Preface

For a number of years, we have been exploring the underpinnings of algorithms for simulating molecular dynamics from a mathematics perspective, i.e. as outsiders. We have learned the tricks of the trade through reading books written for practitioners, websites, documentation in software packages (yes, someone does really read that stuff!), Ph.D. theses (yes, someone does really read that stuff!), and a sometimes bewildering collection of research papers. As we have plumbed some of the deep recesses of the subject, and added our own contributions where we could, we have come to form a composite picture of the topic. Increasing numbers of scientists with an ever wider range of backgrounds are beginning to work within this area, therefore we feel that it is time to have a book that presents the topic at an introductory level and from a mathematical perspective.

Computational molecular dynamics is a field that has evolved in response to the needs of chemists, physicists and more recently biologists, materials engineers, drug designers, etc. The derivation of methods is often based on intuition or appeal to examples, but deeper insight into molecular dynamics methods can be obtained through a mathematical approach. Our book presents molecular dynamics methods from a mathematical perspective and describes numerical methods that form the basis for molecular simulation algorithms and, ultimately, software. Our treatment is far from comprehensive, in the sense that there are things that scientists investigate involving molecules and computers that we will say nothing about. We will largely avoid excessive detail in the formulation of models, and we will steer clear of some of the most recent developments in the field that are still unsettled. By placing the emphasis on foundations our intention is to provide a resource of lasting value for mathematicians and computational scientists, and a guide for those who must rely on molecular dynamics in their research, and for those who may need to develop new methods for complex applications.

The fact that molecular dynamics is sometimes seen as something of a black art is from our point of view the strongest motivation for us to take up the challenge of writing this book. We have sought to place each algorithm on a solid foundation, by introducing geometric integration and backward error analysis to explore the properties of methods for deterministic simulation, and then by

extending these concepts to provide foundations for studying Langevin dynamics integrators. Our presentation follows a simple pattern: give enough theory to describe the models and (deterministic or stochastic) differential equation systems, then discuss the design of numerical methods, and finally explain the performance of the methods using both analysis and numerical experiments. The approximation of averages was originally viewed as a by-product of computing trajectories of a Newtonian dynamical evolution. Our treatment emphasizes a recent change in perspective: numerical methods have their own effective global structures (perturbed Hamiltonians, invariant measures) and the design of methods can be enhanced by using this knowledge, thus providing better (more accurate, more robust) tools for thermodynamic calculation.

Our book emphasizes the construction of paths (deterministic or stochastic) for dynamics and for the computation of long-term properties for atomistic models. The numerical analysis for this has only just reached a state of maturity so that a book like this can be prepared. We aim to provide explicit and precise guidance on the properties of algorithms to aid in selection and implementation. A criticism that could be leveled at our book is that it is not a fully comprehensive treatment of modern work exploring the mathematical foundations of molecular modelling. Since the 1990s, researchers have begun to turn their attention to various topics arising in connection with molecular models, addressing a range of challenges including coarse-graining techniques, methods for computing free energies, metastable processes, rare event techniques, the relationship between quantum and classical models, parallel computing techniques, and the design of multiscale methods that interface molecular dynamics with solid mechanics models. While we make no effort to be complete in our coverage of the recent trends in the field of molecular modelling, the techniques described here are fundamental to most modern theories, and we therefore expect that they will play an important and enduring role in any broader frameworks of molecular simulation. (Below, we provide some references for advanced reading on timely themes.)

The best preparation for reading this book would consist of basic (undergraduate level) courses in elementary analysis, numerical analysis, differential equations and probability, and some knowledge of mechanics (Newton's laws, force balance, harmonic oscillators, the Kepler problem, etc.). The analysis of splitting methods by the Baker-Campbell-Hausdorff expansion can be explained in terms of the exponentials of sums of noncommuting matrices. Concepts such as generalized functions are introduced, but only the Dirac delta function needs really to be explained in this context. The discussion of stochastic differential equations and especially the Fokker-Planck equation (end of Chaps. 6, 7) necessarily gets a bit involved, but the essential results in Chap. 6 have been condensed into a few basic theorems which are stated without proof. For readers with limited exposure to analysis, one might stop in Sect. 6.4 after presenting the irreducibility of discrete Markov chains and then explain in general terms and by analogy the content of the remainder of the section, before continuing to the study of Langevin dynamics methods in Chap. 7. For advanced readers, it is recommended to consult a good book on stochastic processes; see references mentioned below (or other articles

referenced in Sect. 6.4). Results in Chaps. 7 and 8 are relatively new and are presented at a higher level, but still aimed at a broad audience. The book is bound to generate questions from students that cannot be answered entirely using methods and results presented in the book, but this is in many ways rather a strength than a weakness.

The assumption of limited background knowledge has some disadvantages—detailed and rigorous proofs are omitted, for example—but by the same token the book is hoped to be usable for interdisciplinary courses and for independent study by a broad audience. We put the emphasis here on making simple statements regarding algorithms. These statements could, in most cases, be made entirely rigorous with a more formal (but also much longer) presentation. In some cases, we impose more stringent assumptions of smoothness or boundedness than are actually needed in order to simplify the presentation. At the end of the day, it is not just theorems that define this subject. We are well aware that carefully crafted numerical experiments are often more informative than pages of analytical derivations and we therefore describe in detail quite a few simulations.

The appendices provide some notes on force evaluation, a brief review of some elements of basic probability theory that we rely on, and a brief discussion of Monte-Carlo methods.

A course in numerical methods for molecular dynamics could be taught entirely from this book. For a broader perspective on molecular simulation within a course setting, one should include readings on topics such as Monte-Carlo methods, and force calculation which could be found, for example, in [140, 328]. It would also be important to have some experience of using molecular simulation software to perform realistic calculations (see below).

Depending on the background of the reader, or to obtain a fuller picture of various issues, it may be useful to supplement our book with additional texts chosen from among the following:

- *for ordinary differential equations, and dynamical systems:* [51, 177, 216, 362, 386];
- *for classical mechanics:* [15, 16, 57, 151, 212, 249, 297];
- *for statistical mechanics:* [57, 70, 197, 264, 366, 372];
- *for molecular models and algorithms:* [7, 140, 159, 215, 307, 328, 372];
- *for numerical analysis, e.g. numerical ordinary differential equations, geometric integration and numerical methods for stochastic differential equations:* [164, 167, 227, 270];
- *for probability theory, stochastic processes and stochastic differential equations:* [18, 142, 269, 279, 298, 314].

Another excellent book we should mention here is that of Griebel et al. [155], which explores the computational aspects of molecular dynamics in the setting of high performance computing. This would provide an excellent complementary book for a course focussed more on the issues relevant to software, large scale simulation, and parallel computing.

For some readings taking in the modern span of molecular modelling, particularly the "multiscale" context, we propose articles on the string method [115–117, 384], the adaptive biasing force method [76, 91–93], metadynamics [23, 61, 207, 208], other methods of free energy calculation [83, 104, 231], Markov state modelling [282, 332], and atomistic-to-continuum coupling [243]. These articles can serve as starting points for advanced projects and help to put into practice ideas learned in the book. We also emphasize that an excellent follow-on book specialized to mathematical techniques in free energy calculation [233] is available and could help to guide study of this important topic.

Molecular Dynamics Software

It will be difficult to gain a full grasp of molecular dynamics without performing simulations. We have found that small systems (harmonic oscillators, double wells, Lennard-Jones trimers, etc.), suitably chosen, can often provide insight into the behavior of molecular dynamics methods. Experimentation with algorithm design using small toy models can in many cases be performed adequately using a system such as MATLAB.[1] To accompany the release of this book, we have prepared MATLAB code for many of the algorithms we present. These programs are released as the **MD.M** software package on the website linked from http://MolecularDynamics.info. This software (which interfaces to MATLAB as well as to the free alternative Octave[2]) allows easy experimentation with few-atom models (up to a dozen atoms, say). Using a C-language interface in combination with **MD.M** to compile the compute-intensive force field, the code can be accelerated to allow study of up to several hundred atoms.

Eventually, if one wishes to understand the more intricate nuances of numerical simulation, e.g., multiscale phenomena, or to treat larger systems with more realistic force fields, it will be necessary to have access to specialist software with which to try things out or for the purposes of modification. Fortunately, there are a handful of major "freeware" software packages in common use, each of which has a community of researchers and is maintained and developed by a substantial team with extensive funding by government agencies. We mention specifically the following: **AMBER** [393], **CHARMM** [246], **DL_Poly** [365], **GROMACS** [33, 236], **LAMMPS** [303], **NAMD** [302], **TINKER** (http://dasher.wustl.edu/ffe/). These codes are designed for different types of applications. For example, **NAMD**, **AMBER**, and **CHARMM** are very much oriented to biomolecular modelling, whereas **LAMMPS** and **DL_Poly** were originally designed more for materials applications. The **MOIL** code (http://clsb.ices.utexas.edu/web/moil.html) facilitates the study of rare events using molecular dynamics. The **DESMOND**

[1] http://www.mathworks.co.uk/products/matlab/.

[2] GNU Octave is available from http://www.gnu.org/software/octave/

system (http://www.deshawresearch.com/resources_desmond.html) is produced by a commercial company, but provided free for non-commercial use. We also mention the commercial packages "Materials Studio" and "Discovery Studio" which are produced by Accelrys, Inc. The open software projects freely distribute the source code, so it is possible for a keen user to alter the algorithms. Unfortunately, it has to be mentioned that there is something of a trade-off between the efficiency of the codes (especially in the high performance computing setting) and the ease with which algorithms may be changed. **NAMD** and **GROMACS** are, in our experience, a little complicated to modify, whereas **TINKER** and **DL_Poly** are relatively easy. If one were to restrict oneself to a single code system in which all algorithms are to be tried out, then we highlight that **LAMMPS** offers a very convenient mechanism for code modification, making it particularly "developer-friendly".

Acknowledgements

The authors would like to acknowledge some of the valuable materials obtained from colleagues and via the internet, in particular the miscellaneous unpublished lecture notes and lecture slides of Lawrence Evans (Berkeley), Jonathan Goodman (NYU), Martin Hairer (Warwick), Peter Olver (Minnesota), and Mark Tuckerman (NYU). In many cases, our reading of these has been inspirational.

We wish to thank the large number of people who have patiently assisted us in preparing this volume; our consultation was very broad, chosen to span researchers working in numerical analysis, in pure and applied analysis, in probability, and in theoretical chemistry, including Janis Bajars (Nottingham), Stephen Bond (Sandia), Ron Elber (Texas), Radek Erban (Oxford), Jason Frank (Utrecht), Alastair Gillespie (Edinburgh), Ben Goddard (Edinburgh), Martin Hairer (Warwick), Oliver Penrose (Heriot-Watt), Xiaocheng Shang (Edinburgh), Bob Skeel (Purdue), Gabriel Stoltz (Ecole des Ponts, Paris), Andrew Stuart (Warwick), and Mark Tuckerman (New York); their helpful contributions have been integrated into all chapters. Both Greg Pavliotis (Imperial College) and Michael Tretyakov (Nottingham), in addition to providing many helpful comments, were also generous in giving us advance access to their own books-in-progress on specific topics.

Edinburgh, UK — Benedict Leimkuhler
Chicago, Illinois, USA — Charles Matthews
October 2014

http://www.springer.com/978-3-319-16374-1

Molecular Dynamics
With Deterministic and Stochastic Numerical Methods
Leimkuhler, B.; Matthews, C.
2015, XXII, 443 p. 95 illus., 71 illus. in color., Hardcover
ISBN: 978-3-319-16374-1