we shall be encountering.

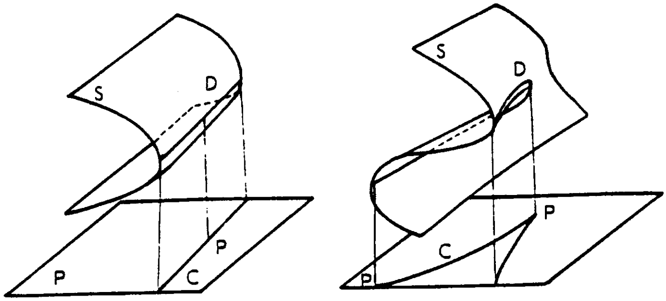
a) Above a general point p of P there is a finite number (possibly zero) of points of S, all with non-vertical tangent plane; as the point p varies each of its inverse images moves continuously (this is once again the Implicit Function Theorem); this 'general' part of the surface can be described as a union of *leaves* that can each be parametrized in the form $z = g(x, y)$.

b) When the point p reaches the apparent outline C at a general point of C the simplest of the catastrophes occurs, namely the *fold* (Fig. 0.2a): two inverse images of p coalesce at a point of S where the tangent plane is vertical, that is at a point of the curve D.

c) When the point p is even more special, that is when it is the projection of a point (x, y, z) of D with vertical tangent and which therefore, as can easily be checked, satisfies the equations

$$f(x,y,z) = 0, \quad f'_z(x,y,z) = 0, \quad f''_{zz}(x,y,z) = 0,$$

then three inverse images of p come together and p is a cusp point of C. This is the second catastrophe in order of complexity, the *cusp* (Fig. 0.2b).



(a) (b)

Fig. 0.2. (a) Fold, (b) cusp

d) If we have one more dimension, for example if S depends on time t and so is in fact given by $f(x, y, z, t) = 0$, we obtain the third catastrophe – called the *swallowtail* – when, for a particular value of t, two cusp points coalesce, with four inverse images.

These examples exhibit the fundamental characteristics of *singularities*; they are in general unavoidable, they are stable and above all they are 'structural': it is they that 'carry' the form of the geometric objects. This explains why differential geometry has evolved historically from the study of regular situations to the study of singularities.

4. In a general way which needs to be made precise in each instance, a geometric object decomposes into *strata*, the situation in the interior of each stratum being regular (that is to say technically validated by the Implicit Function Theorem), and the passage from one stratum to another taking place via one of the elementary catastrophes listed by Thom.

Moreover, when the given object is sufficiently general in its class this decomposition into strata and the catastrophes conforms to the usual dimensional intuitions: each new condition introduced translates into a drop in dimension by 1 of the corresponding stratum (surface, curve, point). This is a more difficult result, based on the Transversality Theorem. We can give an idea of what is involved using the previous example. While it is fairly clear that for general functions f, g and h the three equations $f = 0$, $g = 0$, $h = 0$ define isolated points, it is not immediately obvious why this should also be the case for the equations $f = 0$, $f'_z = 0$, $f''_{zz} = 0$. This is just the type of situation that the said theorem deals with.

A third step is the proof of the stability of the whole analysis under small perturbations. This is even more delicate, and we shall return to it later.

It is one of the central tenets of Thom's philosophy that these phenomena are a part of our everyday observation. It is the case, for example, with the luminous caustics that we can see each morning in our cup of tea: we immediately notice that first of all there have to be cusp points and that secondly their general form is essentially independent of the experimental conditions.

5. Furthermore, the very notion of *stability* is fundamental and it conditions every instance of mathematical modelling. Since the adjective 'stable' is used classically in many contexts, not always compatible with each other, from now on we shall say *structurally stable* to refer to properties not of a particular configuration but of the system as a whole. Thus we shall speak of stable equilibrium or stable orbit, but of a structurally stable differential equation. This notion will be made more precise at the appropriate time.

Now we give an example that uses a theory so simple as to be indisputable, namely that of a spherical mirror. From a certain point of view this may nevertheless seem a little suspect because spherical mirrors do not exist; anything that we can or could ever manufacture would be only approximately spherical. This implies that all that could ever be observed would be the (structurally) stable properties of spherical mirrors, that is to say those properties that remain approximately valid for approximately spherical mirrors. It is curious to note that an argument of this kind – which today seems imposed from the outset on anyone wishing to formulate any physical law whatsoever – leads directly to the present-day mathematical definition of continuity, whereas the historical development of this notion was long and difficult[3]. We remark in

[3] This definition (due in fact to BOLZANO) is named after Cauchy. Augustin CAUCHY (1789-1857) established several fundamental notions of classical analysis in his

passing, without elaborating the point, that the feature of being 'evident *a posteriori*' is common to many scientific concepts.

In parentheses we give, following Arnol'd[4], an amusing illustration of this 'philosophy'. Consider an experiment that consists of releasing a disturbance at the centre of a circular bowl, for example by letting a drop fall into the centre of a cup of tea. Theory says that the circular waves will propagate out from the point of impact, reflect on the boundary, and then reconverge. And this is indeed what is observed, although it is clearly impossible not only to obtain a circular cup but also to hit the centre exactly. In fact, if we consider a circle as an ellipse with coincident foci and then slightly separate the foci we observe that the deformation of the circumference is of second order relative to the separation of the foci, so that conversely a circle can be regarded as an ellipse whose foci are a pair of points, very close to each other and positioned symmetrically about the centre of the circle, but chosen arbitrarily according to the demands of the problem. Therefore what happens in reality is that the drop falls a little to one side of the (alleged) centre of the (alleged) circle and, as must happen in a genuinely elliptic mirror, the waves converge at the other focus, namely the symmetric point with respect to the centre.

6. We continue with these observations on structural stability, but now we consider the opposite case in which the situation that we aim to study really is not structurally stable, the neighbouring situations sharing common properties which are different from those of the initial situation. These are therefore the properties that will in fact be observed.

This remark enables us to understand the phenomenon called *symmetry-breaking* which occurs when the initial situation has symmetry properties which are not preserved under perturbation. Here is a simple example, to which we shall return in Sect. 5.9. It concerns finding the equilibrium states of a bar constrained to move in a vertical plane, freely jointed at its foot and kept in a vertical position by two symmetrical springs (see Fig. 0.3a). Calculations reveal what is called a *bifurcation*. When the spring compression is sufficiently strong the vertical position is the only equilibrium state and it is stable. On the other hand, when the compression is weak the vertical position is unstable and two stable equilibrium states appear symmetrically placed with respect to the vertical. This passage from one regime to another (the 'bifurcation') happens for a given parameter value. Naturally, this is not what is observed (see Fig. 0.3b). In fact the condition of exact symmetry of the two springs cannot be satisfied in practice because *it is not structurally stable.*

---

course (1821) at the Ecole Polytechnique. For further details see [HM], Vol.1, pp.336-345, or [MA], Chapt.2.

4 Vladimir ARNOL'D, Russian mathematician born in 1937, published many works, all fascinating, on the theory of singularities and its applications, notably in mechanics. Their influence on this text is undeniable; they are recommended reading, although it is necessary to be aware of some ideosyncracies in terminology or orthography (particularly in proper names) and the unconstructively polemical (and sometimes seemingly abusive) nature of certain claims of priority.

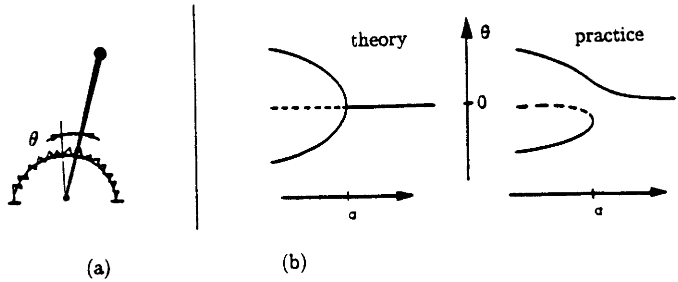


Fig. 0.3. (a) Bar supported by symmetric springs, (b) broken symmetry in the bifurcation

7. However, this in turn opens up another argument: how on earth can we effectively distinguish between harmless simplifying assumptions and those which make the model not structurally stable and therefore quite unrealistic for reasons that we have just seen?

To explain how this question can be answered we have to start by specifying the context a little more accurately. Consider a system (mechanical, physical, chemical, ...) described by a certain number of *state variables* satisfying certain characteristic relations (algebraic or differential, for example) in which some *control variables*[5] are involved in a known way (such as the spring control in the previous example), while an unspecified number, quite probably infinite, of *hidden parameters* are involved in unknown ways which in particular reflect the 'imperfections' of the concrete physical object that is modelled by our abstract system. In other words, all that we know is that our system is near, and as close as we wish (or rather as we are prepared to pay for), to the theoretical system described by our state and control variables and by the relations that connect them. Now, if this theoretical situation is structurally stable, the real system will have behaviour that is close (and even as close as we wish etc. ...) to the theoretical behaviour. On the other hand, if the theoretical system is too special to be structurally stable (supposing that our two springs are exactly symmetric) then the hidden parameters come into play in unpredictable ways; in that case it usually turns out to be enough to incorporate new control variables (so that instead of considering the difference between our two springs as a hidden variable, we take account of it) in order to recover a structurally stable system.

Consequently the structurally unstable symmetric system with one control variable

$$\ddot{\theta} = \sin\theta - a\theta,$$

which (in suitable units) models the bar held by springs that are assumed to be perfectly symmetric, has to be replaced by the asymmetric system with

[5] One contemporary French philosopher uses the excellent terminology *dynamic variables* and *strategic variables*.

two control variables

$$\ddot{\theta} = \sin\theta - a\theta + b,$$

which can be shown to be structurally stable and which therefore correctly takes account of the real situation, whatever (small) hidden perturbation it may be subjected to.

8. In the same spirit, we can present the problem of *linearization* of a differential system as follows. Consider the system

$$(S) \qquad \frac{dx}{dt} = f(x,y), \quad \frac{dy}{dt} = g(x,y),$$

where the vector field

$$X : (x,y) \mapsto (f(x,y), g(x,y))$$

vanishes at the origin, and also the linearized system

$$(L) \qquad \frac{dx}{dt} = ax + by, \quad \frac{dy}{dt} = cx + dy,$$

with

$$f(x,y) = ax + by + \cdots, \quad g(x,y) = cx + dy + \cdots.$$

We wish to compare the trajectories of the system (S) in a neighbourhood of 0 with those of the linear system (L). We see the resemblence to the earlier discussion: with (S) being close to (L), and the nearer to the origin the closer it is to (L) (by Taylor's formula), can we deduce that the trajectories of (S) are close to those of (L)? We are dealing with a 'structural stability' property of (L), and the answer is given by the Hartman-Grobman Theorem (in the special case of dimension 2 as here, a slightly stronger theorem applies[6]): if the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has no purely imaginary eigenvalue then in a neighbourhood of the origin the system (S), which is a perturbation of (L), has its trajectories close to those of the latter.

On the other hand, in the case when these eigenvalues are purely imaginary the trajectories of (L) are circles centred at 0, while those of (S) may have very different global form however close (S) is to (L). In fact in this case (L) is already structurally unstable within linear systems (see Fig. 0.4).

9. These reflections naturally lead us to hope that at the end of the day structural stability will be a common occurrence. To be more precise, in order for the philosophy sketched in Para. 7 above to work effectively we would want it to be the case that structurally stable sytems could be found in the neighbourhood of any system under consideration. In mathematical terms, the structurally stable systems ought to be *dense* in the set of systems, or, to put it another way, a system chosen at random (a so-called 'generic' system)

---

[6] For these two theorems see Sects. 8.7 and 8.11.

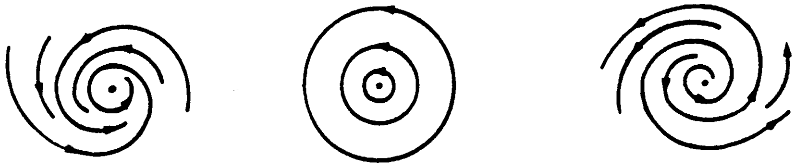


Fig. 0.4. Example of a structurally unstable linear system

ought to be structurally stable. This is what is called somewhat pompously the 'structural stability hypothesis'. We shall see later what we should think of it.

Meanwhile, there is something that obviously has to be made precise in all that we have been saying so far: what do we understand technically by 'nearby properties', 'situation that does not change', 'dense', 'generic', etc.? Take the example of Hartman's Theorem mentioned above. It can be stated as follows: if the linear system (L) has no purely imaginary eigenvalue (in short, if it is 'hyperbolic') then for every system (S) with linearization (L) there is a (fairly regular) transformation of the xy plane in a neighbourhood of 0 which takes the trajectories of (L) to those of (S) preserving the parametrization by time.

Clearly the only content of a theorem of this kind is the assertion of regularity of the transformation: the more regularity we assert, the more significant the theorem becomes and the more worthwhile the operation of linearization will be in practice, but of course the lower the likelihood of the theorem being true. In the precise case of linearization the answer is a little subtle: without additional hypotheses we cannot obtain very much regularity and we can not even ensure differentiability of the transformation, but only a somewhat weaker property (except as already indicated in dimension 2; but even in this case the existence of second derivatives is not guaranteed). On the other hand, with an additional assumption on the eigenvalues (absence of 'resonances'[7], which is a 'generic' assumption) we can assert the infinite differentiability of the transformation: this is Sternberg's Theorem.

10. In these few rather informal paragraphs we have met the essential key-words of the subject: *transversality, genericity, structural stability, linearization, bifurcations, catastrophes, ...* There are at least two missing. The first is *dissipative*; in fact we shall be saying nothing serious about non-dissipative or *conservative* systems whose study is more complicated and involves tools of infinite subtlety. The absence of the other word is more curious. Is this really a text on *geometry*? The question raises another: what is geometry? It is not easy to answer this, and the answers change over time. Nevertheless geometric objects exist, and the things we have been briefly talking about above certainly belong to the universe of geometry. Moreover, it is clear that

[7] See Sects. 8.10 and 8.11.

differentiable objects and their singularities are omnipresent in nature and also in the theoretical models that have been developed to take account of them. This is much less clear for the students of today than for their elders. Teaching geometry at university level nowadays requires picking up the threads of a broken tradition. In particular, the absence of prior knowledge in certain areas (particularly in algebra) and the general state of teaching in schools rule out whole sections of geometry (algebraic geometry, finite geometries, ...). In my view the themes developed here have the advantage of being relevant to both the content and outlook of other texts as well as to current movements in contemporary ideas (in bifurcation, turbulence and so on). The later chapters will lead us to living questions on which research is active and fruitful and which would have seemed totally far-fetched just a few years ago.

In fact a title more in line with the content of this book could have been: 'geometric methods in the study of singularities and bifurcations', or (why not ?) 'from Poincaré to Smale and beyond' ....

11. However, even with the best intentions in the world, it is not possible to do mathematics without technique. The essential characteristic of true statements is that they possess a proof. We would like to be able to delegate to specialists alone the task of checking the existence and validity of proofs. To a large extent this is possible, but there are of course several drawbacks. First of all (and this is the case for the easiest ones) the most economical way to understand the meaning and the limitations of an assertion is often to read a proof of it. But there are more important considerations: in order to be able to apply the conclusions of a theorem it is necessary for the hypotheses to be satisfied (although it would seem that this obvious fact has not penetrated the whole of science, especially the 'inexact'[8] sciences). It is quite common, when confronted with a concrete problem, to find a theorem which 'almost applies'. This is why in order to master a proof it is usually necessary to have a clear idea of the status of the hypotheses that it contains: are they there for circumstantial reasons (convenience of exposition, simplification of the proof, sympathy for the reader ...) or because they are essential? Often a small modification of the premisses entails only a minor change in the consequences (could we then talk about 'stable' or 'generic' results? Unfortunately, a result picked at random has little chance of being interesting! ...); sometimes, on the other hand (could we call these 'critical' results?) a slight weakening of the hypotheses turns into complete collapse of the conclusions. The role of *counterexamples* is precisely to set boundaries on the possible, and it is not due to perversity (or at least not totally) that mathematical texts tend to exhibit monsters[9].

[8] So called in opposition to the 'inhumane' sciences .... On the abuses of pseudo-mathematical modelling in the social sciences, see [BE].

[9] Hermite said, no doubt with a touch of humour, that he "turned away in fear and horror from this lamentable plague of continuous functions that have no

As far as the theorems in this text are concerned, they come in three forms. Some of them, usually the most elementary ones, are accompanied by a complete proof. For others, just the main ideas of the proof are indicated. Finally, for the last type which are too difficult or which appeal to notions too far afield, no indication of the proof is given. Note incidentally that the sign □ marks the end of a proof (or its absence when following directly after the statement).

12. Prerequisites. This book assumes a basic knowledge of linear algebra, general topology and functions of several variables. To help the reader, some of the most important results needed are recalled in the text. Apart from the notational items listed below, these results are placed where they arise most naturally rather than being lumped together in a special section that would inevitably be indigestible. This is the reason why it sometimes happens that we use an auxiliary notion that is not taken up again until later; in that case use of the Index should enable the reader to find the corresponding reference easily.

In places we use small type to give commentaries or include additional observations that may help some readers to make connections between the text and knowledge they may already have from elsewhere. For example, a few variants of terminology are noted in this way. Those who find that these inclusions add to the difficulty can ignore them, at least on the 'first reading'. In the footnotes there is some brief biographical information about mathematicians whose work is quoted. The last section of each chapter aims to give a slightly more global historical overview; this may also contain some biographical pointers. Symmetrically, in the first introductory section of each chapter there is an attempt to motivate or justify the subjects treated in the chapter and the approach chosen. Other comments of this nature can also be found scattered in the text.

13. References. Within each chapter the number of that chapter is to be understood. Thus to refer to Proposition 2.4 in Chapter 3 we shall say "by Proposition 2.4" or "by Proposition 3.2.4" according to whether we are inside Chapter 3 or not. Numbered formulae are referred to in the same way, with the difference that the number is placed in parentheses as in "formula (2.4)" or "formula (3.2.4)", or perhaps just (2.4) or (3.2.4).

Throughout the text the capital letters in square brackets (for example [HM]) refer to the bibliography. The works included vary greatly in length and level. Some of them such as [AS], [A3], [A6] are elementary introductions, while others such as [AA], [BL], [HS], [IR], [PS] are at a level comparable to the present text. The monographs [AM], [A5], [AG], [HA], [GH], [KH], [MV],

---

derivatives", but it was necessary to exhibit some of them if only to put an end to the string of false proofs, often underwritten by famous names, of the fact that every continuous function was differentiable.

[PM], [RO] give an idea of more advanced developments and in some cases the present state of research.

14. I have benefitted from the help of several colleagues in the preparation of this text. I especially thank Marc Chaperon as well as Jean-Pierre Bourguignon, Marc Giusti and Jean Lannes who have shared with me many valuable observations. Several errors in the earlier (French) edition were detected by students in the class of 1984, among whom I must give a special mention to Max Bezard.

http://www.springer.com/978-3-540-52118-1

Bifurcations and Catastrophes
Geometry of Solutions to Nonlinear Problems
Demazure, M.
2000, VIII, 304 p. 1 illus., Softcover
ISBN: 978-3-540-52118-1