Journal of Statistical Physics, Vol. 119, Nos. 5/6, June 2005 (© 2005)
DOI: 10.1007/s10955-005-3770-1

Accelerated Monte Carlo for Optimal Estimation of Time Series

Francis J. Alexander,[1] Gregory L. Eyink,[2] and Juan M. Restrepo[3]

Received September 1, 2004; accepted January 31, 2005

By casting stochastic optimal estimation of time series in path integral form, one can apply analytical and computational techniques of equilibrium statistical mechanics. In particular, one can use standard or accelerated Monte Carlo methods for smoothing, filtering and/or prediction. Here we demonstrate the applicability and efficiency of generalized (nonlocal) hybrid Monte Carlo and multigrid methods applied to optimal estimation, specifically smoothing. We test these methods on a stochastic diffusion dynamics in a bistable potential. This particular problem has been chosen to illustrate the speedup due to the nonlocal sampling technique, and because there is an available optimal solution which can be used to validate the solution via the hybrid Monte Carlo strategy. In addition to showing that the nonlocal hybrid Monte Carlo is statistically accurate, we demonstrate a significant speedup compared with other strategies, thus making it a practical alternative to smoothing/filtering and data assimilation on problems with state vectors of fairly large dimensions, as well as a large total number of time steps.

KEY WORDS: path integral; stochastic processes; time series; hybrid Monte Carlo.

1. INTRODUCTION

Simplicity and generality make Markov-chain Monte Carlo (MC) one of the most powerful and most popular approaches to problems of model,

[1]Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA; e-mail: fja@lanl.gov
[2]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA.
[3]Mathematics Department and Physics Department, University of Arizona, Tucson, AZ, 85721, USA.

parameter, and state estimation. As a result, MC has been applied across a wide range of disciplines for optimal estimation, including finance,[1] climate dynamics,[2] oceanography,[3] meteorology[4] and biometrics.[5,6] See refs. 7–10 for comprehensive reviews.

While powerful, MC does have its drawbacks. In particular, an often slow statistical convergence rate can lead to unacceptably high computational costs, especially for large-scale, spatially distributed systems. Considerable effort has therefore gone into accelerating its statistical convergence.[8,10,11] Ideally, one would like every step in the Markov Chain to yield a new configuration, uncorrelated (as much as possible) from the current one. However, it is difficult to generate new, statistically uncorrelated configurations while at the same time keeping the acceptance rate within reasonable limits.

In this paper we apply a technique based on the hybrid MC method (HMC)[12,13] which distinguishes itself by nonlocal state updates. The method, sometimes referred to as generalized hybrid Monte Carlo (GHMC), was developed by Toral and Ferreira[14] and primarily applied to condensed matter physics, lattice gauge theory and optimization problems in refs. 14–17. Here we describe how it can be effectively used in problems involving optimal estimation.

Another means to achieve speedup is a unigrid-based variant (UMC).[10] Its simplicity, particularly with regard to modifying existing codes, is very attractive. In this study, we take the opportunity to also investigate the relative speedup due to UMC.

The specific estimation problem we consider here is the determination of the statistics of a time series $\mathbf{x}(t)$ taking values in an N-dimensional state space, given incomplete and possibly imprecise observations, $\mathbf{y}(t)$, of that system (see the following refs. 8, 18–21).

The state vector $\mathbf{x}$ is assumed to satisfy a known stochastic process

$$\begin{aligned} \mathbf{dx}(t) &= \mathbf{f}(\mathbf{x}(t), t)dt + (2D)^{1/2}(\mathbf{x}, t)\,\mathbf{dW}(t), \quad t > t_0, \\ \mathbf{x}(t_0) &= \mathbf{x_0}. \end{aligned} \tag{1}$$

For the initial state $\mathbf{x_0}$, either its value or probability distribution is assumed. The deterministic part of the dynamics is given by $\mathbf{f}$. Stochasticity might be inherent in the system dynamics and/or might arise from parametrizations of unknown or unresolved physics, or from ignored degrees of freedom in the dynamics. This is captured by the last term in which the diffusion matrix D acts on a vector-valued, standard Wiener process $\mathbf{W}$. The optimal estimate of the state history is then obtained by "assimilating" observations into the history of the state space statistics, i.e. by conditioning the statistics of the time series on those of the observations.

For simplcity, we consider observations at discrete times which are denoted by the N_y-dimensional vector,

$$\mathbf{y}(t_m) \equiv \mathbf{y}_m, \quad m = 1, 2, \ldots, M_{\text{obs}},$$

where m labels each observation. We assume that these observations are given by

$$\mathbf{y}(t_m) = \mathbf{h}(\mathbf{x}_m) + \boldsymbol{\epsilon}_m, \tag{2}$$

where $\mathbf{h}: \mathbf{R}^N \to \mathbf{R}^{N_y}$, and $\boldsymbol{\epsilon}_m$ is an N_y-dimensional noise vector with a known statistical distribution. We assume that the model noise and measurement errors are uncorrelated and that the initial value of $\mathbf{x}_0$ is a random variable with a known distribution.

The mean history of the state $\mathbf{x}$, conditioned on the measurements,

$$\mathbf{x}_S(t) = E[\mathbf{x}(t)|\mathbf{y}_1, \ldots, \mathbf{y}_M] \tag{3}$$

is the "best estimate" of the state, and the conditional covariance matrix

$$\mathbf{C}_S(t) = E[(\mathbf{x}(t) - \mathbf{x}_S(t))(\mathbf{x}(t) - \mathbf{x}_S(t))^\top | \mathbf{y}_1, \ldots, \mathbf{y}_M]$$

quantifies its uncertainty ($\top$ denotes transpose). The conditional mean $\mathbf{x}_S(t)$ is the quantity which minimizes

$$\text{tr}\mathbf{C}_S(t) = E[|(\mathbf{x}(t) - \mathbf{x}_S(t))|^2 | \mathbf{y}_1, \ldots, \mathbf{y}_M].$$

It is termed the *smoother* estimate.

Stratonovich,[22] Kushner[23,24] and Pardoux[25] constructed a sequential algorithm to solve the above nonlinear estimation problem. We will denote their methodology by KSP. It is reviewed briefly in ref. 26. Unfortunately, the KSP technique cannot be practically implemented except in systems with only a few degrees of freedom. This renders the KSP solution impractical for spatially distributed systems as found in climate/meteorology, hydrology, economics, etc. However, the KSP solution, when implemented on a specific estimation problem, can serve as an ideal benchmark with which to compare the accuracy of alternative methods. It is this way that the KSP will be used here.

2. PROBLEM FORMULATION

We discretize (1) using an explicit Euler–Maruyama scheme.[27] This leads to the following equation:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k + \mathbf{f}(\mathbf{x}_k, t_k)\delta t + (2D)^{1/2}(\mathbf{x}_k, t_k)(\mathbf{W}(t_k+\delta t) - \mathbf{W}(t_k)),\\ &\qquad k = 0, 1, 2, \ldots \\ \mathbf{x}_{k=0} &= \mathbf{x_0}.\end{aligned} \tag{4}$$

The choice of the time discretization is not unique. While we use the standard Euler–Maruyama discretization method here, the acceleration methods that follow can be applied to other discretizations as well.[28–30] The impact of using different time-discretizations is, in fact, an open problem.

The path integral representation assigns weights or probabilities to individual histories. These weights depend on both the stochastic dynamics and the measurements. Roughly speaking, in the absence of measurements, histories unlikely to arise from the dynamics are given a lower weight than histories which are consistent with them. Similarly, when the dynamics are ignored, histories which are far from the measurements are given lower weight than those closer to the measurements. A competition between the noise in the dynamics and the errors in the measurements dictates just how each effect will contribute.

Without loss of generality we will assume the discrete time steps to be equally spaced, and further, that the measurement times are commensurate with δt, the time step interval. Namely, we define $t_k = t_0 + k\delta t$ with $k = 0, 1, \ldots, T$, and $(t_f - t_0)/(T+1) = \delta t$.

The probability of the dynamics generating a given history is simply related to the probability that it experiences a certain noise history

$$\boldsymbol{\eta}(t_k) = (\mathbf{W}(t_k+\delta t) - \mathbf{W}(t_k)), \tag{5}$$

at times t_k, where $k = 0, 1, 2, \ldots, T$ For Gaussian, uncorrelated noise this probability $Prob\{\boldsymbol{\eta}(t), t = t_0, t_1, \ldots, t_T\} \sim \exp(-\frac{1}{2}\sum_k |\boldsymbol{\eta}|^2(t_k))$. Rearranging terms in (5), we can show that the probability of a given history is given by $Prob\{\boldsymbol{\eta}(t), t = t_0, t_1, \ldots, t_T\} \sim \exp(-H_{\text{dynamics}})$, where

$$\begin{aligned}H_{\text{dynamics}} \equiv \sum_{k=0}^{T-1} \frac{\delta t}{4} \{&[(\mathbf{x}_{k+1} - \mathbf{x}_k)/\delta t - \mathbf{f}(\mathbf{x}_k, t_k)]^\top D^{-1}(\mathbf{x}_k, t_k)\\ &\times[(\mathbf{x}_{k+1} - \mathbf{x}_k)/\delta t - \mathbf{f}(\mathbf{x}_k, t_k)]\}.\end{aligned} \tag{6}$$

Here $\top$ means transpose.

Thus far we have only considered the effects of the dynamics. To include the influence of observations, we use Bayes' rule. This modifies the Hamiltonian above by the addition of the following term:

$$H_{\text{obs}} = \frac{1}{2} \sum_{m=1}^{M_{\text{obs}}} [\mathbf{h}(\mathbf{x}(t_m)) - \mathbf{y}(t_m)]^{\top} R^{-1} [\mathbf{h}(\mathbf{x}(t_m)) - \mathbf{y}(t_m)]. \tag{7}$$

Here R is the covariance matrix of the observations. This contribution to the Hamiltonian corresponds to adding a local field or pinning term which, when the measurements are accurate, forces the state variable to be close that of the observation. Here we assume that the errors are Gaussian and uncorrelated with each other and uncorrelated with the state of the system. However, this formalism can, in principle, handle more general error statistics. The total Hamiltonian, including dynamics and observations is then

$$H = H_{\text{dynamics}} + H_{\text{obs}}. \tag{8}$$

The statistics of histories of the time-dependent, stochastic dynamical system have been cast as an equilibrium statistical mechanical system with Hamiltonian (8). As a result, we can apply a host of techniques from equilibrium statistical mechanics to this problem. In particular acceleration methods to sample Gibbs distributions. In addition to being able to handle a variety of error and noise terms, this formalism can also handle nonlinear dynamics.

Rather than sampling the Gibbs distribution, maximum likelihood methods minimize the Hamiltonian (Eq. (8)). The Hamiltonian is in fact the log-likelihood.[31] Maximum likelihood methods determine the mean, the median, or the mode of a distribution, while path-integral methods can be used to determine the mean and other moment statistics. For example, there is a very close relationship between the approach we take here and variational/adjoint, or so-called 4D-var methods, used extensively in the ocean/climate community (see ref. 32). In particular, the spacetime Hamiltonian we construct is identical to the standard cost function for weakly constrained 4D-var.

The known disadvantage of MC-based methods is, however, their slow convergence. Hence, if an MC-based strategy is available with a significant increase in computational efficiency it will then make it a viable methodology for large N-dimensional state estimation problems, whether

these have Gaussian or non-Gaussian statistics and whether they obey linear or nonlinear dynamics. A speed-up is achieved here by reformulating the Gibbs distribution sampling strategy. We thus turn to the sampling problem, showing that a generalized Hybrid Monte Carlo (GHMC) strategy is a fast sampling approach, thus making estimation via hybrid Monte Carlo viable in a large class of problems of physical interest.

3. SAMPLING

Having constructed the Gibbs distribution on histories, we now address its sampling. In order to show the efficiency of the GHMC technique, we will compare it to a standard HMC and a unigrid-based alternative (UMC).

3.1. Hybrid Monte Carlo HMC

Hybrid Monte Carlo was developed to address sampling deficiencies in lattice gauge theory computations in the mid 1980s.[12] Since then it has been applied to a number of physics problems, mostly in statistical mechanics, condensed matter and lattice gauge theory. Neal[13] has written an introductory survey that shows how the technique has been used in other applied probability problems.

HMC derives its name from the fact that it involves a mixture of Hamiltonian molecular dynamics and Metropolis-Hastings Monte Carlo. A time-discretized integration of the molecular dynamics equations is used to propose a new configuration. This configuration is then accepted or rejected by the standard Metropolis-Hastings Monte Carlo criteria. The acceptance/rejection criteria is the change in total energy.

As with many other MC algorithms, HMC in its original formulation was a local updating scheme. Specifically, the proposed change for the value of the field at a point depends only on the current configuration of the field at points near it. For many systems (critical phenomena, for example) this leads to unacceptably slow statistical sampling. Acceleration has been achieved using multigrid,[33] cluster algorithms,[34,35] and Fourier Langevin,[36–38] in statistical mechanics and field theory problems; variants of these have had some impact on state estimation.[10]

Consider a system whose configuration is specified by $T+1$ degrees of freedom (assume each of mass 1) $\mathbf{q}_0, \mathbf{q}_1, \ldots, \mathbf{q}_T$. Here the dimension of each $\mathbf{q}_i$ is N, the state variable dimension. The HMC algorithm works as follows: To each of these variables or generalized coordinates $\mathbf{q}_i$, a conjugate generalized momemtum $\mathbf{p}_i$, is assigned. For our purposes H is the configurational Hamiltonian (8) – the action, or minus the log-likelihood.

The momenta $\mathbf{p}_i$ give rise to a kinetic energy $H_K = \sum_i \frac{|\mathbf{p}_i|^2}{2}$. The total Hamiltonian of the system is then

$$\hat{H} = H + H_K. \tag{9}$$

The standard Hamiltonian dynamics of such a system is

$$\begin{aligned} \frac{d\mathbf{q}_i}{d\tau} &= \mathbf{p}_i \\ \frac{d\mathbf{p}_i}{d\tau} &= \mathbf{F}_i, \end{aligned} \tag{10}$$

where $\mathbf{F}_i = -\frac{\partial H}{\partial \mathbf{q}_i}$ is the force on the ith degree of freedom, and τ is fictitious MC time. A chain of states is generated in the following way: The dynamics are discretized in τ, for example, as

$$\mathbf{q}_i' = \mathbf{q}_i + (\delta\tau)\mathbf{p}_i + (\delta\tau)^2\mathbf{F}_i([\mathbf{q}]) \tag{11}$$

and

$$\mathbf{p}_i' = \mathbf{p}_i + \frac{\delta\tau}{2}(\mathbf{F}[\mathbf{q}]_i + \mathbf{F}[\mathbf{q}']_i), \tag{12}$$

for $i = 0, 1, 2, \ldots, T$. The update from $(\mathbf{q}_i, \mathbf{p}_i)$ to $(\mathbf{q}_i', \mathbf{p}_i')$ will, in general, not conserve energy as a result of the time discretization. The extent to which energy is not conserved is controlled by the step size $\delta\tau$. Detailed balance is achieved if the configuration obtained after evolving J steps is accepted with probability $\min[1, \exp \Delta\hat{H}]$, where

$$\Delta\hat{H} = \hat{H}(\mathbf{q}', \mathbf{p}') - \hat{H}(\mathbf{q}, \mathbf{p}). \tag{13}$$

Thus, the Metropolis step corrects for time discretization errors. The momentum variables are refreshed after every acceptance/rejection stage according to a Gaussian distribution of independent variables $\exp(-H_K)$. The time-marching and acceptance/rejection process represents one *MC trial*.

Much like the Gibbs-sampler, this method allows the system to equilibrate locally, although it would take more than $J = 1$ steps to do so. Configurations with large correlation lengths and long correlation times however would require a significant number of steps to sample the equilibrium. This is analogous to *critical slowing down*.

3.2. Unigrid Monte Carlo UMC

Correlation times are typically longer for longer-wavelength modes. Recognizing this, researchers have developed MC methods to update these modes more efficiently. The multigrid approach (see ref. 33) is one particular strategy. We summarize this approach below and feature it in our sampling comparison. We will specifically implement and compare a simple variant multigrid often called "unigrid".[10]

The unigrid strategy consists of updating the system by taking coherent moves on a number of "length" scales. This is done by decomposing the system into blocks of contiguous points on the lattice, indexed by i. The block sizes range from 1 (i.e., the standard local Metropolis) to the full lattice length. For simplicity we have chosen the lattice to be of size 2^l where l is an integer. Then the blocks to be used are of size $1, 2, 4, \ldots, 2^l$. We assume, of course, that $\mod[T+1, 2]=0$. The update consists of the following: To each site of a block the local value has a random variable $\boldsymbol{\delta\phi}_i$ added to it. This random variable is selected from some specified distribution, e.g., Gaussian with $\langle \delta\phi_i \rangle = 0$, and variance $\langle (|\delta\phi_i|)^2 \rangle = \sigma^2$. The Hamiltonian is computed before and after the change and a Metropolis accept/reject as described previously is performed. The distribution of the proposed change $\boldsymbol{\delta\phi}$ is such that the Markov Chain is ergodic and that the acceptance rates are approximately 0.50.

3.3. Generalized Monte Carlo GHMC

Nonlocality in the sampling can lead to a reduction of the correlation times. Toral and Ferreira proposed a general nonlocal strategy that renders HMC far more efficient.[39] We describe their formulation here and propose that this strategy makes HMC a viable tool for fairly large state estimation problems.

The Hamiltonian dynamics given by (11) is replaced by a more general form:

$$\begin{aligned} \frac{d\mathbf{q}_i}{d\tau} &= \mathcal{A}_{ij}\mathbf{p}_j \\ \frac{d\mathbf{p}_i}{d\tau} &= [\mathcal{A}_{ij}]^\top \mathbf{F}_i, \end{aligned} \tag{14}$$

where $0 \leqslant i, j \leqslant T$ and repeated indices assume summation. Here $\mathcal{A}$ is a $(T+1)\times(T+1)$ matrix. When $\mathcal{A}$ is the identity matrix, the equations reduce to those associated with the standard HMC. The evolution equations in (15) conserve energy, i.e. $d\hat{H}/d\tau = 0$, however, the system may not be Hamiltonian. The discretization of (15) can be schematically expressed as

$$q' = q + \delta\tau \mathcal{A} p + \frac{(\delta\tau)^2}{2} A(A)^T F([q]) \tag{15}$$

and

$$p' = p + \frac{\delta\tau}{2} (\mathcal{A})^T (F[q] + F[q']).$$

The task is then to find a matrix $\mathcal{A}$ that leads to a significant reduction of the correlation time. One example of such a matrix, for a problem with state vector dimension $N = 1$, is the circulant matrix $\mathcal{A} = \text{circ}(1, \exp(-\alpha), \exp(-2\alpha), \ldots, \exp(-T\alpha))$ where the constant α can depend on T. We use this matrix below in the implementation of the GHMC in the nonlinear estimation problem. A particularly convenient characteristic of such a matrix is that the required matrix-vector products can be performed efficiently in $(T+1)\log(T+1)$ operations. This is an important consideration in the overall selection of $\mathcal{A}$, especially when T is large.

4. TEST PROBLEM

For the purposes of comparing to an exact solution we have chosen a problem for which the KSP solution can be calculated numerically (see ref. 26 for details on implementing the KSP solution). In particular, we will compute a statistical estimate of a scalar state variable ($N = 1$) which describes the position on the entire real line of a particle subject to a potential

$$U(x) = -2x^2 + x^4 \tag{16}$$

that has two stable fixed points, in this case at $x = 1$ and $x = -1$ and an unstable fixed point at $x = 0$. The particle is stochastically forced. The model is described by

$$\dot{x}(t) = f(x(t)) + \kappa \eta(t), \tag{17}$$

where

$$f(x) = -U'(x)$$

and $\eta(t)$ is white-noise, with zero mean and covariance $\langle \eta(t)\eta(t') \rangle = \kappa \delta(t - t')$ [see refs. 40, 41]. For specificity we set $\kappa = 0.5$ here. With this noise strength κ, the solution $x(t)$ of (17) fluctuates about the minima in either of the wells with rather long residence times (long with respect to the time it takes to cross the potential barrier) and then, occasionally, undergoes a series of large fluctuations which leads to a transition into the

other well. The steady-state probability distribution of the model, $P_s(x) \propto \exp\left(-\frac{2U(x)}{\kappa^2}\right)$ is bimodal with peaks at $x = \pm 1$, the two stable fixed points of the deterministic dynamics.

The observational data is constructed synthetically: the position of the particle is observed, on average, to occupy one well for a period of time and then transitions to the other one. The actual position of the particle has some degree of uncertainty. In particular, we take a measurement history where the particle is observed to be at $x = -1$ for the first five measurements at integer times and then makes a transition, being observed at $x = 1$ for the next five measurements. This represents a history where the particle has made a transition from one well to another between time $t = 5$ and $t = 6$. This history represents what was observed. The actual state of the system over time, however, is only imperfectly known from these observations. The observation errors are chosen independently at each measurement time, with mean zero and variance $R = 0.04$. Note that the standard deviation $\sqrt{R} \times 100\%$ represents the rms error expressed as a percentage of the size of the equilibrium states at ± 1.

We turn to assessing the effectiveness of the GHMC method at sampling the likelihood function. We will do so by comparing the results of this scheme to HMC and UMC estimation schemes on the double-well problem described above. When compared to a solution of the optimal estimation problem HMC, UMC and GHMC, correctly track the statistics, i.e., there was agreement between MC methods and these in turn agreed with KSP to within discretization and statistical errors. The UMC outcome and the KSP are shown in Fig. 1 (the HMC and GHMC would be identical to the UMC since these cases differ only in the sampling strategy).

To assess the computational cost, however, we need only consider a time-periodic system and no observations. The parameters used in the simulations are specified in the Table I. This table displays the number of effective full-lattice sweeps required for decorrelation for each of the methods as a function of the lattice points $T + 1$. The number of lattice points is the number of Euler-Maruyama time steps.

In what follows we assume that $\mod(T + 1, 2) = 0$. We define a *lattice sweep* as the computational effort involved in finding $(\{\mathbf{Q}\}, \{\mathbf{P}\})'_b$ from $(\{\mathbf{Q}\}, \{\mathbf{P}\})_b$, where $\{\cdot\}_b$ represent blocks of conjugate pairs. Each block of 2^l members, where $0 \leqslant l \leqslant \log_2(T + 1)$, is indexed by $1 \leqslant b \leqslant (T + 1)/2^l$. For HMC and GHMC $(\{\mathbf{Q}\}, \{\mathbf{P}\})_b = (\mathbf{q}, \mathbf{p})$, the blocks are of size 1 and there are $T + 1$ blocks. For UMC, on the other hand, the blocks are of size 2^l and sweeps are required till all block sizes are exhausted. In the case of HMC and GHMC the update might be done J times. So we define then

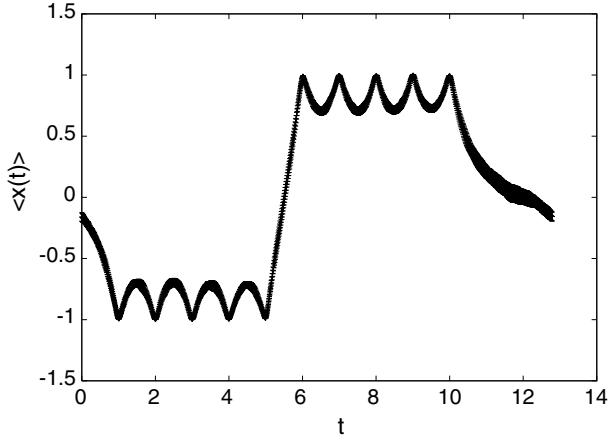

Fig. 1. Comparison between the optimal history, as given by KSP (dashed), and UMC (solid) for the test problem.

Table I. Correlation times in terms of effective lattice sweeps for hybrid Monte Carlo (HMC), unigrid hybrid Monte Carlo (UMC), and generalized hybrid Monte Carlo (GHMC), as a function of the number of Euler-Maruyama time steps

$T+1$	HMC ($J=1$)	HMC ($J=8$)	UMC	GHMC ($J=1$)
8	900(125)	170(7)	800(40)	40(8) [0.20]
16	5300(1600)	560(20)	1040(60)	60(10) [0.10]
32	13300(8300)	2700(140)	1430(100)	200(30) [0.05]
64	30000(7800)	2800(400)	1570(100)	420(70) [0.0245]

In parenthesis we state the standard deviation. J is the number of molecular dynamics integration steps per Metropolis accept/reject. The GHMC column quotes the value of α in the circulant matrix $\mathcal{A}$ used in the calculations in square brackets.

the *effective lattice sweep* as being J times the number of lattice sweeps and report the effective sweep below.

In addition to the cost of the updates GHMC has an added cost related to the matrix-vector multiplication by the matrix $\mathcal{A}$. Hence, when we report the cost of performing GHMC we include this extra work in reporting the effective lattice sweep.

Table I was generated as follows: The MC algorithms were run until the system reached stationarity. The criteria used in establishing stationarity was the same in all methods. We have chosen to report the

correlation times of the sum of the variable x over the total lattice, i.e., $S(t) = \sum_{i=0}^{T} x_i(t)$. The correlation time was found by making an exponential fit to the autocorrelation function, $\langle S(t)S(t')\rangle$. The table makes it clear that a significant speedup is achieved using the particular GHMC implementation described in Section 3.3, when compared to both HMC and the UMC method described earlier. For HMC with $J = 8$, there are 8 lattice sweeps per full HMC step (this is simply a reflection of having $J = 8$ molecular dynamics substeps for each MC accept/reject). Clearly, for $J = 1$ there is only 1 lattice sweep, resulting from only 1 MD integration per MC step.

As mentioned above for GHMC there is the cost of performing the matrix-vector multiplication per step. For the circulant matrix $\mathcal{A}$ we chose, this multiplies the cost by a factor of $\log(T+1)$ (The cost of a matrix-vector multiplication for a circulant matrix is $(T+1)\log(T+1)$, but the linear factor of $T+1$ is already accounted for.

With regard to storage, all three methods require similar amounts of computer resources. If only optimal state estimates are desired, these methods do not require storage of the covariance matrix or its history. All that must be stored is the current history.

5. DISCUSSION AND CONCLUSIONS

The class of stochastic processes covered by (1) is quite large: finite-dimensional approximations to stochastic partial differential equations, stochastic maps, as well as coupled discrete event systems and thus we envision that HMC could be of great utility in a variety of state estimation problems due to the speedup achieved by the generalized sampling strategy: the GHMC strategy could be applied to problems with fairly sizable (space-time) state vectors, say up to 10^7. This estimate is based on the fact that current lattice quantum chromodynamics calculations use HMC for lattices of this size.

Moreover, path-integral MC methods are capable of handling estimation problems for non-Markovian processes by using a two-time axis approach. However, there have been few attempts at applying MC techniques to these kinds of problems due to their excessive computational cost (for most real-world problems, anyway). We think that the GHMC savings on computational cost could make MC viable and applicable to large real-world problems of this sort. For linear or very mildly nonlinear dynamics with Gaussian statistics the conditioned statistics are fully captured by the mean and the variance. However, for strongly nonlinear problems it is often necessary to compute the higher moments. Path integral methods provide detailed information on representative histories,

actually yielding an ensemble of histories. Of the particle methods, some weighted ensemble methods also have this property.[42] There are other approaches that can succesfully handle highly nonlinear/far from Gaussian statistics. For example, the mean field approaches of Eyink *et al.* (see refs. 26, 41, 43), the variational approaches of Bennett and his collaborators (see refs. 44). However, a particularly attractive feature of both the path-integral method and particle methods,[21] is that they can be implemented as "black boxes", accepting whatever model (code) as input. In the other methods mentioned one is required to write fairly extensive auxilliary codes.

ACKNOWLEDGMENTS

We thank T. Bhattacharya, N. Gulbahce, G. Johnson, S. Mitter, and Y. Ch. Paschalidis for useful discussions. This work was carried out in part at Los Alamos National Laboratory under the auspices of the Department of Energy and supported by LDRD-ER 2000047. We also received support from NSF/ITR, Grant DMS-0113649 (JMR, GLE).

REFERENCES

1. B. Eraker, MCMC analysis of diffusion processes with application to finance. *J. Bus. Econ. Stati.*, **19-2**:177–191, (2001).
2. J. C. Hargreaves and J. D. Annan, Assimilation of paleo-data in a simple earth system model, *Clim. Dynam.* **19**:371–381 (2002).
3. C. Wunsch, *The Ocean Circulation Inverse Problem* (Cambridge University Press, Cambridge, UK, 1996).
4. E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, Cambridge, 2003).
5. J. Liu and C. Sabatti, Simulated sintering: MCMC with spaces of varying dimensions, *Bayesian Stat* **6**:386–413 (1998).
6. J. Liu and C. Sabatti, Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation, *Biometrika* **87**:353–369 (2000).
7. H. Sorensen, Parametric inference for diffusion processes observed at discrete points in time: A survey. *Internat. Statist. Rev.* **72**(3):337–354 (2004).
8. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter eds. *Markov Chain Monte Carlo in Practice*. (Chapman and Hall, New York, 1996).
9. Chen, *Bayesian filtering: From Kalman filters to particle filters and beyond*, McMaster University Technical Report, 2003.
10. J. S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer, New York, 2001).
11. M-H. Chen, Q-M. Shao, and J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, (Springer-Verlag, New York, Berlin, 2000).
12. B. Pendleton S. Duane, A. D. Kennedy, and D. Roweth, Hybrid Monte Carlo, *Phys. Lett. B.* **195**:216–222 (1987).
13. R. M. Neal, *Probabilistic inference using Markov chain Monte Carlo methods*, Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto, 1993.

14. R. Toral and A. L. Ferreira, *A general class of hybrid Monte Carlo methods.* (Proceedings of Physics Computing **94**, pp. 265–268, 1994).
15. A. D. Kennedy and B. Pendelton, Cost of the generalised hybrid Monte Carlo algorithm for free field theory, *Nucl. Phys. B*, **B607**:456–510 (2001).
16. R. Salazar and R. Toral, Simulated annealing using hybrid Monte Carlo, *J. Stat. Phys.*, **89**:1047–1060 (1997).
17. R. Salazar and R. Toral, Hybrid simulated annealing, *J. Stat. Phys.* **89**:1047 (1997).
18. A. Gelb. *Applied Optimal Estimation* (MIT Press, Cambridge, MA, 1974).
19. H. Tanizaki, *Nonlinear Filters: Estimation and Applications* (Springer-Verlag, New York, 1996).
20. P. Del Moral, *Feynman-Kac Formulae* (Springer-Verlag, New York, 2004).
21. A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice* (Springer-Verlag, New York, 2002).
22. R. L. Stratonovich, Conditional Markov processes, *Theor. Prob. Appl.* **5**:156–178 (1960).
23. H. J. Kushner, On the differential equations satisfied by conditional probability densities of Markov processes, with applications, *J. SIAM Control Ser.A* **2**:106–119 (1962).
24. H. J. Kushner, Dynamical equations for optimal nonlinear filtering, *J. Diff. Eq.* **3**:179–190 (1967).
25. E. Pardoux, Équations du filtrage non linéaire de la prédiction et du lissage, *Stochastics* **6**:193–231 (1982).
26. G. L. Eyink, J. M. Restrepo, and F. J. Alexander, A mean field approximation in data assimilation for nonlinear dynamics, *Physica D* **194**:347–368 (2004).
27. P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, 1992).
28. F. Langouche, D. Roeckaerts, and E. Tirapegui, On the most probable path for diffusion processes, *J. Phys. A* **11**:L263–L268 (1978).
29. R. Graham, Path integral formulation of general diffusion processes, *Zeitschrift fur Physik* **26**:281–290, (1977).
30. F. Langouche, D. Roeckaerts, and E. Tirapegui, Functional integral methods for stochastic fields, *Physica* **95A**:252–274 (1979).
31. Thomas A. Severini, *Likelihood Methods in Statistics* (Oxford University Press, New York, 2001).
32. O. Talagrand and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Quart. J. Roy. Meteor. Soc.* **113**:1311–1328 (1987).
33. J. Goodman and A. D. Sokal, Multigrid Monte Carlo method. Conceptual foundations, *Phys. Rev. D* **40**:2035–2071 (1989).
34. R. H. Swendsen and J. S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, *Phys. Rev. Lett.* **58**:86–88 (1987).
35. E. Domany and D. Kandel, General cluster Monte Carlo dynamics, *Phys. Rev. B* **43**:8539 (1991).
36. G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Langevin simulations of lattice field theories, *Phys. Rev. D* **32**:2736–2747 (1985).
37. F. J. Alexander, B. M. Boghosian, R. C. Brower, and S. R. Kimura, Fourier acceleration of langevin molecular dynamics, *Phys. Rev. E* 066704 (2001).
38. S. Caterall and S. Karamov, Testing a Fourier accelerated hybrid Monte Carlo algorithm, *Phys. Lett. B* **B528**:301–305 (2002).
39. R. Toral and A. Ferreira, Hybrid Monte Carlo method for conserved-order-parameter systems, *Phys. Rev. E* **47**:3848–3851 (1993).

40. R. N. Miller, M. Ghil, and P. Gauthiez, Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.* **51**:1037–1056 (1994).
41. G. L. Eyink and J. R. Restrepo, Most probable histories for nonlinear dynamics: Tracking climate transitions, *J. Stat. Phys.* **101**:459–472 (2000).
42. S. Godsill, A. Doucet, and M. West, Monte carlo smoothing for non-linear time series, (2001). URL: citeseer.ist.psu.edu/godsill01monte.html.
43. G. L. Eyink, J. M. Restrepo, and F. J. Alexander, Reducing computational complexity using closures in a mean field approach in data assimilation, *submitted* (2002).
44. A. Bennett, *Inverse Modeling of the Ocean and Atmosphere* (Cambridge University Press, Cambridge, UK, 2002).