Metastability and Markov State Models in Molecular Dynamics
Modeling, Analysis, Algorithmic Approaches
Metastability and Markov State Models in Molecular Dynamics
Metastability and Markov State Models in Molecular Dynamics

Modeling, Analysis, Algorithmic Approaches
# Contents

Introduction vii

Chapter 1. Transfer Operator Approach to Conformation Dynamics 1

Chapter 2. Dynamics 5
   2.1. Markov Processes and Transition Functions 5
   2.2. Markov Processes in Molecular Dynamics 7

Chapter 3. Metastability 15
   3.1. Hitting Times and Committors 15
   3.2. Exit Times and Exit Rates 21
   3.3. Metastable Full Partitions and Almost Invariance 23
   3.4. Definitions of Metastability 24

Chapter 4. Transfer Operators and Generators 25
   4.1. Transfer Operators in Molecular Dynamics 28
   4.2. Mean Hitting Times and Committors 30
   4.3. Dominant Eigenvalues 33
   4.4. Spectral Approach to Metastability 56
   4.5. Numerical Experiments 57

Chapter 5. Projected Transfer Operators 61
   5.1. Galerkin Projection 61
   5.2. Density Propagation Error 68
   5.3. Eigenvalue Error 78
   5.4. Committor Projection and Core Sets 82
   5.5. Numerical Experiments 91

Chapter 6. Transition Path Theory 99
   6.1. Reactive Trajectories 100
   6.2. Illustrative Examples 109

Chapter 7. Concluding Remarks 115

Appendix A. Some Mathematical Aspects of Transfer Operators 117

Appendix B. Definition of Exit Rates 121

Bibliography 123
Introduction

This is a book about mathematics. At the same time it is a book about molecular dynamics. Typical books with this combination of topics try to give a mathematical framework for what is done in standard molecular dynamics simulations or consider the temporal discretization schemes used in molecular dynamics. This book will not. Instead it will be concerned with one of the key challenges in molecular dynamics: How can one analyze molecular function if the required simulation timescales are infeasible?

For readers who are not familiar with molecular dynamics a few words explaining this key challenge seem appropriate: Applications in modern biotechnology and molecular medicine require simulation of biomolecular systems in atomic representation with immense length and timescales that are far beyond the capacity of computer power currently available. The reason for this is that many of the processes that constitute molecular function are rare event processes appearing on timescales that are many orders of magnitude, say 10–15 orders of magnitude, longer than the typical time steps of the numerical simulation. As a consequence, there is an increasing need for reduced models of the dynamical behavior of molecular systems that still allow describing the relevant dynamical properties while at the same time being less complex and thus allowing simulation on longer timescales. However, whether such reduced descriptions reproduce the correct rare event statistics is a question that in principle can only be answered mathematically.

In most molecular systems the biologically interesting and computationally problematic rare events belong to so-called conformation changes. Conformations are metastable sets of the dynamical behavior of the molecule, that is, regions of the molecule’s state space that are attractive for the dynamics in the sense that typical trajectories remain within such regions for long periods of time before exiting towards other metastable sets. This rough description may already be sufficient for indicating that any mathematical approach to efficient rare event simulation in molecular dynamics should address the issue of metastability of the dynamical behavior of molecular systems. Consequently, this book will be concerned with understanding the origin of metastability and with its mathematical description. This, however, is just the first step. We will see how to exploit the existence of metastable sets for constructing a reduced molecular dynamics model with good approximation properties on the long timescales. In this reduced model, the so-called Markov state model (MSM), the main metastable sets will form the states of a Markov chain or Markov jump process while the transition probabilities or transition rates of the Markov model are given by the transition statistics of the original
molecular dynamics process. Mathematically, the process of reducing the original molecular dynamics process to the MSM process will be a Galerkin discretization of the so-called transfer operator of the molecular dynamics process. Thus, the mathematics of projected transfer operators is central for our undertaking.

That is exactly what the book will focus on: metastability, MSMs, and the mathematics of transfer operators that allows the analysis of the relation between one and the other. The rough outline is as follows: We start with a more detailed introduction to the idea behind MSMs in Chapter 1. Next we will outline the different molecular dynamics models and the mathematical framework in which they all can be handled together (Chapter 2). Then we can turn to the mathematical concepts of metastability in Chapter 3 before we introduce transfer operators and the spectral approach to metastability (Chapter 4), and the construction of MSMs via projected transfer operators in Chapter 5. Last, we will investigate how to use an existing MSM to get information on the transition pathways that dominate the transitions from one metastable set into another (Chapter 6).

Recent years have seen an ever increasing publication activity on how to construct MSMs for very different molecular systems ranging from peptides to proteins, from RNA to DNA, and via molecular sensors to molecular aggregation (see the next section for references). Presently, several books are under preparation that will address these practical approaches in all algorithmic detail but not their mathematical background. Therefore, the present book concentrates on the latter.

The material presented in the following has resulted from more than 12 years of research on MSMs and metastability that started with the first publications by the first author and his coworkers in the last millennium; see [15, 16, 95, 96]. It has been compiled for mathematicians as well as practical computational scientists interested in modern molecular dynamics.

The book is predominantly addressed to graduate students and researchers wanting to get an overview of the theoretical background of MSMs and associated concepts in molecular dynamics. Therefore the mathematical technicalities have been limited to the level that the authors believe to be accessible for theory-interested nonmathematicians. The text includes many examples and helpful illustrations. Everyone interested in deeper insights into the matter will find comprehensive references to the relevant literature. The authors hope that the present book may help to bridge the still existing gap between mathematical research on molecular dynamics and its practical use for realistic molecular systems by providing readers with tools for performing more in-depth analysis of simulation and data-analysis methods in molecular dynamics.

Many extraordinarily talented people have contributed to the mathematics of metastable systems in relation to the transfer operator approach, its algorithmic realization, and its applications. We wish to acknowledge the efforts and excellent skill of the following mathematicians: Peter Deuflhard (who has contributed from the very beginnings in the last millennium till today), Eric Vanden-Eijnden, Marcus Weber, and Frank Noe. Their contributions have been invaluable. Furthermore, we want to thank the DFG Research Center Matheon “Mathematics for
key technologie” and the Freie Universität Berlin for making this work possible by providing funding and the creative atmosphere we treasure so highly.
Bibliography


Published Titles in This Series

24 Christof Schütte and Marco Sarich, Metastability and Markov State Models in Molecular Dynamics, 2013
23 Jerome K. Percus, Mathematical Methods in Immunology, 2011
22 Frank C. Hoppensteadt, Mathematical Methods for Analysis of a Complex Disease, 2011
21 Frank C. Hoppensteadt, Quasi-Static State Analysis of Differential, Difference, Integral, and Gradient Systems, 2010
20 Pierpaolo Esposito, Nassif Ghoussoub, and Yujin Guo, Mathematical Analysis of Partial Differential Equations Modeling Electrostatic MEMS, 2010
19 Stephen Childress, An Introduction to Theoretical Fluid Mechanics, 2009
18 Percy Deift and Dimitri Gioev, Random Matrix Theory: Invariant Ensembles and Universality, 2009
17 Ping Zhang, Wigner Measure and Semiclassical Limits of Nonlinear Schrödinger Equations, 2008
15 Emil Artin, Algebra with Galois Theory, 2007
14 Peter D. Lax, Hyperbolic Partial Differential Equations, 2006
13 Oliver Bühler, A Brief Introduction to Classical, Statistical, and Quantum Mechanics, 2006
12 Jürgen Moser and Eduard J. Zehnder, Notes on Dynamical Systems, 2005
11 V. S. Varadarajan, Supersymmetry for Mathematicians: An Introduction, 2004
10 Thierry Cazenave, Semilinear Schrödinger Equations, 2003
9 Andrew Majda, Introduction to PDEs and Waves for the Atmosphere and Ocean, 2003
8 Fedor Bogomolov and Tihomir Petrov, Algebraic Curves and One-Dimensional Fields, 2002
7 S. R. S. Varadhan, Probability Theory, 2001
6 Louis Nirenberg, Topics in Nonlinear Functional Analysis, 2001
5 Emmanuel Hebey, Nonlinear Analysis on Manifolds: Sobolev Spaces and Inequalities, 2000
2 Jalal Shatah and Michael Struwe, Geometric Wave Equations
Applications in modern biotechnology and molecular medicine often require simulation of biomolecular systems in atomic representation with immense length and timescales that are far beyond the capacity of computer power currently available. As a consequence, there is an increasing need for reduced models that describe the relevant dynamical properties while at the same time being less complex. In this book the authors exploiting the existence of metastable sets for constructing such a reduced molecular dynamics model, the so-called Markov state model (MSM), with good approximation properties on the long timescales.

With its many examples and illustrations, this book is addressed to graduate students, mathematicians, and practical computational scientists wanting an overview of the mathematical background for the ever-increasing research activity on how to construct MSMs for very different molecular systems ranging from peptides to proteins, from RNA to DNA, and via molecular sensors to molecular aggregation. This book bridges the gap between mathematical research on molecular dynamics and its practical use for realistic molecular systems by providing readers with tools for performing in-depth analysis of simulation and data-analysis methods.

For additional information and updates on this book, visit www.ams.org/bookpages/cln-24