Final Technical Report
Large Deviation Methods for the Analysis and Design of
Monte Carlo Schemes in Physics and Chemistry
DE-SC0002413

PI: Paul Dupuis
Co-PIs: Jimmie Doll, Hui Wang
Institution: Brown University

1 Objectives of the project

This proposal is concerned with applications of Monte Carlo to problems in physics and chemistry where rare events degrade the performance of standard Monte Carlo. One class of problems is concerned with computation of various aspects of the equilibrium behavior of some Markov process via time averages. The problem to be overcome is that rare events interfere with the efficient sampling of all relevant parts of phase space. A second class concerns sampling transitions between two or more stable attractors. Here, rare events do not interfere with the sampling of all relevant parts of phase space, but make Monte Carlo inefficient because of the very large number of samples required to obtain variance comparable to the quantity estimated.

The project uses large deviation methods for the mathematical analyses of various Monte Carlo techniques, and in particular for algorithmic analysis and design. This is done in the context of relevant application areas, mainly from chemistry and biology.

2 Participants

The senior researchers supported by this grant were:

- Paul Dupuis, PI
- Hui Wang, PI
- Nuria Plattner, postdoc
- Konstantinos Spiliopoulos, postdoc
- Xiang Zhou, postdoc
- Yufei Liu, senior PhD student (currently at Google)
- Dane Johnson, senior PhD student

3 Infinite swapping limit for parallel tempering

3.1 Problem formulation, issues and results

An important component of the research was the development of efficient simulation methods for the equilibrium distributions of high dimensional complex systems arising from chemistry and physics. For such systems, exact or approximate analytical solutions are nearly impossible and simulation is usually the only way to extract quantitative and qualitative information. There are two inter-related shortcomings in standard simulation-based approaches to these issues: (1) rigorous theoretical analyses of proposed algorithms are lacking; (2) the performance criteria typically used are not ideally suited to the problem. To alleviate the former, one typically conducts a large number of empirical studies. However, the degree of reliability or generality that can be inferred from such studies is questionable. As for the latter, two popular performance measures are asymptotic variance and distributional convergence rate, which is connected with the spectral properties of the underlying dynamics. However, simulation algorithms generally use an empirical measure to produce estimates, and important time averaging effects associated with empirical measures are not captured by these two criteria.

We introduced a performance criterion based on large deviation theory for the convergence of empirical measures. Since the efficiency of an algorithm is largely determined by how well it deals with the rare transitions between different parts of the state space, the theory of large deviations seems to be particularly suitable for performance analysis. We applied this criterion to a classical simulation scheme called parallel tempering, and developed a new and much more efficient algorithm. Parallel tempering (also known as replica exchange sampling) is one of the main tools used for the study of physical systems today. Hence a significant improvement over existing methods can be expected to find wide use.

In parallel tempering simulations are conducted in parallel for a set of temperatures, and the key improvement over standard Monte Carlo is a swap mechanism that exchanges configurations between these parallel simulations at a given rate. The mechanism is designed to allow the low temperature system to escape from deep local energy minima where it might otherwise be trapped, via those swaps with the higher temperature components. Based on large deviation theory, we argued that the rate of convergence of the empirical measure is a monotone increasing function of the swap rate. This suggested that one should raise the swap frequency in order to improve effi-

ciency, but this is eventually counter-productive since eventually most of the computational effort is directed towards swapping and little towards moving the process dynamics. However, it turned out that one could construct a simulation scheme that is equivalent to the limit of the parallel tempering schemes in the sense of distributions, but which involves no swapping at all. With this scheme, which we call *infinite swapping* (INS), the effect of the swapping is captured by a collection of weights that influence both the dynamics and the empirical measure.

While the infinite swapping scheme optimizes the convergence rate as described above, it has practical limitations in implementation due to the complexity of the weights when the number of temperatures is large (>7). Complex problems in chemistry and biology often involve scores of temperatures, and so it was critical to overcome this limitation. We developed an approximation to the full infinite swapping which is based on alternating between *partial infinite swapping* (PINS) algorithms, which can be shown to approximate (theoretically and practically) the INS scheme.

The mathematical theory for the INS and PINS is developed in [6], and is based on large deviation results from [3]. Numerical studies on fairly complex Leonard-Jones systems are conducted in reference [8]. Improvements of three orders of magnitude in performance over conventional parallel tempering were observed at an increased computational cost of 5-15%. Applications to complex problems from biology are developed in [14], with similar improvements in performance. An easily implemented diagnostic for convergence is introduced and analyzed in [5].

4 Importance sampling for multiscale dynamics

4.1 Problem formulation, issues and results

This portion of the project considered Monte Carlo simulation methods for complex systems, but the quantities of interest are the probability of transitions between different parts of the state space (and are not equilibrium properties). Specifically, we developed accelerated Monte Carlo methods such as importance sampling and spitting schemes to deal with rough energy landscapes. Such energy landscapes can have a relatively small number of macroscopic local minima [also referred to as “wells”]. By “macroscopic” we mean the minima would persist even if the landscape were smoothed to some degree. Superimposed are many more “microscopic” minima that would be removed by smoothing. It is important to understand transitions between the macroscopic minima. The presence of the microscopic local

minima introduces a homogenization effect that alters the effective diffusiveness of the system. Even though the primary determining factor of a good simulation scheme should be the properties of the macroscopic wells, the homogenization effect due to local minima should not be overlooked in the design of Monte Carlo schemes.

The design of Monte Carlo numerical schemes for the estimation of small probabilities was for a long time guided by heuristics that were ultimately found to be valid only in special circumstances. The PIs established, in a series of papers that focused on processes without multiscale effects, that necessary and sufficient conditions for the design of effective schemes could be phrased in terms of what are called “subsolutions” to an associated partial differential equation. The first step in such an approach is to understand the large deviation theory for the underlying process model. Large deviation theory gives approximations to probabilities of rare events, and has been found to be crucial in designing and understanding fast simulation techniques such as importance sampling. The paper [10] develops the large deviation theory for a class of multiscale stochastic differential equations. To be specific, the paper considers the following d -dimensional small noise diffusion model which is often used to describe systems with rough energy landscapes:

$$dX_t^\varepsilon = \left[\frac{\varepsilon}{\delta} b \left(X_t^\varepsilon, \frac{X_t^\varepsilon}{\delta} \right) + c \left(X_t^\varepsilon, \frac{X_t^\varepsilon}{\delta} \right) \right] dt + \sqrt{\varepsilon} \sigma \left(X_t^\varepsilon, \frac{X_t^\varepsilon}{\delta} \right) dW_t.$$

Here ε is the small noise parameter which is often determined by the temperature, and δ is the fast oscillation parameter that measures the “roughness” of the energy landscape. The coefficients were assumed to be periodic in the second (fast) variable, and a large deviations analysis was carried out in [10] for all possible limits of ε/δ as $\varepsilon \rightarrow 0$ and $\delta \rightarrow 0$ (problems from physical systems usually correspond to the case $\varepsilon/\delta \rightarrow \infty$).

With the large deviation theory understood, the next step was to adapt methods developed by the PIs which yield necessary and sufficient conditions for the construction of good sampling schemes to problems with multiscale effects. Since the large deviation theory has only (to date) been developed for the multi scale model above with periodic coefficients, the theoretical results of [10] regarding performance of the importance sampling schemes are valid only for this class of systems. However, one can formally adapt the schemes for the periodic case to more realistic models, such as those where the fast oscillations of the coefficients are given in terms of an ergodic random field. Both types of models were tested numerically in [7], and the resulting Monte Carlo schemes performed as theoretically predicted. The solution

of something known as a “cell problem,” which was essential in analyzing the large deviation properties, not surprisingly is a needed ingredient in the design of nearly optimal importance sampling schemes. The paper [1] focuses on use of “moderate deviations” as a tool for the development of algorithms when events are rare but not too rare, and [4] is the first paper to consider the particularly difficult situation where a rest point is within the domain of the simulation.

5 Robustness for uncertainty quantification

Much of uncertainty quantification, including the Monte Carlo methods described previously, focus on determining the effect of variables modeled probabilistically, and with a known distribution, on some physical or engineering system. In [2] and [9] we developed methods to obtain information on the system when the distributions of some variables are known exactly, others are known only approximately, and perhaps others are not modeled as random variables at all. The main tool used in [9] is the duality between risk-sensitive integrals and relative entropy. We obtained explicit bounds on standard performance measures (variances, exceedance probabilities) over families of distributions whose distance from a nominal distribution is measured by relative entropy. These are robust bounds for the impact of modeling error on ordinary cost structures, i.e., expected values and probabilities determined by events that are not rare. An extension to robust bounds on expected values and probabilities determined by rare events, which would have relevance for the first two topics of this project, is given in [2]. It is based on the use of Rényi entropy, and will be further developed.

6 Preprints and publications

1. Moderate deviations for recursive stochastic algorithms, (P. Dupuis and D. Johnson), submitted.
2. Robust bounds on risk-sensitive functionals via Rényi divergence, (R. Atar, P. Dupuis and K. Chowdhary), submitted.
3. On the large deviation rate function for the empirical measures of reversible jump Markov processes, (P. Dupuis and Y. Liu), to appear in the *Annals of Probability*.
4. Escaping from an attractor: Importance sampling and rest points I, (P. Dupuis, K. Spiliopoulos and X. Zhou) submitted.

5. Rare-event sampling: occupation-based performance measures for parallel tempering and infinite swapping Monte Carlo methods, (J. D. Doll, N. Plattner, D. L. Freeman, Y. Liu and P. Dupuis), *J. of Chemical Physics*, **137**, (2012), 204112.
6. On the infinite swapping limit for parallel tempering (P. Dupuis, Y. Liu, N. Plattner, J.D. Doll), *SIAM J. Multiscale Model. and Simul.*, **10**, (2012), 986–1022.
7. Importance sampling for multiscale diffusions (P. Dupuis, K. Spiliopoulos and H. Wang), *SIAM J. Multiscale Model. and Simul.*, **10**, (2012), 1–27.
8. An infinite swapping approach to the rare-event sampling problem (N. Plattner, J.D. Doll, P. Dupuis, H. Wang, Y. Liu and J. Gubernatis), *J. of Chemical Physics*, **135**, (2011), 134111.
9. Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification (K. Chowdhary and P. Dupuis), *ESAIM: Mathematical Modelling and Numerical Analysis*, **47**, (2013), 635–662.
10. Large deviations for multiscale diffusions via weak convergence methods (P. Dupuis and K. Spiliopoulos), *Stochastic Processes and Their Applications*, **122**, (2012), 1947–1987.
11. Spatial averaging applied to small molecule diffusion in condensed phase environments, (N. Plattner, J. D. Doll, and M. Meuwly), *J. Chem. Phys.* **133** 044506 (2010).
12. Convergence characteristics of the cumulant expansion for Fourier path integrals, (S.D. Kuniyev, D.L. Freeman and J.D. Doll), *Phys. Rev. E*, **81**, 066707 (2010).
13. Numerical investigation of the cumulant expansion for Fourier path integrals, (N. Plattner, S. Kuniyev, D.L. Freeman, and J.D. Doll), *PARA 2010*, Part II, Lecture Notes on Computer Science, vol. 7134, K. Jónasson (ed.) (Springer-Verlag Berlin Heidelberg (2012)), pp. 13–22.
14. Overcoming the rare-event sampling problem in biological systems with infinite swapping, (N. Plattner, J.D. Doll and M. Meuwly), *J. Chem. Theory Comput.*, **9**, 4215–4224, (2013).