Preface

The computation of free-energy differences is a very important and active research field in computational statistical physics. Being applied mathematicians, we were unable to find a textbook that suited our needs when we first approached numerical problems in computational statistical physics. The references we found were either theoretical statistical physics books, focusing on model problems treated analytically and perturbations thereof; or textbooks for practitioners, with recipes and comments on the physical models, but almost no mathematical analysis of the numerical techniques. We hope that the present book contributes to filling this gap and usefully complements some of the recent references concerning the long-time integration of Hamiltonian dynamics (such as [Hairer *et al.* (2006)]).

The audience we have in mind while writing these lines is composed both of mathematicians and scientists from the applied communities (physics, chemistry, biology, etc.), who use free-energy techniques as one tool among many to study the complex systems they are interested in. We conceive these notes as a self-contained presentation of what we believe are currently the most standard computational methods for free-energy computations. We hope that this presentation will be of interest to researchers in the field, while still being accessible for graduate students.

Free-energy computation is an opportunity for mathematicians to study many theoretical concepts and numerical strategies, such as techniques to sample multimodal probability measures, constrained stochastic dynamics to sample measures on submanifolds, adaptive importance sampling strategies, etc. In this book, we insist on the numerical analysis of the methods at hand, giving error estimates or rates of convergence.

Besides, for those interested in the applications, this book is an opportunity to learn more about the mathematical underpinning of the numerical techniques used on a daily basis for the computation of free-energy differences. We want to highlight the similarities between techniques presented as very different in nature in the current literature. We hope that a more abstract viewpoint on the problems at hand will be inspiring for practitioners. To this audience, this book may also be seen as a companion book, more biased towards mathematical analysis, of the recent review book on free energy methods [Chipot and Pohorille (2007b)].

Let us also insist at this point on the many possible application fields of the techniques presented in this book. Besides the obvious application domains where a free-energy difference is an important quantity *per se* since it has an experimental meaning (biology, physics, etc.), there also exist scientific fields where ratios of partition functions are required for computational purposes. An example is computational statistics (in particular Bayesian statistics). Moreover, it is often the case that some adequate importance sampling function is needed to enhance the sampling of multimodal probability distributions, and it is in general difficult to think of good candidates for high-dimensional problems. The free energy can be used as an efficient and automatic importance sampling function, once some "slowly evolving", "frustrated", or "metastable" degree of freedom has been identified. In particular, adaptive methods are very interesting since they provide an adaptive importance sampling strategy.

The book is organized linearly, but sometimes we need to anticipate on notions presented later on to motivate some techniques or concepts (especially in the introductory chapter). The notation is homogeneous throughout the book, and, for the reader's convenience, it is summarized in the Appendix. This will hopefully help the reader to keep track of the objects manipulated, in particular in the sections on constrained processes.

We now briefly describe the structure of the book, which is based on the mathematical classification explained at the end of Chapter 1. The introductory chapter presents the notions of statistical physics which will be constantly used in this book. It also gives the definition of *free energy.* This presentation is deliberately very different from that of standard physics textbooks, since our aim is primarily to describe how to compute average properties predicted by statistical physics, rather than motivating the physical relevance of these expressions. Chapter 2 presents in its first two sections standard techniques to compute averages in computational statistical physics, before describing a first set of techniques to compute free-energy differences using only these standard methods (*free-energy* *perturbation*, *histogram methods*). Chapter 3 is more technical, and deals mainly with *constrained processes*, which are a fundamental tool for the so-called *thermodynamic integration* method. *Nonequilibrium processes* are considered in Chapter 4, with variations and extensions around the famous *Jarzynski equality*. *Adaptive methods* are at the heart of Chapter 5. Our understanding of these recently proposed methods is not complete to date. We hope that our general mathematical presentation will motivate further research. The last chapter is a short presentation of *selection strategies*, which can be used as a complement to other methods. Chapter 5 is the part of the book where most of the open problems are listed.

Software deserves some comment. Most research groups in computational statistical physics are developing an in-house code, or are organized around the use of a well-known simulation package, dedicated to a family of applications. This code is often efficient, flexible, and allows one to treat challenging systems of current interest. These characteristics mean that the source files are not easily understood and modified, and that it is difficult to precisely know which numerical methods are used, and how they are actually implemented. For some users, the code is a black-box, whose results are too quickly trusted and too rarely questioned. We believe that running a code as a black-box and blindly trusting the outcome is risky. Questioning the validity of the results is a necessity. Besides, the reproducibility of the numerical simulations is as fundamental a rule as the reproducibility of any experiment in experimental sciences. For these reasons, we propose a series of codes for the two running examples considered throughout the book. They are freely available on Gabriel Stoltz's webpage[1] so that the numerical computations presented throughout this book can be double-checked. Our programs are not implemented in a computationally optimal way. On the other hand, the simplicity of the code and the documentation should make it simple enough for the reader to fully understand the details of one possible implementation of the methods and to test different numerical strategies.

We conclude this preface by emphasizing that computational statistical physics in general, and free-energy based techniques in particular, are a relatively recent research domain in applied mathematics. Many questions therefore remain open. Any comments about the scientific content, the pedagogical or non-pedagogical approaches used, typos, etc., are welcome! We

---

[1] Visit the webpage http://cermics.enpc.fr/∼stoltz/ or download directly the file from http://cermics.enpc.fr/∼stoltz/FreeEnergyCodes.tar.gz.

are very much indebted to many of our coworkers, for a careful re-reading of parts of this book: Manuel Athènes, Chris Chipot, John Chodera, Claude Le Bris, Frédéric Legoll, Kimiya Minoukadeh, and Raphaël Roux. We finally acknowledge the hospitality of the Hausdorff Institute of Mathematics, in Bonn (Germany), where we three authors attended a research program in spring 2008 and where the writing of this book was initiated. We are also grateful to the Banff International Research Station (Canada), where we were invited to a timely and stimulating workshop on "Mathematical and Numerical Methods for Free Energy Calculations in Molecular Systems" in June 2008.

Tony Lelièvre, Mathias Rousset and Gabriel Stoltz
Paris, January 2010

FREE ENERGY COMPUTATIONS - A Mathematical Perspective