Thermodynamic integration methods, infinite swapping, and the calculation of generalized averages

J. D. Doll, P. Dupuis, and P. Nyquist

Citation: The Journal of Chemical Physics 146, 134111 (2017);
View online: https://doi.org/10.1063/1.4979493
View Table of Contents: http://aip.scitation.org/toc/jcp/146/13
Published by the American Institute of Physics

Articles you may be interested in

A Gaussian theory for fluctuations in simple liquids
The Journal of Chemical Physics 146, 134507 (2017); 10.1063/1.4979659

Perspective: Dissipative particle dynamics
The Journal of Chemical Physics 146, 150901 (2017); 10.1063/1.4979514

Surface tension of a Yukawa fluid according to mean-field theory
The Journal of Chemical Physics 146, 134701 (2017); 10.1063/1.4979203

Cheap but accurate calculation of chemical reaction rate constants from ab initio data, via system-specific, black-box force fields
The Journal of Chemical Physics 147, 161701 (2017); 10.1063/1.4979712

Grand canonical electronic density-functional theory: Algorithms and applications to electrochemistry
The Journal of Chemical Physics 146, 114104 (2017); 10.1063/1.4978411

Nucleation of urea from aqueous solution: Structure, critical size, and rate
The Journal of Chemical Physics 146, 134501 (2017); 10.1063/1.4979141
Thermodynamic integration methods, infinite swapping, and the calculation of generalized averages

J. D. Doll, P. Dupuis, and P. Nyquist

1Department of Chemistry, Brown University, Providence, Rhode Island 02912, USA
2Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912, USA

(Received 1 November 2016; accepted 18 March 2017; published online 6 April 2017)

In the present paper we examine the risk-sensitive and sampling issues associated with the problem of calculating generalized averages. By combining thermodynamic integration, stationary phase Monte Carlo, and infinite swapping techniques, we develop an approach for such problems and explore its utility for a prototypical class of applications. Published by AIP Publishing. [http://dx.doi.org/10.1063/1.4979493]

I. INTRODUCTION

Monte Carlo methods are a valuable tool for the study of many-dimensional problems in a variety of disciplines. By providing a general means of investigating the properties of well-defined, physically realistic models without resorting to untestable numerical approximations, they are an essential element in obtaining the insight necessary for the construction of valid conceptual models of complex phenomena.

An important use of Monte Carlo methods is providing numerical estimates of averages of the type that arise naturally in a variety of classical and quantum statistical-mechanical contexts. In the present work, we wish to consider the computation of generalized averages of the form

$$\langle e^{bx}\rangle_S = \frac{\int e^{-S(x,\alpha)} e^{bx} dx}{\int e^{-S(x,\alpha)} dx}, \quad (1.1)$$

where $S$ and $b$ are both potentially complex. In Eq. (1.1), $x$ and $\alpha$ represent the coordinate(s) and system parameters of the problem, respectively. Real valued moment generating functions of this type are common in the calculation of equilibrium thermodynamic properties. As discussed in Section V, generalized averages in which $S$ is complex arise in quantum dynamical applications. Averages of the type in Eq. (1.1) provide a useful prototype for both classes of problems. Moreover, their treatment represents a general demonstration of the ability to compute averages of any functions that can be written in Fourier or Laplace form. For convenience we utilize a pseudo-one-dimensional notation in the following with the understanding that multidimensional generalizations of all results are readily obtained.

In confronting generalized averages of the type in Eq. (1.1), there are a number of core issues. Chief among them are the choice of an appropriate sampling density and the selection/design of sampling methods to assure that all regions of importance for that density are properly included in the final average. In the present work, we combine stationary phase Monte Carlo (SPMC) and infinite swapping (INS) techniques to accomplish these twin tasks. The INS computational ensemble in the present developments is based on a spatial rather than a thermal control parameter. In addition to the utilization of improved sampling methods, we find that a reformulation of the underlying problem using Kirkwood style thermodynamic integration techniques proves advantageous.

The remainder of this paper is organized as follows. In Section II we examine the risk sensitive nature of moment generating functions of the type in Eq. (1.1) and discuss the advantages of a Kirkwood approach for their evaluation. Methods for implementing the Kirkwood approach for complex generalizations of Eq. (1.1) are presented in Section III, and illustrative examples are presented and discussed in Section IV.

II. RISK SENSITIVITY AND THE KIRKWOOD FORMULATION

In considering the evaluation of averages of the type in Eq. (1.1) it is useful to note that there are a number of possible approaches. One is the direct application of Monte Carlo methods. Specifically, if $S(x,\alpha)$ is real, a natural route is to replace the continuous average in Eq. (1.1) with the discrete average of the integrand, $\exp(bx)$, over a finite set of points obtained from a Monte Carlo sampling of the problem’s natural density, $\exp(-S(x,\alpha))$. If $S(x,\alpha)$ is not real, a case examined in greater detail in Sec. III, the choice of an appropriate density for use as an importance function becomes more subtle. In either situation rather than approaching Eq. (1.1) directly, it proves useful to consider restructuring it.

Using techniques familiar from “thermodynamic integration” methods, the average we seek can be recast exactly as

$$\langle e^{bx}\rangle_S = \exp\left(\int_0^b \langle \chi \rangle_d d\lambda\right), \quad (2.1)$$

where

$$\langle \chi \rangle_d = \frac{\int e^{-S_d(x,\alpha)} dx}{\int e^{-S(x,\alpha)} dx}, \quad (2.2)$$

...
and where

\[ S_1(x, \alpha) = S(x, \alpha) - \lambda x. \]  \hspace{1cm} (2.3)

We refer to expressions such as Eq. (2.1) in what follows as the “Kirkwood” form of the original average, Eq. (1.1).

While the Kirkwood and original forms of the average are equally valid, they can differ significantly in their sensitivities to statistical noise. Anticipating an ultimate evaluation by stochastic means, this difference in sensitivity is potentially an important practical matter. To illustrate the issues involved, it is instructive to compare the sensitivity of the direct and Kirkwood approaches for a simple example where \( b \) is real and \( S \) is a quadratic. Assuming \( S \) to be of the form

\[ S(x, x_0) = \frac{1}{2} (x - x_0)^2, \]  \hspace{1cm} (2.4)

where \( x_0 \) is a real constant, the average specified by Eq. (1.1) is given analytically by

\[ \langle e^{bx} \rangle_{\text{Exact}} = e^{bx_0 + \frac{\sigma^2}{2}}, \]  \hspace{1cm} (2.5)

while the standard deviation of \( e^{bx} \) with respect to the natural density of the problem, \( \exp(-S(x, x_0)) \), is given by

\[ \sigma = \left[ \langle e^{2bx} \rangle_S - \langle e^{bx} \rangle_S^2 \right]^{1/2} = e^{bx_0 + \frac{\sigma^2}{2}} \left( e^{\lambda^2} - 1 \right)^{1/2}. \]  \hspace{1cm} (2.6)

From Eqs. (2.5) and (2.6), we see that the error associated with a direct, N-point Monte Carlo (DMC) estimate of Eq. (1.1) scales poorly with respect to the parameter \( b \). In particular, assuming \( N \) independent Monte Carlo sample points drawn from the density associated with \( \exp(-S(x, x_0)) \), the DMC estimate of the moment generating function, \( \langle e^{bx} \rangle_{\text{DMC}} \), is given by the exact value plus a random variable whose standard deviation is \( \sigma \sqrt{N} \). In other words, the ratio of the DMC estimate of the moment generating function to its exact value scales as

\[ \frac{\langle e^{bx} \rangle_{\text{DMC}}}{\langle e^{bx} \rangle_{\text{Exact}}} = 1 + \xi_{\text{DMC}}, \]  \hspace{1cm} (2.7)

where \( \xi_{\text{DMC}} \) is a random variable whose standard deviation is given by \( (e^{b^2} - 1)^{1/2}/\sqrt{N} \). This result makes the direct approach unworkable except for small values of \( b \). Such extreme sensitivity to noise is a general characteristic of “risk-sensitive” problems, in which the variance of the integrand in question is dominated by regions of the underlying integration, which are in the tails of the importance function.

In contrast with the results for the direct Monte Carlo approach, the ratio analogous to that in Eq. (2.7) for the Kirkwood approach in the large \( N \) limit is given by

\[ \frac{\langle e^{bx} \rangle_{K}}{\langle e^{bx} \rangle_{\text{Exact}}} = 1 + \xi_K, \]  \hspace{1cm} (2.8)

where \( \xi_K \) is a random variable whose standard deviation scales as \( \sqrt{b}/N \). This result follows from the assumptions that the number of quadrature points required for the numerical integration in Eq. (2.1) increases linearly with the size of the integration domain, \( b \), and that the variance in each of the \( \langle x \rangle \) terms in Eq. (2.1) is independent of \( \lambda \). The computational moral to this story is that the Kirkwood and direct approaches can differ significantly with respect to their inherent sensitivities to Monte Carlo noise. Either alone or in combination with traditional variance reduction methods discussed in the first portion of Appendix A, the Kirkwood approach thus offers a potential means for avoiding/dealing with risk-sensitive issues that arise.

III. METHODS FOR COMPLEX AVERAGES

In computing averages of the type in Eq. (1.1) or in its Kirkwood equivalent, Eqs. (2.1) and (2.2), a number of practical issues arise when \( S(x, \alpha) \) is complex. A major one is the choice of an importance function. A tempting choice is the modulus, \( \exp(-S) \). The (potentially) highly oscillatory nature of the integrand, however, means that the important regions of the integrand are no longer dictated exclusively by \( \exp(-S) \), but rather by a competition between that modulus and the stationary phase regions of the problem.

Stationary phase Monte Carlo (SPMC) techniques, described in detail elsewhere, have been developed for dealing with integrals of highly oscillatory integrands. Briefly summarized, these approaches are based on the observation that there exists a group of transformations of the integrands, which leave the value of the associated integral unchanged. For integrals over an infinite or periodic domain, which converge sufficiently rapidly such that orders of integration can be interchanged, integrals of a function and its convolution with an arbitrary probability density are equal. That is, for a function \( f(x) \) and a normalized probability density \( P_\epsilon(y) \),

\[ \int f(x)dx = \int \langle f(x) \rangle_\epsilon dx, \]  \hspace{1cm} (3.1)

where

\[ \langle f(x) \rangle_\epsilon = \int P_\epsilon(y)f(x + \gamma)dy. \]  \hspace{1cm} (3.2)

Although the integrands of the left and right hand sides of (3.1) differ, the corresponding integrals are equal. Thinking of \( P_\epsilon(y) \) as a density of dimension \( \epsilon \), the “pre-averaging” process in Eq. (3.2) serves to damp the integrand’s oscillations on a controllable length scale.

Applying the SPMC approach to the generalized average in Eq. (2.2), we have

\[ \langle x \rangle_\epsilon = \frac{\int \left\{ e^{S_1(x, \alpha)} \right\}_\epsilon dx}{\int \left\{ e^{S_1(x, \alpha)} \right\}_{\text{Exact}} dx}. \]  \hspace{1cm} (3.3)

The result in Eq. (3.3) is formally correct, independent of the parameter(s) \( \epsilon \), and of a form that is readily amenable to importance sampling.

To make progress, we now seek suitable importance function(s) for the evaluation of Eq. (3.3) and rewrite that expression as

\[ \langle x \rangle_\epsilon = \frac{\int W_N(x) \left\{ e^{S_1(x, \alpha)} \right\}_N dx}{\int W_N(x) \left\{ e^{S_1(x, \alpha)} \right\}_{\text{Exact}} dx}. \]  \hspace{1cm} (3.4)

The optimal variance reduction generally requires that we choose different importance functions for \( W_N(x) \) and \( W_D(x) \)
in Eq. (3.4). Provided that the variation of $x$ relative to that of $\exp(-S(x,a))$ is small on the length scale $\varepsilon$; however, it is both reasonable and convenient to utilize a common importance function, denoted $W_\varepsilon(x)$, for both tasks accepting that such a choice may lead to sub-optimal variance reduction.

As discussed in the later portions of Appendix A, a generalization of the standard arguments for optimal variance reduction provides a natural choice for the importance function in the denominator of Eq. (3.4), $W_\varepsilon(x)$. Those arguments show that

$$W_\varepsilon(x) = \left|\left(e^{-S_1(x,a)}\right)\right|_\varepsilon.$$

Using this choice in both the numerator and denominator, Eq. (3.4) becomes

$$\langle \lambda \rangle \approx \frac{\int W_\varepsilon(x)\left(e^{-S_1(x,a)}\right) / W_\varepsilon(x) dx}{\int W_\varepsilon(x)\left(e^{-S_1(x,a)}\right) / W_\varepsilon(x) dx}. \tag{3.6}$$

Consistent with the use of a common importance function for both numerator and denominator portions of Eq. (3.6), it is reasonable to assume that the variation of $x$ is small relative to that of $\exp(-S)$ on the length scale $\varepsilon$. This means that

$$\left(e^{-S_1(x,a)}\right)_\varepsilon \approx \left(e^{-S_1(x,a)}\right) x. \tag{3.7}$$

Together, the overall effect of these steps is tantamount to rewriting Eq. (3.3) as

$$\langle \lambda \rangle = \frac{\int \left(e^{-S_1(x,a)}\right) \cdot x dx}{\int \left(e^{-S_1(x,a)}\right) dx}, \tag{3.8}$$

an expression for which the common importance function specified in Eq. (3.5) is useful.

For present purposes, we compute $\langle \exp(-S) \rangle_\varepsilon$ in Eq. (3.8) by approximate means and use its modulus as the relevant importance function to obtain a Monte Carlo estimate of the desired $\langle \lambda \rangle$, values, taking care to examine the effects of such approximations on the final results. Unlike $\exp(-S)$, if the $\varepsilon$-length scale is properly chosen, the modulus of $\langle \exp(-S) \rangle_\varepsilon$ provides a suitable importance function for the evaluation of the average in Eq. (3.8). In practical terms, $\varepsilon$ must be chosen small enough that the approximation in Eq. (3.7) is valid, but large enough that the irrelevant, non-stationary phase regions are suppressed and the important regions are emphasized. Issues related to the choice of $\varepsilon$ have been discussed previously and are considered in greater detail in Sec. IV. It should be noted that the approximation in Eq. (3.7) can be removed at the expense of introducing secondary Monte Carlo estimates of the $\varepsilon$-averages involved in Eq. (3.6).

Gradient methods are a convenient means for constructing the $\varepsilon$-averages required to implement the SPMC pre-average in the present approach. Assuming a Gaussian form for $P_\varepsilon(y)$ (cf. Eq. (3.2)) through second-order the gradient approach gives

$$\langle e^{-S(x,a)} \rangle_\varepsilon = \frac{\exp \left\{ -S(x,a) + \frac{1}{2}(eS'(x,a))^2 / (1 + e^2 S''(x,a)) \right\} }{(1 + e^2 S''(x,a))^{1/2}}. \tag{3.9}$$

where $S'$ and $S''$ denote the first and second derivatives of $S$, respectively. Analogous first-order and multidimensional approximations are easily obtained.

Monte Carlo applications of the type under discussion frequently involve sparse sampling issues. When the probability density that underlies the average in question is composed of isolated or weakly connected regions, special care must be exercised to assure that all regions of importance are properly included. Failure of the sampling method to provide a proper accounting is, in practice, both computationally destructive and difficult to detect. Such difficulties, present in conventional real-valued forms of Eq. (1.1), become even more problematic in analogous complex averages where the potentially highly oscillatory nature of the integrand plays a key role. In essence, the SPMC approach is the exchange of a problem involving severe phase oscillations for one of sparse sampling.

A variety of techniques have been developed for dealing with the general sparse sampling problem. One approach, parallel tempering, utilizes a computational ensemble composed of the product of densities for a set of control parameters (typically the temperature). Rather than studying the various ensemble members individually, parallel tempering studies the entire ensemble in unison. Ordinary random walk displacements are augmented with trial moves based on attempted swaps of configurations between the different data streams. By demanding that detailed balance be preserved for such swaps, the resulting approach provides a practical means for using information from the more highly connected members of the ensemble to improve the efficiency of sampling for the more weakly connected ones.

The recently developed infinite swapping (INS) approach is a sparse sampling strategy based on a large deviation analysis of parallel tempering. It represents the extreme limit of parallel tempering in which swaps involving all possible temperatures are attempted at an infinitely rapid rate, a limit the large deviation analysis proves to be optimal. Operationally, the method utilizes a probability density composed of a symmetrized sum of parallel tempering like product densities, a form that is more highly connected than the original. The INS approach represents the conscious use of symmetry as a tool for dealing with the sparse sampling problem. Practical methods for implementing the approach for arbitrary sized ensembles have been developed and discussed in detail elsewhere and are briefly summarized in Appendix B.

Combining the items discussed above, we propose an approach to the construction of averages of the type in Eq. (2.1) that consists of three elements:

- a Kirkwood-like formulation of the problem to deal with its risk sensitive aspects;
- SPMC methods to suppress phase oscillations and to produce a suitable importance function (Eq. (3.5)); and
- INS techniques to treat the sparse sampling issues arising from use of the SPMC approach.

In conventional parallel tempering simulations, the system temperature is typically utilized as the control parameter for the creation of the expanded computational ensemble. The various data streams within such simulations thus produce estimates of thermodynamic properties for the different
temperatures within the ensemble. In the present work the control parameter for the INS ensemble is the SPMC length scale, \( \epsilon \). Recalling that the overall SPMC results are independent of the choice of this length scale (cf. Eq. (3.1)), we see that the different data streams in the present approach are estimates of the same computational object. Because they correspond to different SPMC length scales, however, the statistical variance of these estimates will generally differ.

IV. RESULTS AND DISCUSSION

In this section, we illustrate the current approach with an application to a model average of the form

\[
\phi(\eta) = \langle e^{i \eta x} \rangle_S = \frac{\int e^{-S(x)} e^{i \eta x} dx}{\int e^{-S(x)} dx}, \tag{4.1}
\]

where \( S(x) \) is a complex quantity specified by

\[
S(x) = \frac{1}{2} \left( \frac{x - x_0}{\sigma} \right)^2 - i \left( \frac{x^3}{3} \right). \tag{4.2}
\]

To streamline the notation in Eq. (4.2) and in the following discussion, the explicit \( x_0 \) and \( \sigma \) labels in the expression for \( S(x) \) will be omitted. Simple enough that key aspects of the method can be readily investigated, the present model is nonetheless sufficiently complex to reflect the general computational challenges involved.

The nature of the underlying average in Eq. (4.1) changes character as a function of the parameters \( \eta \) and \( \sigma \). In the small \( \eta \) limit, \( \exp(-S) \) covers the important regions of the problem making direct Monte Carlo methods generally applicable. In the limit that \( \eta \) is large and negative, however, the phase oscillations for the integrand in the numerator of Eq. (4.1) become severe, and the stationary phase regions at \( \pm (\eta)^{3/2} \) play a dominant role. Depending on the value of \( \sigma \), these stationary phase regions may or may not fall within the natural range of \( \exp(-S) \). In any case in a conventional Monte Carlo approach, the irrelevant regions of the problem would be established in an inefficient, after-the-fact manner through numerical cancellation involving poorly placed Monte Carlo points. In contrast, if the SPMC length scale is properly chosen, the importance function, \( \{ e^{-S(x)/\lambda} \}^\lambda \), is concentrated in the regions that dominate the final result, and the inefficient numerical cancellation issue is avoided.

The Kirkwood form of Eq. (4.1) is

\[
\phi(\eta) = \exp \left( \int_0^\eta \langle x \rangle_\lambda d\lambda \right), \tag{4.3}
\]

where

\[
\langle x \rangle_\lambda = \frac{\int e^{-S(x)} x dx}{\int e^{-S(x)} dx}, \tag{4.4}
\]

and where

\[ S_\lambda(x) = S(x) - \lambda x. \tag{4.5} \]

To produce an estimate of the original computational objective in the Kirkwood approach, the essential numerical tasks are to evaluate \( \langle x \rangle_\lambda \) on a grid of \( \lambda \)-values and to perform the associated one-dimensional \( \lambda \)-integration. To calculate the necessary \( \langle x \rangle_\lambda \) values, we rewrite Eq. (4.4) using Eq. (3.8) of Section III as

\[
\langle x \rangle_\lambda = \frac{\int \left( e^{-S_\lambda(x)} \right) x dx}{\int \left( e^{-S_\lambda(x)} \right) dx}. \tag{4.6}
\]

INS techniques can then be used to evaluate \( \langle x \rangle_\lambda \) as a function of \( \lambda \), for each of the \( \epsilon \)-values in the ensemble. Once the necessary \( \langle x \rangle_\lambda \) values are prepared, conventional numerical quadrature techniques can be used to perform the \( \lambda \)-integration in Eq. (4.3). Unless otherwise noted, all numerical results presented in the present studies utilize

- the Gauss-Airy model defined by Eq. (4.2),
- second-order gradient approximations (Eq. (3.9)) for the necessary SPMC averages,
- a 5-member INS ensemble based on a range of \( \epsilon \) values \((\epsilon = (0.00,0.05,0.10,0.20,0.40))\) chosen by methods outlined below,
- Metropolis single-variable techniques in combination with the heat bath method outlined previously\(^\text{10}\) and in Appendix B to perform the necessary sampling, and
- trapezoidal quadrature to perform the one-dimensional \( \lambda \)-integration in Eq. (4.3).

We begin by first investigating the quality of the gradient level implementation of the approach and the \( \epsilon \)-independence of the results of Eq. (4.6). From the discussion in Section III, we know that the results of Eq. (3.3) are formally independent of the choice of the \( \epsilon \) parameter. Table I examines the extent to which this is also true of the approximate result in Eq. (4.6) for the set of five \( \epsilon \)-values and system parameters used in the present studies. Shown in Table I are the numerical values of \( \langle \chi \rangle_\lambda \) for the present model as a function of \( \epsilon \) for two representative, large negative values of \( \lambda \). All results in Table I are computed for \( \sigma = 1 \) and \( x_0 = 0.5 \) using Mathematica to perform the necessary integrations. The simplicity of the present model problem permits the use of such conventional methods to provide an unambiguous test of the level of \( \epsilon \)-independence of Eq. (4.6). More generally, the presence or absence of such \( \epsilon \)-independence will in practice be signaled by the internal consistency of the calculated results for the various INS ensemble members. We see from Table I that the \( \langle \chi \rangle_\lambda \) results display only a very weak dependence on \( \epsilon \) over the range studied, thus justifying the quality of the gradient level methods and the use of Eq. (4.6).

![Table I. Shown are the values of \( \langle \chi \rangle_\lambda \).](image)

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>( \langle \chi \rangle_\lambda (\lambda = -16) )</th>
<th>( \langle \chi \rangle_\lambda (\lambda = -25) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>3.822 - 0.463 i</td>
<td>4.926 - 0.412 i</td>
</tr>
<tr>
<td>0.05</td>
<td>3.822 - 0.463 i</td>
<td>4.926 - 0.412 i</td>
</tr>
<tr>
<td>0.10</td>
<td>3.822 - 0.463 i</td>
<td>4.926 - 0.412 i</td>
</tr>
<tr>
<td>0.20</td>
<td>3.822 - 0.461 i</td>
<td>4.926 - 0.411 i</td>
</tr>
<tr>
<td>0.40</td>
<td>3.894 - 0.442 i</td>
<td>4.975 - 0.435 i</td>
</tr>
</tbody>
</table>
involved are listed in the figure caption. To facilitate the comparison of the various results, all densities in Fig. 1 are normalized to unity.

The changing character of the importance functions with SPMC length scale is evident in the results of Fig. 1. For smaller $\varepsilon$, $W_\varepsilon(x)$ results reflect the real portions of $S(x)$ and are thus essentially unimodal Gaussians centered on the corresponding values of $x_0$. As $\varepsilon$ increases, this initial unimodal density tends to increase in width and then ultimately to develop a structure that reflects the underlying stationary phase regions of the problem. The sharpness of the resulting stationary phase feature(s) varies with $\varepsilon$, being most highly focused when the $\varepsilon$ length scale matches the natural width(s) of those region(s).

The $\varepsilon$-dependence of the SPMC density is conveniently summarized by the information entropy, $S_{\text{info}}$, associated with the SPMC importance function, $W_\varepsilon(x)$, and defined by

$$S_{\text{info}} = -\frac{\int W_\varepsilon(x) \ln(W_\varepsilon(x)) dx}{\int W_\varepsilon(x) dx} + \ln \left( \int W_\varepsilon(x) dx \right).$$

Shown in Fig. 2, for example, are the plots of the information entropy for the densities of the system described in panel (a) of Fig. 1 for $\lambda = -16$ and $-25$ as a function of the SPMC length scale, $\varepsilon$. The increases in entropy visible in the small $\varepsilon$ regions of Fig. 2 correspond to the broadening of the initial unimodal Gaussian densities centered at $x_0$. As $\varepsilon$ continues to increase, the information entropies peak, go through minima, and ultimately increase as the initial unimodal densities first split, sharpen, and then broaden. The peak in the information entropy as a function of $\varepsilon$ thus serves as a rough indicator of the $\varepsilon$ value for which the SPMC importance function begins to reflect qualitatively the inherent stationary phase character of the problem, roughly 0.2 for the systems in Fig. 2. The minimum in the information entropy, on the other hand, provides a practical guide for the $\varepsilon$ value that produces the maximally compressed SPMC density, roughly 0.4 for the systems in Fig. 2. Such considerations form the basis for the selection of the INS computational ensemble for the present example. In general applications, information entropy differences rather than the absolute entropies provide a more readily computed basis for such decisions.

Shown in Fig. 3 are the real and imaginary parts of $\langle x \rangle_\varepsilon$, as a function of $\lambda$, computed from Eq. (4.6) using the present INS/SPMC approach. The $\langle x \rangle_\varepsilon$ results shown are those for $\sigma = 1, \varepsilon = 0.40$ for two different choices of $x_0$ obtained using $10^6$ single particle Monte Carlo moves for each of a discrete grid of $\lambda$-values (grid spacing = 0.05). As can be seen in the large $\varepsilon$-results of Fig. 1, the stationary phase regions that dominate the present averages for large negative $\lambda$-values are isolated and represent a small fraction of the total integration volume. The resolution and lack of statistical noise in the results of Fig. 3 (independently computed for each of the grid points involved) demonstrate that the INS approach is effective in dealing with the rare-event sampling issues involved.

To validate the results for the chosen model and to understand better the performance of the present computational approach, it is useful to examine selected sequences of results. Shown in Fig. 4(a) are a number of $\langle x \rangle_\varepsilon$ results obtained using various system parameters and numbers of Monte Carlo points. For simplicity we display only the real portions of $\langle x \rangle_\varepsilon$. The behavior of the analogous imaginary quantities is qualitatively similar in all cases. We consider first the $\langle x \rangle_\varepsilon$ values for a
FIG. 4. (a) Additional detail of Re($\langle x \rangle_\lambda$) for Gauss/Airy model of Fig. 3 ($x_0 = 0.50$) as a function of $\lambda$ for various $\epsilon$-values. Results for $\epsilon = 0.40$ (black) and 0.20 (red) are computed using 100 loops of $10^4$ points. Results are shown for $\epsilon = 0.10$ computed using 100 loops of $10^5$ points (green), $10^6$ points (blue), $10^7$ points (orange), and $10^8$ points (brown). $\sigma = 1$ for all results. For clarity, the error bars are not shown but can be inferred from the level of jitter in the results. (b) Blowup of large negative $\lambda$ region of (a). Shown are the real parts of $\langle x \rangle_\lambda$ as a function of $\lambda$ for $x_0 = 0.50$ obtained using $\epsilon = 0.40$ (black) and $\epsilon = 0.20$ (red). Both simulations utilized 100 loops of $10^4$ MC points and $\sigma = 1$. For clarity, the error bars are not shown but can be inferred from the level of jitter in the results.

fixed number of Monte Carlo points for varying $\epsilon$-values and then examine analogous $(x)_\lambda$ results for varying numbers of Monte Carlo points for a fixed value of $\epsilon$. The blue, red, and black curves in Fig. 4(a) show the real portions of $(x)_\lambda$ values obtained for $x_0 = 0.5$ and $\sigma = 1$ for three of the five ensemble $\epsilon$ values, $\epsilon = (0.10, 0.20, 0.40)$, respectively, using $10^6$ Monte Carlo points. The associated results for $\epsilon = 0.05$ and 0.00 (not shown) are qualitatively similar to those of $\epsilon = 0.10$. Although “noisier,” the $(x)_\lambda$ results for $\epsilon = 0.20$ (red curve) are in basic agreement for those for $\epsilon = 0.40$ (black curve—obscured by red curve) over the entire $\lambda$-range shown in Fig. 4(a). The level of the agreement between the $\epsilon = 0.20$ and 0.40 results is shown in greater detail in Fig. 4(b). The $(x)_\lambda$ results for $\epsilon = 0.10$ (blue curve) in Fig. 4(a), on the other hand, agree with those of the larger $\epsilon$ values for the smaller $\lambda$-range (albeit with greater noise), but exhibit systematic errors for large negative $\lambda$-values. At first glance these systematic errors for large negative $\lambda$-values seem inconsistent with the results of Table I.

It is important to note, however, that the results in Table I utilize high-precision, direct quadrature while those in Fig. 4(a) are Monte Carlo estimates based on a fixed number of points ($10^6$). From Fig. 1 we see that the importance function for $\epsilon = 0.10$ poorly reflects the relevant stationary phase regions. The $\epsilon = 0.10$ importance function has appreciable density in the non-stationary phase regions, regions whose unimportance must then be retroactively established by the use of more sample points. Thus, while Table I tells us that the calculated $(x)_\lambda$ values are, in principle, independent of the choice of $\epsilon$, the results of Fig. 4(a) tell us that in practice a statistical estimate made using a fixed number Monte Carlo points is dependent upon the quality of the associated importance function. In the present case, $10^6$ Monte Carlo points are insufficient to produce all the phase cancellations necessary to overcome the qualitatively incorrect importance function associated with $\epsilon = 0.10$. This conclusion is reinforced by the brown, orange, blue, and green curves of Fig. 4(a). These curves denote the real portions of the $(x)_\lambda$ results computed for $\epsilon = 0.10$ for $10^4$, $10^5$, $10^6$, and $10^7$ Monte Carlo points, respectively. We see that the onset of systematic errors in these $(x)_\lambda$ results correlates with the number of Monte Carlo points used in the corresponding simulation. As more points are used, more of the deficiencies of the underlying importance functions are
overcome and the \( \langle x \rangle_\lambda \) values are computed reliably for larger negative \( \lambda \)-values. As illustrated by the black and green curves in Fig. 4(a), however, improving the underlying importance function is generally a more efficient option than the brute-force approach. Finally, it is important to note that while they individually may have computational shortcomings, the small \( \varepsilon \)-values of the ensemble actually play a critical role in the INS approach. In particular, they provide the “connective tissue” that bridges the otherwise sparsely connected densities associated with other control parameters. In general, the level of agreement between results computed for different ensemble control parameters serves as a practical internal quality control indicator for the overall simulation.

Figures 5 and 6 show the real and imaginary parts of \( \phi(\eta) \) computed from Eq. (4.3) using the \( \varepsilon = 0.40 \) \( \langle x \rangle_\lambda \) results of the type shown in Fig. 3. These results illustrate the variation of the \( \phi(\eta) \) results for different \( x_0 \) values (0.25 = black, 0.50 = red, 1.00 = green) for a fixed value of \( \sigma \) (1.00 = black, 2.00 = red, 3.00 = green). In general, the results of the type in Figs. 5 and 6 are accurate over the \( \eta \)-range for which the corresponding \( \langle x \rangle_\lambda \) results are \( \varepsilon \)-independent.

Figure 7 documents the ability of the present approach to compute \( \phi(\eta) \) accurately for large, negative \( \eta \)-values, regions hard to treat with direct Monte Carlo methods. For the choice of \( x_0 = 0.50 \) and \( \varepsilon = 0.40 \), the red curve in Fig. 7 shows \( \text{Re}(\phi(\eta)) \) obtained using the present approach, while the black curve shows the corresponding results obtained using direct Monte Carlo methods with the same number of points \( (10^6) \). The corresponding results for \( \text{Im}(\phi(\eta)) \) are of similar quality.

Finally, Figs. 8 and 9 display the real and imaginary portions of \( \phi(\eta) \) computed for a fixed value of \( x_0 \) (0.5) and varying values of \( \sigma \) (1.00 = black, 2.00 = red, 3.00 = green) using the present approach. All results utilize \( 10^6 \) Monte Carlo points for the evaluation of the necessary \( \langle x \rangle_\lambda \) results.

V. SUMMARY AND FUTURE DIRECTIONS

In the present work, we have explored the problem of calculating generalized averages. We have presented an approach that combines thermodynamic integration and stationary phase Monte Carlo techniques to cope with the risk-sensitive and rare-event sampling issues involved and have explored its application to a prototypical class of problems.

We close by noting that the methods developed in the present work may offer a potential tool for the study of real-time quantum dynamics. In particular, we note that a generic, coordinate-space, equilibrium time correlation function, \( G_{AB}(t) \), can be expressed as

\[
G_{AB}(t) = \frac{\int dx dx' \rho(x) P(x \rightarrow x', t) A(x) B(x')}{{\int dx dx' \rho(x) P(x \rightarrow x', t)}}, \tag{5.1}
\]

where \( \rho(x) \) is proportional to the probability of \( x \) and \( P(x \rightarrow x', t) \) is the conditional probability density that the system that starts at \( x \) at time zero arrives at \( x' \) a time \( t \) later. If one can sample these “initial” and “final” conditions, \( G_{AB}(t) \) can be estimated as

\[
G_{AB}(t) \approx \frac{1}{N} \sum_{n=1}^{N} A(x_n) B(x'_n), \tag{5.2}
\]

where the points \( \{x_n\} \) are a random sample of \( \rho(x) \) and the points \( \{x'_n\} \) are a random sample of the conditional probability \( P(x_n \rightarrow x'_n, t) \).

Generating a sampling of the initial positions, \( \{x_n\} \), is a standard equilibrium problem, one for which well established classical and quantum-mechanical approaches exist. Techniques for sampling the conditional probability involved are well established for classical systems, but generally lacking for quantum-mechanical ones.
The minimal information needed to sample the conditional probability, \( P(x \rightarrow x', t) \), with respect to the final position, \( x' \), is knowledge of ratios of the form

\[
|R|^2 = \frac{P(x \rightarrow x', t)}{P(x \rightarrow x'', t)}. \tag{5.3}
\]

As noted previously, \( R \) for finite temperature quantum-mechanical problems can be expressed in the Kirkwood form. The combination of such Kirkwood expressions for \( R \), the current INS/SPMC approach for their evaluation, and a method such as the existing penalty Monte Carlo technique for sampling from a noisy distribution may offer a possible approach to the conditional, quantum-dynamical sampling problem. Time and future investigations will reveal if this is so.

**ACKNOWLEDGMENTS**

The authors wish to acknowledge support from the National Science Foundation Award No. DMS-1317199 and from the DARPA EQUiPS Award No. W911NF-15-2-0122.

**APPENDIX A: OPTIMAL IMPORTANCE SAMPLING**

The sampling error of a Monte Carlo estimate of an average of the form

\[
\langle f \rangle_S = \int \exp(-\tilde{S}(x)) f(x) dx, \tag{A1}
\]

where \( \exp(-\tilde{S}(x)) \) is a conventional, real-valued normalized density that depends on the variance of \( f(x) \) about its average value. As is well known, \( ^1 \) by introducing a different density, \( W(x) \), in the design and construction of such estimates it is often possible to reduce the variance involved. It has been shown, \( ^1,18 \) that the “optimal” variance reduction is achieved by choosing the density \( W_{opt}(x) \) given by

\[
W_{opt}(x) = \frac{\exp(-\tilde{S}(x)) |f(x)|}{\int \exp(-\tilde{S}(x)) |f(x)| dx}, \tag{A2}
\]

and that the associated variance is given by

\[
\text{var}(f)_{W_{opt}} = \left( \int \exp(-\tilde{S}(x)) |f(x)| dx \right)^2 - \left( \int \exp(-\tilde{S}(x)) f(x) dx \right)^2. \tag{A3}
\]

If \( f(x) \) is everywhere strictly non-negative or non-positive, then the variance is reduced to zero. More generally, information about the integrand can be used to guide the selection of the importance function and thereby to achieve a variance reduction. Recent work by Ceriotti et al. has examined the limitations and weakness of reweighting approaches. \( ^19 \)

We now examine the choice of an optimal importance function in situations involving complex averages. The SPMC form of Eq. (A1) is (cf. Eqs. (A1) and (3.1))

\[
\langle f \rangle_S = \left\langle \exp(-\tilde{S}(x)) f(x) \right\rangle_x dx. \tag{A4}
\]

Mimicking the sequence of steps taken in the purely real case that lead to Eqs. (A2) and (A3), \( ^1,18 \) we seek the density, \( W(x) \), that minimizes

\[
\text{var}(f)_W = \int W(x) \left| \frac{\exp(-\tilde{S}(x)) f(x)}{W(x)} - \langle f \rangle_S \right|^2 dx \tag{A5}
\]

subject to the constraint that \( W(x) \) is normalized. The resulting minimization reveals that the optimal importance function is given by

\[
W_{opt}(x) = \left\langle \left| \frac{\exp(-\tilde{S}(x)) f(x)}{W(x)} \right| \right\rangle_x dx \tag{A6}
\]

and the associated variance is

\[
\text{var}(f)_{W_{opt}} = \left( \int \left| \frac{\exp(-\tilde{S}(x)) f(x)}{W(x)} \right| dx \right)^2 - \left( \int \frac{\exp(-\tilde{S}(x)) f(x)}{W(x)} dx \right)^2. \tag{A7}
\]

**APPENDIX B: BRIEF SUMMARY OF INFINITE SWAPPING METHOD**

In order to make the present discussion somewhat more self-contained, we include a brief outline of the infinite swapping (INS) approach. Details of the formal development of the method as well as its implementation and sample applications can be found elsewhere. \( ^8–10,20–22 \)

In the notation of Section III, we consider a computational ensemble corresponding to \( K \) values of a designated control parameter (e.g., temperature or length scale). If the density for coordinate(s), \( x_1 \), and control parameter(s), \( \varepsilon_1 \), is denoted \( \pi(x_1, \varepsilon_1) \), for coordinate(s), \( x_2 \), and control parameter(s), \( \varepsilon_2 \), by \( \pi(x_2, \varepsilon_2) \), etc., then the fully symmetrized density for the ensemble, \( \mu(x, \varepsilon) \), is given by

\[
\mu(x, \varepsilon) = \sum_P P \left[ \pi(x_1, \varepsilon_1) \pi(x_2, \varepsilon_2) \ldots \pi(x_K, \varepsilon_K) \right], \tag{B1}
\]

where the sum over \( P \) denotes the sum over all possible permutations of coordinates and control parameters. Estimates of properties corresponding to any of the individual control parameters in the computational ensemble can be obtained from a sampling of \( \mu(x, \{ \varepsilon_k \}) \). Although nominally more complicated than its individual components, the ensemble density in Eq. (B1) tends to be more highly connected in sparse sampling situations (cf. Section II of Ref. 8) and thus more easily sampled.

In the present studies, the number of control parameters in the computational ensemble is relatively small. Consequently, the ensemble density in Eq. (B1) can be sampled directly using conventional methods such as ordinary Metropolis procedures or sum-sampling techniques. \( ^10 \) For larger computational ensembles, the factorial growth of the number of permutations limits approaches based on the direct use of the full ensemble density and other methods are required.

Generally applicable techniques have been developed and documented for dealing with applications beyond the reach of simple, brute force methods. One approach, partial infinite swapping (PINS), is discussed in detail in Appendix A of Ref. 8. In this method, the computational ensemble is
partitioned into non-overlapping “blocks” of control parameters individually small enough so that complete symmetrization within each block is feasible. This partitioning is then performed in two distinct ways producing two, partially symmetrized distributions. Individually, the two resulting forms are consistent with multiple invariant distributions, namely, those that have the proper relative weights of the permutations within the various symmetrized blocks. There is, however, a single, unique distribution shared by the two densities, the fully symmetrized distribution, Eq. (B1). By combining a complete symmetrization within the blocks of the two separate densities with a suitably designed exchange of information between them, it is possible to devise a practical procedure that samples the original, fully symmetrized distribution without ever explicitly evaluating it. While neither partially symmetrized density alone can accomplish this task, the two working in tandem can.