GSHMC: An efficient Markov chain Monte Carlo sampling method

Sebastian Reich in collaboration with Elena Akhmatskaya (Fujitsu Laboratories Europe)

1. Motivation

In the first lecture, we started from a given differential equation with initial conditions distribute according to some probability density function (PDF) $\rho_0(\mathbf{x})$. Furthermore, we assumed that $\rho_0 = \mathrm{N}(\mathbf{x}_b, \mathbf{B})$ is Gaussian.

The ensemble Kalman filter requires that we generate an appropriate ensemble $\mathbf{x}_i$ of solutions with law ρ_0. How to do this in case ρ_0 is not Gaussian?

Furthermore, is it possible to generalize the Kalman analysis step to non-Gaussian PDFs? This leads us to consider particle filters.

In both cases, we may consider Monte Carlo methods to sample from a given PDF. How to do this efficiently?

2. Monte Carlo Methods

Given a probability density function (PDF) π over a configuration space $\mathbb{R}^n$, we are often interested in expectation values

$$\mathbb{E}[f] = \int_{\mathbb{R}^n} f(\mathbf{x})\, \pi(\mathbf{x})\, d\mathbf{x}$$

for a given function (observable) f. If $n \gg 1$, numerical quadrature can become prohibitively expensive and Monte Carlo methods are often the only alternative.

The basic Monte Carlo approach assumes that we can draw independent and identically distributed (i.i.d) random samples $\mathbf{x}_i$ from the PDF π. Then an approximation to $\mathbb{E}[f]$ can be obtained as

$$\bar{f}_N = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i).$$

How to generate samples from a given PDF π? Almost all methods rely on the assumption that we can draw samples from the uniform distribution U$[0,1]$ and/or the normal distribution N$(0,1)$.

If we give up the request that samples should be *independent*, then the much wider class of Markov chain Monte Carlo (MCMC) methods can be considered.

Under the assumption of geometric ergodicity we still obtain an $\mathcal{O}(N^{1/2})$ convergence of

$$\bar{f}_N = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i)$$

to the mean $\mathbb{E}[f]$, but the prefactor in $\mathcal{O}(N^{1/2})$ will be reduced from i.i.d. samples. On the other hand, generation of samples might be much easier.

3. Markov Chain Monte Carlo (MCMC) Methods

The following strategy allows one to produce statistically dependent samples from a given PDF in an elegant manner (Metropolis et al, 1953). It is based on the idea that one can construct a Markov process with the desired π as the (only) invariant PDF.

Let $A(\mathbf{y}|\mathbf{x})$ denote the transition rule of a Markov process, i.e., given a PDF ρ_i the next iterate is provided by

$$\rho_{i+1}(\mathbf{y}) = \int_{\mathbb{R}^n} \rho_i(\mathbf{x})\, A(\mathbf{y}|\mathbf{x})\, d\mathbf{x}.$$

and invariance of π amounts to

$$\pi(\mathbf{y}) = \int_{\mathbb{R}^n} \pi(\mathbf{x})\, A(\mathbf{y}|\mathbf{x})\, d\mathbf{x}.$$

A stronger assumption is that of detailed balance

$$\pi(\mathbf{x})\, A(\mathbf{y}|\mathbf{x}) = \pi(\mathbf{y})\, A(\mathbf{x}|\mathbf{y}).$$

Metropolis-Hastings algorithm. Given the current state $\mathbf{x}_i$:

- Draw $\mathbf{y}$ from the proposal distribution $T(\mathbf{y}|\mathbf{x}_i)$.

- Draw $u \sim U[0,1]$ and update

$$\mathbf{x}_{i+1} = \begin{cases} \mathbf{y}, & \text{if } u \leq r(\mathbf{y}, \mathbf{x}_i) \\ \mathbf{x}_i & \text{otherwise.} \end{cases}$$

Here

$$r(\mathbf{y}, \mathbf{x}) = \frac{\delta(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})\, T(\mathbf{y}|\mathbf{x})}$$

and $\delta(\mathbf{y}, \mathbf{x})$ is any symmetric function in $\mathbf{x}$ and $\mathbf{y}$ that makes $r(\mathbf{x}, \mathbf{y}) \leq 1$ for all $\mathbf{x}$, $\mathbf{y}$.

The actual transition probability from $\mathbf{x}$ to $\mathbf{y} \neq \mathbf{x}$ is given by

$$A(\mathbf{y}|\mathbf{x}) = T(\mathbf{y}|\mathbf{x})\, r(\mathbf{y},\mathbf{x}) = \pi(\mathbf{x})^{-1}\, \delta(\mathbf{y},\mathbf{x}).$$

Similarly,

$$A(\mathbf{x}|\mathbf{y}) = T(\mathbf{x}|\mathbf{y})\, r(\mathbf{x},\mathbf{y}) = \pi^{-1}(\mathbf{y})\, \delta(\mathbf{x},\mathbf{y}).$$

Since $\delta(\mathbf{x},\mathbf{y}) = \delta(\mathbf{y},\mathbf{x})$ the detailed balance condition

$$\pi(\mathbf{x})\, A(\mathbf{y}|\mathbf{x}) = \pi(\mathbf{y})\, A(\mathbf{x}|\mathbf{y})$$

follows.

The most popular choice for $\delta(\mathbf{y},\mathbf{x})$ is

$$\delta(\mathbf{y},\mathbf{x}) = \min(\pi(\mathbf{x})\, T(\mathbf{y}|\mathbf{x}), \pi(\mathbf{y})\, T(\mathbf{x}|\mathbf{y}))$$

leading to

$$r(\mathbf{y},\mathbf{x}) = \min\left(1, \frac{\pi(\mathbf{y})\, T(\mathbf{x}|\mathbf{y})}{\pi(\mathbf{x})\, T(\mathbf{y}|\mathbf{x})}\right).$$

A simple random walk MCMC algorithm is provided by the proposal step

$$\mathbf{y} = \mathbf{x}_i + \varepsilon_i$$

where ε_i are i.i.d. according to some given PDF $\rho(\mathbf{x})$. Provided that $\rho(\mathbf{x}) = \rho(-\mathbf{x})$, we obtain $T(\mathbf{y}|\mathbf{x}) = T(\mathbf{x}|\mathbf{y})$ and the acceptance function reduces to

$$r(\mathbf{y}, \mathbf{x}) = \min\left(1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right).$$

An efficient MCMC method should lead to a rapid decorrelation of the accepted samples $\{\mathbf{x}_i\}_{i=1}^N$.

4. Geometric ergodicity (Meyn & Tweedy)

Minorization condition. There is a compact subset C of phase space $\mathcal{X}$, a probability measure ν on $\mathcal{X}$ and a constant $\epsilon > 0$ such that

$$\int \mathbf{1}_\Omega(\mathbf{y})\, A(\mathbf{y}|\mathbf{x})\, \mathrm{d}V(\mathbf{y}) \geq \epsilon\nu(\Omega)$$

for all $\mathbf{x} \in C$ and all measurable sets $\Omega \subset \mathcal{X}$.

Drift condition. There is a scalar (Lyapunov) function W with $W(\mathbf{x}) \geq 1$ and $W(\mathbf{x}) \to \infty$ as $\|\mathbf{x}\| \to \infty$ and numbers $\alpha \in (0, 1)$, and $\beta \in [0, \infty)$ such that

$$\int W(\mathbf{y})\, A(\mathbf{y}|\mathbf{x})\, \mathrm{d}V(\mathbf{y}) \leq \alpha W(\mathbf{x}) + \beta \mathbf{1}_C(\mathbf{x})$$

for the MCMC method with transition density $A(\mathbf{y}|\mathbf{x})$.

Theorem. If a Markov chain transition density $A(\mathbf{y}|\mathbf{x})$ satisfies appropriate drift and minorization conditions, then there is a unique invariant PDF Π and

$$|\mathbb{E}_{\mathbf{x}_0}[f(\mathbf{x}_i)] - \mathbb{E}_{\Pi}[f]| \leq \kappa r^i W(\mathbf{x}_0)$$

for all f with $|f| \leq W$, where $r \in (0,1)$ and $\kappa \in (0,\infty)$.

See Meyn & Tweedy (1993) for geometric ergodicity; Rosenthal (1995,2002) for relatively elementary proofs using drift conditions; Roberts & Tweedie (1996), Mengersen & Tweedie (1996), Mattingly, Stuart and Higham (2001) for applications to MCMC methods and numerical solutions of SDEs.

A practical measure is provided by the integrated auto-correlation function $\tau_{int}(h)$ for a given observable $h(\mathbf{x})$.

Let $\sigma^2 = \text{var}[h]$ and compute

$$\rho_j = \text{corr}\{h(\mathbf{x}_i), h(\mathbf{x}_{i+j})\},$$

which becomes independent of i once the MCMC has equilibrated, i.e., all $\mathbf{x}_i$ are assumed to follow the law π. Then

$$\begin{aligned} N\text{var}\left\{\frac{h(\mathbf{x}_1)+\cdots+h(\mathbf{x}_N)}{N}\right\} &= \sigma^2\left[1+2\sum_{j=1}^{N-1}\left(1-\frac{j}{N}\right)\rho_j\right] \\ &\approx \sigma^2\left[1+2\sum_{j=1}^{N-1}\rho_j\right] =: \sigma^2\,\tau_{int}(h). \end{aligned}$$

Hence the variance in the MCMC estimator is equal to that of

$$\frac{N}{\tau_{int}(h)}$$

independent samples.

5. Dynamial systems motivated Markov chains

Given a PDF $\rho(\mathbf{q})$, we introduce the potential $V(\mathbf{q})$ through

$$C\,\exp(-V(\mathbf{q})) = \rho(\mathbf{q}) \quad \rightarrow \quad V(\mathbf{q}) = -\log(\rho(\mathbf{q})) + C,$$

and the "guide Hamiltonian"

$$\mathcal{H} = \frac{1}{2}\mathbf{p}\cdot\left[M^{-1}\mathbf{p}\right] + V(\mathbf{q}).$$

The "mass matrix" M can be used as a "preconditioner" to enhance sampling.

We introduce the state variable $\mathbf{x} = (\mathbf{q}^T, \mathbf{p}^T)^T$ and consider the associated Hamiltonian equations of motion

$$\dot{\mathbf{q}} = M^{-1}\mathbf{p}, \qquad \dot{\mathbf{p}} = -\nabla V(\mathbf{q}).$$

Properties:

(i) The canonical density

$$\pi(\mathbf{x}) \propto \exp(-\mathcal{H}(\mathbf{q},\mathbf{p})) \propto \exp(-\mathbf{p}^T M^{-1}\mathbf{p}/2) \times \rho(\mathbf{q})$$

is an invariant of the dynamics (but not the only one!). We will replace the task of sampling from $\rho(\mathbf{q})$ by sampling from $\pi(\mathbf{x})$ and marginalization (i.e., we just ignore the $\mathbf{p}$'s).

(ii) The equations of motion conserve volume and are time-reversible, i.e., the involution

$$\mathcal{F} : (\mathbf{q},\mathbf{p}) \to (\mathbf{q},-\mathbf{p})$$

reverses the direction of time.

Numerical implementation:

We will use the 2nd order Störmer-Verlet (SV) method

$$\begin{aligned}
\mathbf{p}^{n+1/2} &= \mathbf{p}^n - \frac{\Delta t}{2}\nabla V(\mathbf{q}^n), \\
\mathbf{q}^{n+1} &= \mathbf{q}^n + M^{-1}\mathbf{p}^{n+1/2}, \\
\mathbf{p}^{n+1} &= \mathbf{p}^{n+1/2} - \frac{\Delta t}{2}\nabla V(\mathbf{q}^{n+1})
\end{aligned}$$

The SV method conserves volume in phase space and is time-reversible.

But since the SV method *does not* conserve energy $\mathcal{H}$, the canonical distribution $\pi(\mathbf{x}) \propto \exp(-\mathcal{H}(\mathbf{x}))$ is not an invariant!

However, since the SV method is symplectic it conserves a modified energy $\mathcal{H}_{\Delta t}$ to arbitrarely high order!

More precisely, let us denote the numerical propagator by $\Psi_{\Delta t}$. Then

$$\mathcal{H}(\Psi_{\Delta t}(\mathbf{x})) - \mathcal{H}(\mathbf{x}) = \mathcal{O}(\Delta t^2)$$

for the SV method. But we can find modified energies $\mathcal{H}_{\Delta t}$ such that

$$\mathcal{H}_{\Delta t}(\Psi_{\Delta t}(\mathbf{x})) - \mathcal{H}_{\Delta t}(\mathbf{x}) = \mathcal{O}(\Delta t^p)$$

for p arbitrarely large as $\Delta t \to 0$ (Neishtadt, Benettin & Giorgilli, Hairer & Lubich, Reich).

Efficient methods for computing modified energies are available (Hardy & Skeel).

6. Generalized hybrid Monte Carlo method

A Markov chain will converge to some distribution of configurations if it is constructed out of Markov chain Monte Carlo (MCMC) updates each of which has the desired distribution as an invariant PDF, and which taken together are ergodic.

The generalized hybrid Monte Carlo (GHMC) algorithm of Horowitz (1991) and Kennedy & Pendleton (2001) for sampling from the canonical ensemble with density function

$$\pi(\mathbf{q}, \mathbf{p}) \propto \exp(-\mathcal{H}(\mathbf{q}, \mathbf{p})),$$

is defined as the concatenation of a classical mechanics Monte Carlo (CMMC) and a partial momentum refreshment Monte Carlo (PMMC) step.

Classical mechanics Monte Carlo (CMMC):

(i) *Classical mechanics* (CM): Solution of Hamilton's equations with a time-reversible and volume conserving method $\Psi_{\Delta t}$ (e.g., the SV method) over L steps and step-size Δt.

The resulting time-reversible and sympletic (and hence volume conserving) map from the initial to the final state is denoted by $U_\tau : (\mathbf{q}, \mathbf{p}) \to (\mathbf{q}', \mathbf{p}')$, $\tau = L\,\Delta t$.

(ii) A *momentum flip* $\mathcal{F} : (\mathbf{q}, \mathbf{p}) \to (\mathbf{q}, -\mathbf{p})$.

(iii) *Monte Carlo* (MC): a Metropolis accept/reject test

$$(\mathbf{q}', \mathbf{p}') = \begin{cases} \mathcal{F}U_\tau(\mathbf{q}, \mathbf{p}) & \text{with probability } \min(1, \exp(-\delta H)) \\ (\mathbf{q}, \mathbf{p}) & \text{otherwise} \end{cases},$$

with

$$\delta H := \mathcal{H}(U_\tau(\mathbf{q}, \mathbf{p})) - \mathcal{H}(\mathbf{q}, \mathbf{p}) = \mathcal{H}(\mathcal{F}U_\tau(\mathbf{q}, \mathbf{p})) - \mathcal{H}(\mathbf{q}, \mathbf{p}).$$

Partial momentum refreshment Monte Carlo (PMMC)

We first apply an extra momentum flip $\mathcal{F}$ so that the trajectory is reversed upon an CMMC rejection (instead of upon an acceptance). The momenta $\mathbf{p}$ are now mixed with a normal (Gaussian) i.i.d. distributed noise vector $\mathbf{u} \in \mathbb{R}^m$ and the complete partial momentum refreshment step is given by

$$\begin{pmatrix} \mathbf{u}' \\ \mathbf{p}' \end{pmatrix} = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ -\mathbf{p} \end{pmatrix} \tag{1}$$

where

$$\mathbf{u} = \mathcal{M}(\mathbf{q})^{1/2}\xi, \quad \xi = (\xi_1, \ldots, \xi_m)^T, \quad \xi_i \sim \mathrm{N}(0,1),$$

and $0 \leq \phi \leq \pi/2$. Here $\mathrm{N}(0,1)$ denotes the normal distribution with zero mean and unit variance.

Detailed balance of CMMC step:

Given two subsets A and B of phase space $\mathbb{R}^{2m}$, let π_{AB} denote the probability to go from A to B. Then

$$\begin{aligned}
\pi_{AB} &= \int_B \int_A A(\mathbf{x}'|\mathbf{x})\,\pi(\mathbf{x})\,d\mathbf{x}\,d\mathbf{x}' \\
&= \int_{\mathbb{R}^{2m}} \chi_A(\mathbf{x})\,\chi_B(\mathcal{F}U_\tau(\mathbf{x}))\,\min\left\{1, \frac{\pi(\mathcal{F}U_\tau(\mathbf{x}))}{\pi(\mathbf{x})}\right\}\pi(\mathbf{x})\,d\mathbf{x} \\
&= \int_{\mathbb{R}^{2m}} \chi_A(\mathbf{x})\,\chi_B(\mathcal{F}U_\tau(\mathbf{x}))\,\min\left\{\pi(\mathbf{x}), \pi(\mathcal{F}U_\tau(\mathbf{x}))\right\}d\mathbf{x} \\
&= \int_{\mathbb{R}^{2m}} \chi_A(\mathcal{F}U_\tau(\hat{\mathbf{x}}))\,\chi_B(\hat{\mathbf{x}})\,\min\left\{\pi(\mathcal{F}U_\tau(\hat{\mathbf{x}})), \pi(\hat{\mathbf{x}})\right\}d\hat{\mathbf{x}} \\
&= \pi_{BA}
\end{aligned}$$

where we used the substitution $\hat{\mathbf{x}} = \mathcal{F}U_\tau(\mathbf{x})$ (volume conserving!) and $(\mathcal{F}U_\tau)^2 = \mathrm{id}$!

Comments:

The GHMC method can be viewed as a Metropolis corrected time-stepping method for second-order Langevin dynamics provided $L = 1$ (number of time-steps per MC step) and $\phi = \sqrt{2\Delta t\gamma}$ (angle in the momentum refreshment step).

The special case $L = 1$ and $\phi = \pi/2$ leads to the hybrid Monte Carlo (HMC) method and can be viewed as a Metropolis corrected time-stepping method for Brownian dynamics with step-size $h = \sqrt{\Delta t}$.

Neither the CMMC nor the PMMC step on its own are geometrically ergodic. Roberts & Tweedie (1996) and Mengersen & Tweedie (1996) have discussed geometric ergodicity of HMC method. We are working on the extension of these results to the generalized hybrid MC method.

7. Analysis of acceptance rates

We now analyze the acceptance rate of the classical mechanics Monte Carlo step.

Step 1. We obtain

$$\begin{aligned}\langle \exp(-\delta H)\rangle &= \frac{1}{C}\int_{-\infty}^{\infty} \exp(-\delta H)\, \exp(-\mathcal{H})\, d\mathbf{z} \\ &= \frac{1}{C}\int_{-\infty}^{\infty} \exp(-\mathcal{H}\circ U_\tau)\, d\mathbf{x} = \frac{1}{C}\int_{-\infty}^{\infty} \exp(-\mathcal{H})\, d\mathbf{x}' = 1\end{aligned}$$

Step 2. Cumulants:

$$\begin{aligned}0 &= \log\left(\langle \exp(-\delta H)\rangle\right) \\ &= \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} = -\langle \delta H\rangle + \frac{1}{2}\langle (\delta H - \langle \delta H\rangle)^2\rangle + \ldots\end{aligned}$$

Hence

$$\langle \delta H\rangle \approx \frac{1}{2}\langle (\delta H - \langle \delta H\rangle)^2\rangle = \mathcal{O}(m\,\Delta t^{2p})$$

$p \geq 1$ the order of the method and m the number of DOFs.

Step 3. Hence we may assume that δH is N$(\sigma_0^2/2, \sigma_0^2)$ distributed with $\sigma_0^2 = \langle(\delta H - \langle\delta H\rangle)^2\rangle = 2\langle\delta H\rangle$. The average acceptance rate is determined by

$$\begin{aligned} \langle P_{acc}\rangle &= \frac{1}{\sqrt{2\pi}\sigma_0}\int_{-\infty}^{+\infty} \text{min}\,(1, \text{exp}(-x))\ \text{exp}\left(-\frac{(x-\sigma_0^2/2)^2}{2\sigma_0^2}\right) dx \\ &= \text{function of } \sigma_0^2 \end{aligned}$$

Step 4. To keep the average acceptance rate constant as the system size changes we have to keep the variance σ_0^2 constant, i.e.,

$$\Delta t^{2p}\, m = \text{constant.}$$

However, the CMMC step becomes increasingly computationally demanding as we either decrease Δt (i.e., increase the number of time-steps L for $\tau = L\Delta t = const.$) or increase the order p of the method. We got $p = 2$ for the Störmer-Verlet method.

8. Generalized shadow hybrid Monte Carlo methods

Our work has been motivated by the desire to use

$$\widetilde{\pi} \propto \exp(-\mathcal{H}_{\Delta t}),$$

where $\mathcal{H}_{\Delta t}$ is a modified/shadow energy computed by the methods described earlier.

This choice achieves an optimal acceptance rate P_{acc} in the CMMC part (arbitrarily close to one even as the system size m increases).

What about the momentum refreshment step? The GHMC methodology allows one to implement the partial momentum refreshment (PMMC) step as a Markov chain Monte Carlo method with invariant density $\widetilde{\pi}$.

In the partial momentum refreshment step, we consider $\mathcal{H}_{\Delta t}$ for fixed $\mathbf{q}$ as a function of $\mathbf{p}$ and introduce $\mathcal{H}_{\Delta t,\mathbf{q}}(\mathbf{p})$ as a short hand to emphasize this point.

The idea is to construct a standard hybrid Monte Carlo method in the "position" vector $\mathbf{p}$ and a set of "conjugate momenta" $\xi \in \mathbb{R}^m$. The required "Hamiltonian" is provided by

$$H(\mathbf{p}, \xi) = \mathcal{H}_{\Delta t,\mathbf{q}}(\mathbf{p}) + \frac{1}{2}\xi^T\xi$$

and defines the reference density

$$\widehat{\pi} \propto \exp(-H) \propto \exp(-\mathcal{H}_{\Delta t,\mathbf{q}}) \times \exp(-\xi^T\xi/2)$$

for the Metropolis acceptance criterion.

The generalized shadow HMC (GSHMC) method of Akhmatskaya & Reich (2005,2008) achieves a quasi m-independent acceptance rate in both the CMMC and PMMC steps for given fixed step-size Δt in the classical mechanics propagator $\Psi_{\Delta t}$.

GSHMC can be viewed as an importance sampling method using the modified ensemble

$$\widetilde{\pi} \propto \exp(-\mathcal{H}_{\Delta t}).$$

The computational overhead compared to GHMC consists in two evaluations of the modified energy $\mathcal{H}_{\Delta t}$ per time step (and hence in a small number of additional force evaluations).

Other forms of the momentum update under a modified Hamiltonian have been discussed by Izaguirre, Skeel Hampton and Sweet (2004,2009).

9. A nonlinear Kalman analysis step

A natural application of GSHMC is in Bayesian parameter estimation (and other statistical inference problems) (Neal, 1996).

In Bayesian estimation, we consider the unknown parameters $\mathbf{x}$ as random variables with some *a priori* distribution $\rho(\mathbf{x})$. The *a priori* information is combined with the measurement $\mathbf{y}$ through a conditional density function $\rho(\mathbf{y}|\mathbf{x})$.

In case of the Kalman filter, we had

$$\rho(\mathbf{x}) = \mathrm{N}(\mathbf{x}_f, \mathbf{P}_f)$$

and

$$\rho(\mathbf{y}|\mathbf{x}) = \mathrm{N}(\mathbf{y} - \mathbf{H}\mathbf{x}, \mathbf{R}).$$

Recall that Bayes' rule is given by

$$\rho(\mathbf{x}|\mathbf{y}) = \frac{\rho(\mathbf{y}|\mathbf{x})\,\rho(\mathbf{x})}{\rho(\mathbf{y})},$$

where

$$\rho(\mathbf{y}) = \int_{-\infty}^{+\infty} \rho(\mathbf{y}|\mathbf{x})\,\rho(\mathbf{x})\,d\mathbf{x}$$

is just a normalization constant since $\mathbf{y}$ is a set of known quantities.

If all involved PDFs are Gaussian then

$$\rho(\mathbf{x}|\mathbf{y}) = \mathsf{N}(\mathbf{x}_a, \mathbf{P}_a)$$

and Kalman's formulas follow.

For the general case, Monte Carlo methods can be used to generate samples $\{\mathbf{x}_i\}$ from the distribution $\rho(\mathbf{x}|\mathbf{y})$.

10. Summary

Sampling of a given PDF can be achieved by embedding into a classical mechanics system. This allows for proposal steps that lead "far away" from the current state (on a given energy level).

Time-stepping artifacts can be eliminated by applying a Metropolis acceptance criterion. The rejection rate increases with system size, however.

Acceptance rate can be increased by importance sampling with respect to a modified energy. The computational overhead is minor.

These methods could be used for nonlinear generalizations of Kalman filters (particle filters).

Geometric ergodicity might not hold unless the potential energy function V is globally Lipschitz and growths sufficiently fast at infinity. This can be cured by time-reversible variable step-size methods.

High energy barriers might also lead to a slow exploration of phase space. Here an artificial temperature can often help (simulated annealing / parallel tempering).

References:

[1] E. Akhmatskaya and S. Reich, GSHMC: An efficient method for molecular simulation. J. Comp. Phys. 227, 4934-4954, 2008.

[2] Chze Ling Wee, Mark S.P. Sansom, S. Reich and E. Akhmatskaya, Improved sampling for simulations of interfacial membrane proteins: Application of GSHMC to a peptide toxin/bilayer system. J. Phys. Chem. B. 112 , 5710-5717, 2008.

[3] B. Leimkuhler and S. Reich, A Metropolis adjusted Nose-Hoover thermostat. M2AN, in press.

[4] E. Akhmatskaya, N. Bour-Rabee and S. Reich, A comparison of generalized hybrid Monte Carlo methods with and without momentum flip. J. Comput. Phys. 228, 2256–2265, 2009.