Irreversible Langevin samplers and variance reduction: a large deviations approach

Konstantinos Spiliopoulos

Department of Mathematics & Statistics, Boston University
Partially supported by NSF-DMS 1312124 and career award NSF-DMS 1550918

Outline

Part I

Motivation: Accelerating Monte Carlo

Motivation

- Sampling from a given high dimensional distribution is a classical problem.
- One knows the target distributions only up to normalizing constants. Hence approximations are necessary.
- Often, such approximations are based on constructing Markov processes that have the target distribution as their target distribution, e.g., MCMC.

Motivation

- Sampling from a given high dimensional distribution is a classical problem.
- One knows the target distributions only up to normalizing constants. Hence approximations are necessary.
- Often, such approximations are based on constructing Markov processes that have the target distribution as their target distribution, e.g., MCMC.
- The degree of how good the approximation is depends on
  1. the approximating Markov process, and
  2. on the criterion used for comparison.

Problem formulation-Steady state simulation

Let us assume that target distribution is of Gibbs type

$$\bar{\pi}(dx) = \frac{e^{-2U(x)}dx}{\int_E e^{-2U(x)}dx}$$

Often one is interested in quantities of the form

$$\bar{f} \equiv \int_E f(x)\bar{\pi}(dx)$$

One may consider a Markov process X_t which has $\bar{\pi}$ as its invariant distribution and under the assumption that X_t is positive recurrent, the ergodic theorem gives

$$\frac{1}{t}\int_0^t f(X_s)ds \to \int_E f(x)\bar{\pi}(dx), \text{ a.s. as } t \to \infty\,, \tag{1}$$

for all $f \in L^1(\bar{\pi})$. Hence the estimator $f_t \equiv \frac{1}{t}\int_0^t f(X_s)ds$ can be used to approximate the expectation $\bar{f}$.

Detailed Balance Condition

Most of the times, a Markov chain is constructed that is time-reversible or in other words satisfies the detailed balance condition (DBC).
If for example that target stationary distribution is $\pi = (\pi_1, \cdots, \pi_N)$, then a sufficient condition to guarantee that

$$\lim_{t\to\infty} P_i(t) = \pi_i$$

is the detailed balance condition

$$\pi_i p_{ij} = \pi_j p_{ji} \text{ or } \pi(x)P(x,y) = \pi(y)P(y,x)$$

But DBC is only sufficient and not necessary!!

Questions.

1. What is the best Markov process X_t ?
2. What would be a reasonable criterion of optimality?

Questions.

1. What is the best Markov process X_t ?
2. What would be a reasonable criterion of optimality?

- *Can we use large deviations theory in an effective way? Connections with variance reduction?*

Part II

Investigation of convergence criteria

Overdamped Langevin equation.

To sample the Gibbs measure $\bar{\pi}$ on the set E

$$\bar{\pi}(dx) = \frac{e^{-2U(x)}dx}{\int_E e^{-2U(x)}dx}$$

one can consider the (time-reversible) Langevin equation

$$dX_t = -\nabla U(X_t)dt + dW_t\,. \tag{2}$$

Overdamped Langevin equation.

To sample the Gibbs measure $\bar{\pi}$ on the set E

$$\bar{\pi}(dx) = \frac{e^{-2U(x)}dx}{\int_E e^{-2U(x)}dx}$$

one can consider the (time-reversible) Langevin equation

$$dX_t = -\nabla U(X_t)dt + dW_t\,. \tag{2}$$

How should we describe the rate of convergence

$$\frac{1}{t}\int_0^t f(X_s)ds \to \int_E f(x)\bar{\pi}(dx), ? \tag{3}$$

Important remark: Need to use a performance measure that works directly with the empirical measure, which is what is used in practice!!!!

Standard measures of performance and their problems.

Spectral gap and 2nd eigenvalue. Consider the transition kernel

$$p(t, x_0, dx) = \mathbb{P}\left[X_t \in dx | X_0 = x_0\right]$$

Under the appropriate conditions, we have that the limit

$$p(t, x_0, dx) \to \pi(dx)$$

is determined by the spectral gap of the generator corresponding to X and the rate is exponential. In other words:

$$\left\| E_{\cdot} f(X_t) - \bar{f} \right\|_{L^2(\bar{\pi})} \leq C_0 \left\| f - \bar{f} \right\| e^{-\lambda t}$$

where

$\lambda = \inf\{$real part of non-zero eigenvalues in the spectrum of the operator$\}$.

Standard measures of performance and their problems.

Spectral gap and 2nd eigenvalue. Consider the transition kernel

$$p(t, x_0, dx) = \mathbb{P}\left[X_t \in dx | X_0 = x_0\right]$$

Under the appropriate conditions, we have that the limit

$$p(t, x_0, dx) \to \pi(dx)$$

is determined by the spectral gap of the generator corresponding to X and the rate is exponential. In other words:

$$\left\| E_{\cdot} f(X_t) - \bar{f} \right\|_{L^2(\bar{\pi})} \leq C_0 \left\| f - \bar{f} \right\| e^{-\lambda t}$$

where

$\lambda = \inf\{$real part of non-zero eigenvalues in the spectrum of the operator$\}$.

Is spectral gap the most appropriate measure to characterize convergence?

Another standard measures of performance.

Asymptotic variance: Hard to compute and it is a property of the algorithm only when we are at equilibrium. This is ok if we are interested in steady-state simulation.

$$t^{1/2}\left(\frac{1}{t}\int_0^t f(X_s)ds - \int f d\bar{\pi}\right) \Rightarrow N(0,\sigma_f^2)$$

and the asymptotic variance σ_f^2 is given in terms of the integrated autocorrelation function,

$$\sigma_f^2 = 2\int_0^\infty \mathbb{E}_{\bar{\pi}}\left[\left(f(X_0)-\bar{f}\right)\left(f(X_t)-\bar{f}\right)\right] dt$$

Hard to compute.

Question: Are there other possible Markov processes to use? How to compare their performance?

Overdamped Langevin equation.

To sample the Gibbs measure $\bar{\pi}$ on the set E

$$\bar{\pi}(dx) = \frac{e^{-2U(x)}dx}{\int_E e^{-2U(x)}dx}$$

one can consider the (time-reversible) Langevin equation

$$dX_t = -\nabla U(X_t)dt + dW_t \,. \tag{4}$$

Overdamped Langevin equation.

To sample the Gibbs measure $\bar{\pi}$ on the set E

$$\bar{\pi}(dx) = \frac{e^{-2U(x)}dx}{\int_E e^{-2U(x)}dx}$$

one can consider the (time-reversible) Langevin equation

$$dX_t = -\nabla U(X_t)dt + dW_t\,. \tag{4}$$

There are however many other stochastic differential equations with the same invariant measure and we may consider instead the family of equations

$$dX_t = \left[-\nabla U(X_t) + C(X_t)\right] dt + dW_t$$

where the vector field $C(x)$ satisfies the condition

$$\mathrm{div}(Ce^{-2U}) = 0$$

In this case the Markov process is not time-reversible!!

Overdamped Langevin equation.

There are many such C, indeed since $\mathrm{div}(Ce^{-2U}) = 0$ is equivalent to

$$\mathrm{div}(C) = 2C\nabla U\,,$$

so that we, for example, can choose C to be both divergence free and orthogonal to ∇U.

- One can always pick $C = S\nabla U$ for any antisymmetric matrix S.
- More generally, it is proved in [Barbarosie, 2011] that in dimension d any divergence free vector field can be written, locally, as the exterior product $C = \nabla V_1 \wedge \cdots \nabla V_{n-1}$ for some for $V_i \in \mathcal{C}^1(E;\mathbb{R})$. Therefore we can pick C of the form

$$C = \nabla U \wedge \nabla V_2 \cdots \nabla V_{n-1}\,.$$

for arbitrary $V_2, \cdots V_{n-1}$.

Donsker-Varadhan and Gärtner large deviations theory.

From a practical Monte-Carlo point of view one is interested in the distribution of the ergodic average $t^{-1}\int_0^t f(X_s)\,ds$ and how likely it is that this average differs from $\int f d\bar{\pi}$.

Donsker-Varadhan and Gärtner large deviations theory.

From a practical Monte-Carlo point of view one is interested in the distribution of the ergodic average $t^{-1}\int_0^t f(X_s)\,ds$ and how likely it is that this average differs from $\int fd\bar{\pi}$.

Define the empirical measure

$$\pi_t \equiv \frac{1}{t}\int_0^t \delta_{X_s}\,ds$$

which converges to $\bar{\pi}$ almost surely. If we have a large deviation for the family of measures π_t, which we write, symbolically as

$$\mathbb{P}\left\{\pi_t \approx \mu\right\} \asymp e^{-tI_C(\mu)}$$

Donsker-Varadhan and Gärtner large deviations theory.

From a practical Monte-Carlo point of view one is interested in the distribution of the ergodic average $t^{-1}\int_0^t f(X_s)\,ds$ and how likely it is that this average differs from $\int f d\bar{\pi}$.

Define the empirical measure

$$\pi_t \equiv \frac{1}{t}\int_0^t \delta_{X_s}\,ds$$

which converges to $\bar{\pi}$ almost surely. If we have a large deviation for the family of measures π_t, which we write, symbolically as

$$\mathbb{P}\left\{\pi_t \approx \mu\right\} \asymp e^{-tI_C(\mu)}$$

Note that rate function $I_C(\mu)$ quantifies the exponential rate at which the random measure π_t converges to $\bar{\pi}$. Clearly, the larger I_C is, the faster the convergence occurs.

Standard measures of performance and their problems.

How should we describe the rate of convergence

$$\frac{1}{t}\int_0^t f(X_s)ds \to \int_E f(x)\bar{\pi}(dx), ? \tag{5}$$

Important remark: None of the standard measures of performance works directly with the empirical measure, which is what is used in practice!!!!

Spectral gap and 2nd eigenvalue. Consider the transition kernel

$$p(t, x_0, dx) = \mathbb{P}\left[X_t \in dx | X_0 = x_0\right]$$

Under the appropriate conditions, we have that the limit

$$p(t, x_0, dx) \to \pi(dx)$$

is determined by the spectral gap of the generator corresponding to X and the rate is exponential.

Standard measures of performance and their problems.

How should we describe the rate of convergence

$$\frac{1}{t}\int_0^t f(X_s)ds \to \int_E f(x)\bar{\pi}(dx), ? \tag{5}$$

Important remark: None of the standard measures of performance works directly with the empirical measure, which is what is used in practice!!!!

Spectral gap and 2nd eigenvalue. Consider the transition kernel

$$p(t, x_0, dx) = \mathbb{P}\left[X_t \in dx | X_0 = x_0\right]$$

Under the appropriate conditions, we have that the limit

$$p(t, x_0, dx) \to \pi(dx)$$

is determined by the spectral gap of the generator corresponding to X and the rate is exponential.

Is spectral gap the most appropriate measure to characterize convergence?

Other standard measures of performance.

- **Asymptotic variance:** Hard to compute and it is a property of the algorithm only when we are at equilibrium. This is ok if we are interested in steady-state simulation.
- **Large deviations rate function.** It quantifies the exponential rate at which the empirical measure π_t converges to the Gibbs measure $\bar{\pi}$. The larger the rate function is, the fastest the convergence is (e.g., Dupuis, Liu, Plattner, and J. D. Doll (2012)).

Moreover, it turns out that we can characterize asymptotic variance, using the large deviations rate function!

Standard measures of performance.

Is spectral gap the most appropriate measure to characterize convergence?
The problem with spectral gap is that the information is on the density at fixed time t and not on the empirical measure. But empirical measure depends on sample path. Hence, spectral gap neglects potentially significant effects of time averaging in empirical measure.

Counter example. Let us consider the family of diffusions

$$dX_t = \delta dt + dW_t$$

on the circle S^1 with generator

$$\mathcal{L}_\delta = \Delta + \delta\nabla$$

For any $\delta \in \mathbb{R}$ the Lebesgue measure on S^1 is invariant but the diffusion is reversible only if $\delta = 0$.

Standard measures of performance.

The eigenvalues and eigenfunctions of $\mathcal{L}_\delta$ are

$$e_n = e^{inx}, \quad \lambda_n = -n^2 + in\delta, \quad n \in \mathbb{Z}\,.$$

The spectral gap is -1 for all $\delta \in \mathbb{R}$, i.e. the spectral gap does not move. However the asymptotic variance does decrease. For any real-valued function f with $\int_{S^1} f dx = 0$ we have for the asymptotic variance of the estimator

$$\sigma_f^2(\delta) = \int_0^\infty \langle e^{t\mathcal{L}} f(x)\,, f(x)\rangle_{L^2(dx)}\, dt \;=\; \langle \mathcal{L}_\delta^{-1} f\,, f\rangle_{L^2(dx)}$$

where $\mathcal{L}_\delta^{-1}$ is the inverse of $\mathcal{L}_\delta$ on the orthogonal complement of the eigenfunction 1. Expanding in the eigenfunctions we find

$$\sigma_f^2(\delta) \;=\; \sum_{n\in\mathbb{Z}, n\neq 0} \frac{|c_n|^2}{n^2 + in\delta} \;=\; \sum_{n=1}^{\infty} \frac{2|c_n|^2}{n^2+\delta^2}.$$

In this example, even though the spectral gap does not decrease at all, the variance not only decreases, but it can be made as small as we want by increasing δ^2.

What is known in the literature?

- Spectral gap decreases under a natural non-degeneracy condition on adding some irreversibility The corresponding eigenspace should not be invariant under the action of the added drift C (Hwang, Hwang-Ma and Sheu, (2005)).
- Let $U = 0$ and consider a one-parameter family of perturbations $C = \delta C_0$ for $\delta \in \mathbb{R}$ and C_0 is some divergence vector field. If the flow is weak-mixing then the second largest eigenvalue tends to 0 as $\delta \to \infty$ (Constantin-Kiselev-Ryshik-Zlatos, (2008)).
- Detailed analysis of the Gaussian case, i.e., when $U(x) = \frac{1}{2}x^T Ax$ and $C = JAx$ for a antisymmetric J can be found in Hwang-Ma-Sheu (1993) and Lelievre-Nier-Pavliotis (2012).
- Evidence that violation of detailed balance accelerates relaxation in recent physics literature, Ichiki-Ohzeki (2013).
- Use of large deviations to analyze parallel tempering type of algorithms in Dupuis, Liu, Plattner, and J. D. Doll (2012) (infinite swapping algorithm)

Part III

What can large deviations theory say?

Approximation via diffusions

To sample the Gibbs measure $\bar{\pi}$ on the set E

$$\bar{\pi}(dx) = \frac{e^{-2U(x)}dx}{\int_E e^{-2U(x)}dx}$$

one can consider the (time-reversible) Langevin equation

$$dX_t = -\nabla U(X_t)dt + dW_t\,. \tag{6}$$

Approximation via diffusions

To sample the Gibbs measure $\bar{\pi}$ on the set E

$$\bar{\pi}(dx) = \frac{e^{-2U(x)}dx}{\int_E e^{-2U(x)}dx}$$

one can consider the (time-reversible) Langevin equation

$$dX_t = -\nabla U(X_t)dt + dW_t\,. \tag{6}$$

There are however many other stochastic differential equations with the same invariant measure and we may consider instead the family of equations

$$dX_t = \left[-\nabla U(X_t) + C(X_t)\right]dt + dW_t$$

where the vector field $C(x)$ satisfies the condition

$$\mathrm{div}(Ce^{-2U}) = 0$$

In this case the Markov process is not time-reversible!! Can we somehow optimize by choosing C?

Donsker-Varadhan and Gärtner theory

Define the empirical measure

$$\pi_t \equiv \frac{1}{t} \int_0^t \delta_{X_s} \, ds$$

which converges to $\bar{\pi}$ almost surely. If we have a large deviation for the family of measures π_t, which we write, symbolically as

$$\mathbb{P}\left\{\pi_t \approx \mu\right\} \asymp e^{-tI_C(\mu)}$$

The information in $I_C(\mu)$ can be used to study observable: we have for $f \in \mathcal{C}(E;\mathbb{R})$ the large deviation principle

$$\mathbb{P}\left\{\frac{1}{t} \int_0^t f(X_s) \, ds \approx \ell \right\} \asymp e^{-t\tilde{I}_{f,C}(\ell)}$$

where

$$\tilde{I}_{f,C}(\ell) = \inf_{\mu \in \mathcal{P}(E)} \left\{ I_C(\mu) : \langle f, \mu \rangle = \ell \right\} ,$$

Donsker-Varadhan and Gärtner theory

In particular, if $\mathcal{A}$ is the generator of the Markov process and $\mathcal{D}$ its domain, then the Dosnker-Varadhan functional takes the form

$$I(\mu) = -\inf_{u \in \{u \in \mathcal{D}, u > 0\}} \int_E \frac{\mathcal{A}u}{u} d\mu$$

A more explicit formula due to Gärtner:

Theorem (Gärtner).

Consider the SDE

$$dX_t = b(X_t) + dW_t$$

on a compact manifold E with $b \in \mathcal{C}^1(E; \mathbb{R}^d)$. The Donsker-Vardhan rate function $I(\mu)$ takes the form

$$I(\mu) = \frac{1}{2} \int_E |\nabla \phi(x)|^2 \, d\mu(x) \tag{7}$$

where ϕ is the unique (up to constant) solution of the equation

$$\Delta \phi + \frac{1}{p} (\nabla p, \nabla \phi) = \frac{1}{p} \mathcal{L}^* p \tag{8}$$

Simplifications

In the special case where $b = -\nabla U$ is a gradient, then $\phi(x) = \frac{1}{2}\log p(x) + U(X) + \text{constant}$ and we get

$$I(\mu) = \frac{1}{2}\int_E \left|\frac{1}{2}\frac{\nabla p(x)}{p(x)} + \nabla U(x)\right|^2 d\mu(x) \tag{9}$$

which the usual explicit formula for the rate function in the reversible case. Motivated, by this if we set $\phi(x) = \frac{1}{2}\log p(x) + \psi(x)$, then we get the following representation.

Lemma

We have

$$I(\mu) = \frac{1}{8}\int_E \left|\frac{\nabla p(x)}{p(x)}\right|^2 d\mu(x) + \frac{1}{2}\int_E |\nabla\psi(x)|^2 d\mu(x) - \frac{1}{2}\int_E \frac{b\nabla p}{p} d\mu(x)$$

where ψ is the unique (up to constant) solution of the equation

$$\mathrm{div}\left[p\left(b + \nabla\psi\right)\right] = 0.$$

Behavior of rate function

Recall that we are comparing

$$
\begin{aligned}
dX_t &= [-\nabla U(X_t)]\, dt + dW_t \\
dX_t &= [-\nabla U(X_t) + C(X_t)]\, dt + dW_t
\end{aligned}
$$

Same invariant measure but reversible versus irreversible!

Theorem

Assume that $C \neq 0$ such that $\mathrm{div} C = 2C\nabla U$. For any $\mu \in \mathcal{P}(E)$ we have $I_C(\mu) \geq I_0(\mu)$. If $\mu(dx) = p(x)dx$ is a measure with positive density $p \in \mathcal{C}^{(2+\alpha)}(E)$ for some $\alpha > 0$ and $\mu \neq \bar{\pi}$ then we have

$$
I_C(\mu) = I_0(\mu) + \frac{1}{2}\int_E |\nabla\psi_C(x) - \nabla U(x)|^2\, d\mu(x)\,.
$$

where ψ_C is the unique solution (up to a constant) of the equation

$$
\mathrm{d}iv\, [p\,(-\nabla U + C + \nabla\psi_C)] = 0.
$$

Moreover we have $I_C(\mu) = I_0(\mu)$ if and only if the positive density $p(x)$ satisfies $\mathrm{div}\,(p(x)C(x)) = 0$. Equivalently such p have the form $p(x) = e^{2G(x)}$ where G is such that $G + U$ is an invariant for the vector field C (i.e., $C\nabla(G+U) = 0$).

Behavior of rate function

To obtain a slightly more quantitative result let us consider a one-parameter family $C(x) = \delta C_0(X)$ where $\delta \in \mathbb{R}$ and $C_0 \neq 0$ such that $\mathrm{div} C_0 = 2C_0 \nabla U$.

Theorem

Assume that $C_0 \neq 0$ such that $\mathrm{div} C_0 = 2C_0 \nabla U$. Consider the measure $\mu(dx) = p(x)dx$ with positive density $p \in \mathcal{C}^{(2+\alpha)}(E)$ for some $\alpha > 0$. Then we

$$I_{\delta C_0}(\mu) = I_0(\mu) + \delta^2 K(\mu)\,.$$

where the functional $K(\mu)$ is strictly positive if and only if $\mathrm{div}\,(p(x)C(x)) \neq 0$.

Namely, rate function, is quadratic in δ!

What about observables?

For $f \in \mathcal{C}(E, \mathbb{R})$, by the contraction principle,

$$\mathbb{P}\left\{ \frac{1}{t} \int_0^t f(X_s)\, ds \approx \ell \right\} \asymp e^{-t \tilde{I}_{f,C}(\ell)}$$

where

$$\tilde{I}_{f,C}(\ell) = \inf_{\mu \in \mathcal{P}(E)} \left\{ I_C(\mu) : \langle f, \mu \rangle = \ell \right\}$$

Theorem

Consider $f \in \mathcal{C}^{(\alpha)}(E)$ and $\ell \in (\min_x f(x), \max_x f(x))$ with $\ell \neq \int f d\bar{\pi}$. Fix a vector field C as in assumption (**H**). Then we have

$$\tilde{I}_{f,C}(\ell) \geq \tilde{I}_{f,0}(\ell)\,.$$

Moreover if there exists ℓ_0 such that for this particular field C, $\tilde{I}_{f,C}(\ell_0) = \tilde{I}_{f,0}(\ell_0)$ then we must have

$$\widehat{\beta}(\ell_0) f \;=\; \frac{1}{2}\Delta(G+U) + \frac{1}{2}|\nabla G|^2 - \frac{1}{2}|\nabla U|^2\,, \qquad (10)$$

where G is such that $G+U$ is invariant under the particular vector field C.

What about observables?

- Letting $\mathcal{L}_0$ denote the infinitesimal generator of the reversible process X_t (i.e., when $C = 0$), we get that (10) can be rewritten as a nonlinear Poisson equation of the form

$$\mathcal{H}(G+U) = \widehat{\beta}(\ell_0) f \,, \tag{11}$$

where

$$\mathcal{H}(G+U) = e^{-(G+U)}\mathcal{L}_0 e^{G+U} = \frac{1}{2}\Delta(G+U) + \frac{1}{2}|\nabla G|^2 - \frac{1}{2}|\nabla U|^2.$$

- This result does not mean that there are observables f, for which variance reduction cannot be attained. It only means, that for a given observable, one should choose a vector field C, such that there is no G that satisfies both $C\nabla(G+U) = 0$ and (10).

Sketch of the proof 1/5.

Overview of the proof:

1. Recall that we already know that $I_C(\mu) > I_0(\mu)$.
2. Since $\tilde{I}_{f,C}(\ell) = \inf_{\mu \in \mathcal{P}(E)} \{I_C(\mu) : \langle f, \mu \rangle = \ell\}$, is the minimizer μ achieved?
3. The mimimizer μ is achieved for "good" functions f.
4. Exploiting these properties, a contradiction argument gives us that $\tilde{I}_{f,C}(\ell) \geq \tilde{I}_{f,0}(\ell)$ and the condition under which the inequality is strict.

Sketch of the proof 2/5.

The proof of the **existence and characterization of the minimizer** μ is based on the representation of the rate function in terms of a Legendre transform

$$\tilde{I}_{f,C}(\ell) = \sup_{\beta \in \mathbb{R}} \{\beta\ell - \lambda(\beta f)\} .$$

where the eigenvalue $\lambda(\beta f)$ is a smooth strictly convex function of β

$$\lambda(\beta f) = \lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}\left[e^{\int_0^t \beta f(X_s) ds}\right] .$$

If ℓ belongs to the range of f we have

$$\tilde{I}_{f,C}(\ell) = \widehat{\beta}\ell - \lambda(\widehat{\beta} f), \quad \text{with } \widehat{\beta} \text{ given by } \ell = \frac{d}{d\beta}\lambda(\widehat{\beta} f) .$$

Sketch of the proof 3/5.

Since $f \in \mathcal{C}^{(\alpha)}$, $\lambda(\beta f)$ is the maximal eigenvalue of $\mathcal{L}_C + \beta f$

$$(\mathcal{L}_C + \beta f)u(\beta f) = \lambda(\beta f)u(\beta f)\,, \tag{12}$$

and is a smooth convex function of β. Here $u(\beta f)$ is the corresponding eigenfunction.

With $u(\beta f) = e^{\phi(\beta f)}$, the eigenvalue equation can be equivalently written as

$$\mathcal{L}_C \phi(\beta f) + \frac{1}{2}\left|\nabla \phi(\beta f)\right|^2 = \lambda(\beta f) - \beta f \tag{13}$$

Differentiating with respect to β and setting $\psi(\beta f) = \frac{\partial \phi}{\partial \beta}(\beta f)$ we see that $\psi(\beta f)$ satisfies the equation

$$\mathcal{L}_C \psi(\beta f) + (\nabla \phi(\beta f), \nabla \psi(\beta f)) = \lambda_f'(\beta) - f$$

Sketch of the proof 4/5.

Equivalently

$$\mathcal{L}_{C+\nabla\phi(\beta f)}\psi = \lambda'_f(\beta) - f$$

Thus, the constraint $\langle f, \mu\rangle = \ell$, implies that in order to have $\ell = \lambda'_f(\hat{\beta})$ for some $\hat{\beta}$, $\mu_{\hat{\beta}}$ should be the invariant measure for the process with generator $\mathcal{L}_{C+\nabla\phi(\hat{\beta}f)}$.
Since $\nabla\phi \in \mathcal{C}^{(1+\alpha)}$ the corresponding invariant measure $\mu_{\hat{\beta}}$ is strictly positive and has a density $p(x) \in \mathcal{C}^{(2+\alpha)}$.
To conclude the proof, by Gärtner's result we have
$I_C(\mu_{\hat{\beta}}) = \mu(\hat{\beta}f) - \lambda(\hat{\beta}f)$. But since $\mu(f) = \ell$ this is also equal to $I_{f,C}(\ell)$.

Sketch of the proof 5/5.

Hence, we have obtained

$$\tilde{I}_{f,C}(\ell) = I_C(\mu_{C,\hat{\beta}}) \text{ with } \mu_{C,\hat{\beta}}(dx) = p_{C,\hat{\beta}}(x)dx\,.$$

The rest is standard.

1. Let $\mathrm{div}(Cp_{C,\hat{\beta}}) \neq 0$. Assume that the rate functions with $C = 0$ and $C \neq 0$ are equal and get a contradiction.
2. Let $\mathrm{div}(Cp_{C,\hat{\beta}}) = 0$. Let us write $p_{C,\hat{\beta}} = e^{-2G}$, so we must have $C \cdot \nabla G = 2\mathrm{div}(C)$. Keeping in mind that $p_{C,\hat{\beta}}(x)$ is invariant density corresponding to a known operator gives us that we must have $p_{C,\hat{\beta}} = p_{0,\hat{\beta}} = e^{2(\phi(\beta f)-U)+const}$. Thus $\phi(\beta f) = G + U$ and $C \cdot \nabla\phi(\beta f) = 2\mathrm{div}(C)$ and (13) reduces to $(\mathcal{L}_0 + \hat{\beta}f)e^{\phi} = \lambda(\hat{\beta}f)e^{\phi}$. Solving for f gives the nonlinear Poisson equation (11):

$$\hat{\beta}f = e^{-\phi}\mathcal{L}_0 e^{\phi} + \mathrm{const}. \tag{14}$$

Part IV

What about variance reduction?

Asymptotic variance

Under our assumptions the central limit theorem holds for the ergodic average f_t and we have

$$t^{1/2}\left(\frac{1}{t}\int_0^t f(X_s)ds - \int fd\bar{\pi}\right) \Rightarrow N(0, \sigma_f^2)$$

and the asymptotic variance σ_f^2 is given in terms of the integrated autocorrelation function,

$$\sigma_f^2 = 2\int_0^\infty \mathbb{E}_{\bar{\pi}}\left[\left(f(X_0) - \bar{f}\right)\left(f(X_t) - \bar{f}\right)\right] dt$$

Asymptotic variance

This is a convenient quantity from a practical point of view since there exists easily implementable estimators for σ_f^2. On the other hand the asymptotic variance σ_f^2 is related to the curvature of the rate function $I_f(\ell)$ around the mean $\bar{f}$ we have

$$\tilde{I}_f''(\bar{f}) = \frac{1}{2\sigma_f^2}\,.$$

From previous theorem it follows immediately that $\sigma_{f,C}^2 \leq \sigma_{f,0}^2$ but in fact the addition of an irreversible drift strictly generically decreases the asymptotic variance.

Theorem

Assume that $C \neq 0$ is a vector field such that $\mathrm{div} C = 2C\nabla U$. Let $f \in \mathcal{C}^{(\alpha)}(E)$ such that for some $\epsilon > 0$ and $\ell \in (\bar{f} - \epsilon, \bar{f} + \epsilon) \setminus \{\bar{f}\}$ we have $\tilde{I}_{f,C}(\ell) > \tilde{I}_{f,0}(\ell)$. Then we have

$$\sigma_{f,C}^2 < \sigma_{f,0}^2.$$

Sketch of the proof

It is clear that the relation $\sigma_f^2 = \frac{1}{2\tilde{I}_f''(\bar{f})}$ implies that it is enough to prove that for $C \neq 0$ and $f \in \mathcal{C}^{(\alpha)}(E)$

$$\tilde{I}_{f,C}''(\bar{f}) - \tilde{I}_{f,0}''(\bar{f}) > 0$$

The proof of this statement follows by precise computation of first and then second order Gâteaux derivatives.

Sketch of the proof

It is clear that the relation $\sigma_f^2 = \frac{1}{2\tilde{I}_f''(\bar{f})}$ implies that it is enough to prove that for $C \neq 0$ and $f \in \mathcal{C}^{(\alpha)}(E)$

$$\tilde{I}''_{f,C}(\bar{f}) - \tilde{I}''_{f,0}(\bar{f}) > 0$$

The proof of this statement follows by precise computation of first and then second order Gâteaux derivatives.

The formula of the second order derivative $\tilde{I}''_{f,C}(\bar{f})$ that is derived, provides a natural optimization problem for the choice of C.

Part V

Increasing irreversibility and diffusion on graphs

Increasing Irreversibility.

Recall the result:

- Let $U = 0$ and consider a one-parameter family of perturbations $C = \frac{1}{\epsilon} C_0$ for $\epsilon \in \mathbb{R}$ and C_0 is some divergence vector field. If the flow is weak-mixing then the second largest eigenvalue tends to 0 as $\epsilon \to 0$ (Constantin-Kiselev-Ryshik-Zlatos, (2008)).

Questions. (1): What happens when $U \neq 0$? (2): How does the underlying process behaves? (3): How does the asymptotic variance behaves as $\epsilon \to 0$? (4): What about metastability effects?

Increasing Irreversibility.

Consider

$$dX_t^\epsilon = \left[-\nabla U(X_t^\epsilon) + \frac{1}{\epsilon} C(X_t^\epsilon) \right] dt + dW_t$$

where the vector field $C(x)$ satisfies the condition

$$\mathrm{div}(Ce^{-2U}) = 0$$

Theorem

Assume that $C \neq 0$ is a vector field such that $\mathrm{div} C = 2C\nabla U$. If $\tilde{I}_{f,\frac{1}{\epsilon}C}(\ell) > \tilde{I}_{f,0}(\ell)$ in a neighborhood of $\bar{f} = \int_E f(x)\pi(dx)$, but excluding $\bar{f}$, then the map $|\epsilon| \mapsto \sigma^2_{f,\frac{1}{\epsilon}}$ is a monotone increasing function and thus its limit as $\epsilon \to 0$ exists.

In order to understand the limiting behavior of $\sigma^2_{f,\frac{1}{\epsilon}}$ as $\epsilon \to 0$, we need to first understand the limiting behavior of X_t^ϵ as $\epsilon \to 0$.

Limiting behavior of the underlying process.

- Assume that U has finitely many non-degenerate critical points and that each connected level set component of U contains at most one critical point.
- Following Freidlin and Wenztell, let us consider a finite graph Γ, which represents the structure of the level sets of the potential function U on E.
- Identify the points that belong to the connected components of each of the level sets of U.
- Each of the domains that is bounded by the separatrices gets mapped into an edge of the graph. At the same time the separatrices gets mapped to the vertexes of Γ.
- Exterior vertexes correspond to minima of U, whereas interior vertexes correspond to saddle points of U. Edges of Γ are indexed by $I_1, \cdots, I_m$
- Each point on Γ is indexed by a pair $y = (z, i)$ where z is the value of U on the level set corresponding to y and i is the edge number containing y. Clearly the pair $y = (z, i)$ forms a global coordinate on Γ.
- For any given point $x \in \mathbb{R}^d$, let $Q : E \mapsto \Gamma$ by $Q(x) = (U(x), i(x))$ be the corresponding projection on the graph.

Limiting behavior of the underlying process.

Theorem (Freidlin and Wenztell)

Under appropriate assumptions, for any $0 < T < \infty$, the process $Y_t^\epsilon = Q(X_t^\epsilon)$, converges weakly in $\mathcal{C}([0, T], \Gamma)$ to a certain Markov process, denoted by Y_t, on Γ with continuous trajectories, which is also exponentially mixing.

Limiting behavior of the asymptotic variance.

Theorem

Let us assume that the Condition of the previous theorem hold and let Y_t be the continuous Markov process on the graph Γ indicated in the previous theorem. Let $f \in \mathcal{C}^{2+\alpha}(E)$ such that $\bar{f} = 0$. For $(z, i) \in \Gamma$, define $\widehat{f}(z, i)$ to be the average of f on the graph Γ over the corresponding connected component of the level set U. Namely, let

$$\widehat{f}(z,i) = \oint_{d_i(z)} f(x) m_{z,i}(x) \ell(dx) = \frac{1}{T_i(z)} \oint_{d_i(z)} \frac{f(x)}{|\nabla U(x)|} m(x) \ell(dx)$$

Then, we have that $\sigma_f^2(0) = \lim_{\epsilon \to 0} \sigma_f^2(\epsilon)$, where

$$\sigma_f^2(0) = 2 \int_0^\infty \mathbb{E}_\mu \left[\widehat{f}(Y_0) \widehat{f}(Y_t) \right] dt \tag{15}$$

and $\mu = \pi \circ \Gamma^{-1}$ is the invariant measure of the process Y on Γ.

It is straightforward to see that this is the asymptotic variance of an ergodic average on the graph. In particular, we have

$$\sigma_f^2(0) = \lim_{t \to \infty} t \mathrm{Var} \left(\frac{1}{t} \int_0^t \widehat{f}(Y_s) ds \right). \tag{16}$$

Part VI

Related Multiscale Integrators

A numerical method

Consider

$$dX_t^\epsilon = \left[-\nabla U(X_t^\epsilon) + \frac{1}{\epsilon} C(X_t^\epsilon)\right] dt + \sqrt{2\beta} dW_t$$

where the vector field $C(x)$ satisfies the condition

$$\mathrm{div}\, C = C \nabla U = 0$$

Consider a split-step time integrator where for the short time step τ we use the whole SDE and for the long time step $\delta - \tau$ we neglect the irreversible drift.

$$\bar{X}_{t_n+\tau} - \bar{X}_{t_n} = -\tau \nabla U(\bar{X}_{t_n}) + \frac{\tau}{\epsilon} C(\bar{X}_{t_n}) + \sqrt{2\beta\tau}\, \xi_n; \tag{17a}$$

$$\bar{X}_{t_n+\delta} - \bar{X}_{t_n+\tau} = -(\delta - \tau) \nabla U(\bar{X}_{t_n+\tau}) + \sqrt{2\beta(\delta - \tau)}\, \xi_n', \tag{17b}$$

where ξ_n and ξ_n' are independent standard normal random variables.

Convergence of the numerical method

Theorem

Assume that $\epsilon, \delta, \tau \downarrow 0$ are such that $\frac{\delta\epsilon}{\tau}, \frac{\tau}{\epsilon}, \left(\frac{\tau}{\epsilon}\right)^{3/2}\frac{1}{\delta} \downarrow 0$. Then, for $\tau < \delta < \frac{\tau}{\epsilon} \ll 1$ sufficiently small, the process $Q(\bar{X}^{\epsilon}_{n\delta}) = (U(\bar{X}^{\epsilon}_{n\delta}), i(\bar{X}^{\epsilon}_{n\delta}))$ converges in distribution to the process on the tree $Y_{\cdot}$. In addition, convergence to the invariant measure μ of the Y process holds, in the sense that for any bounded and uniformly Lipschitz test function f we have that for all $t > 0$

$$\lim_{h\downarrow 0} \lim_{\epsilon,\delta,\frac{\delta\epsilon}{\tau},\frac{\tau}{\epsilon},\left(\frac{\tau}{\epsilon}\right)^{3/2}\frac{1}{\delta}\downarrow 0} \frac{1}{h}\int_{t}^{t+h} E_{\pi} f(\bar{X}^{\epsilon}_{s})ds = E_{\mu}\widehat{f}(Y_t),$$

where π is the invariant measure of the continuous process X^{ϵ}.

Part VII

On general homogeneous Markov process

General theory based on generators.

- Let $X_0(t)$ be an ergodic time reversible continuous-time Markov process on the state space S with invariant measure π.
- Let $L^2_{\mathbb{R}}(\pi)$ be the real Hilbert space with scalar product $\langle f, g\rangle = \int f(x)g(x)\pi(dx)$. We denote by T_0^t the corresponding strongly continuous Markov semigroup as an operator on $L^2_{\mathbb{R}}(\pi)$ with infinitesimal generator $\mathcal{L}_0$ with domain $D(\mathcal{L}_0)$.
- We have for all $f, g \in D(\mathcal{L}_0)$, $\langle f, \mathcal{L}_0 g\rangle = \langle \mathcal{L}_0 f, g\rangle$.
- Assume the semigroup T_0^t has a spectral gap in $L^2_{\mathbb{R}}(\pi)$, i.e., there exists $\lambda_0 < 0$ such that $\sigma(\mathcal{L}_0) \setminus \{0\} \subset (-\infty, \lambda_0]$.

General theory based on generators.

Consider any type of perturbation of the operator $\mathcal{L}_0$ of the form

$$\mathcal{L} = \mathcal{L}_0 + S + A$$

which maintains the invariant measure. Here S is a negative definite reversible perturbation and A is an irreversible perturbation. Then, as it is proven in [Rey-Bellet &S. 2016]

- spectral gap, asymptotic variance, large deviations behavior all improve!
- degree of improvement depends on the perturbation applied.

We have implemented this scheme for some concrete situations:

- Reversible and irreversible perturbations of continuous time Markov chains and diffusion processes.
- Reversible perturbations of Markov jump processes.

General developed theory covers previous partial results, such the famous Peskun and Tierney constructions, as special cases.

Part VIII

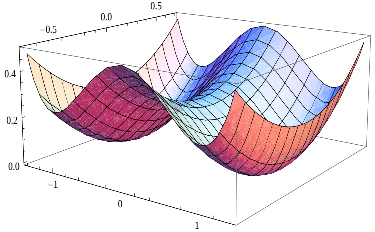
Simulation results

Example.

Consider that we want to sample from the stationary distribution

$$\bar{\pi}(dxdy) = \frac{e^{-\frac{U(x,y)}{D}}}{\int_{\mathbb{R}^2} e^{-\frac{U(x,y)}{D}} dxdy} dxdy$$

where D is some constant and $U(x,y) = \frac{1}{4}(x^2-1)^2 + \frac{1}{2}y^2$,



Example.

- Consider the Markov process

$$dZ_t = [-\nabla U(Z_t) + C(Z_t)]\,dt + \sqrt{2D}dW_t, \quad Z_0 = 0$$

where for $z = (x, y)$, $U(x, y) = \frac{1}{4}(x^2 - 1)^2 + \frac{1}{2}y^2$.

- Let $D = 0.1$ and $C(x, y) = \delta C_0(x, y)$ with $C_0(x, y) = J\nabla U(x, y)$. Here, $\delta \in \mathbb{R}$, I is the 2×2 identity matrix and J is the standard 2×2 antisymmetric matrix, i.e., $J_{12} = 1$ and $J_{21} = -1$.

- Notice that for any $\delta \in \mathbb{R}$, the invariant measure is

$$\bar{\pi}(dxdy) = \frac{e^{-\frac{U(x,y)}{D}}}{\int_{\mathbb{R}^2} e^{-\frac{U(x,y)}{D}}dxdy}dxdy$$

Example.

Let us suppose that we want to compute the following observables

$$\bar{f}_i = \int_{\mathbb{R}^2} f_i\,(x,y)\,\bar{\pi}(dxdy), \quad i = 1,2$$

where

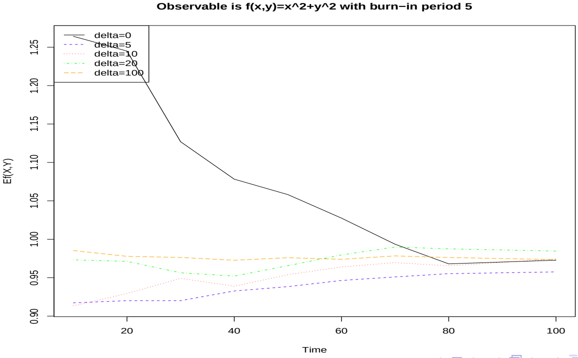
$$f_1(x,y) = x^2 + y^2, \quad f_2(x,y) = U(x,y) = \frac{1}{4}(x^2-1)^2 + \frac{1}{2}y^2$$

Speed of convergence

It is known that an estimator for $\bar{f}_i$ is given by

$$\hat{\bar{f}}_i(t) = \frac{1}{t-v}\int_v^t f_i\left(X_s, Y_s\right) ds$$

where v is some burn-in period that is used with the hope that the bias has been significantly reduced by time v.



Variance reduction

In general, a central limit theorem holds and takes the following form

$$t^{1/2}\left(\hat{\bar{f}}(t)-\bar{f}\right)\Rightarrow N(0,\sigma_f^2)$$

In order to estimate σ_f^2, we use the well established method of batch means Then for $\kappa=1,\cdots,m$ (m is number of batches) we define

$$\hat{\bar{f}}(t;\kappa)=\frac{1}{t/m}\int_{(\kappa-1)t/m}^{\kappa t/m} f\left(X_s,Y_s\right)ds,$$

$$\hat{\bar{f}}(t)=\frac{1}{m}\sum_{\kappa=1}^{m}\hat{\bar{f}}(t;\kappa)$$

and

$$s_m^2(t)=\frac{1}{m-1}\sum_{\kappa=1}^{m}\left(\hat{\bar{f}}(t;\kappa)-\hat{\bar{f}}(t)\right)^2$$

Variance reduction

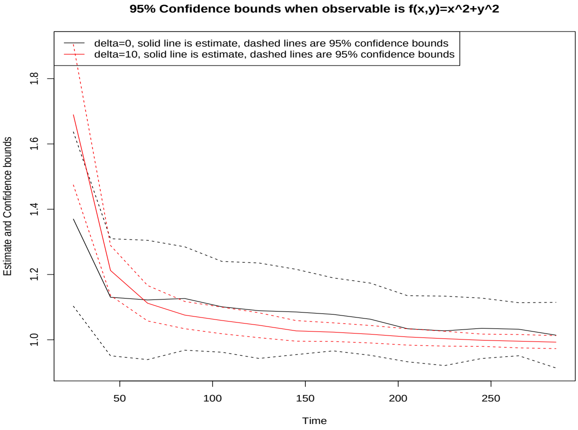
Then, we have in distribution

$$\sqrt{m}\frac{\hat{\bar{f}}(t) - \bar{f}}{s_m(t)} \Rightarrow T_{m-1}, \quad \text{as } t \to \infty$$

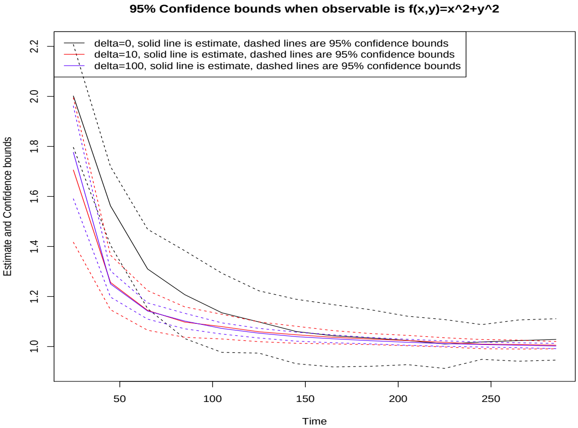
where T_{m-1} is the Student's T distribution with $m-1$ degrees of freedom. So, a $(1-\alpha)\%$ confidence interval is given by

$$\left(\hat{\bar{f}}(t) - t_{\alpha/2,m-1}s_m(t)/\sqrt{m}, \hat{\bar{f}}(t) + t_{\alpha/2,m-1}s_m(t)/\sqrt{m}\right)$$

Variance reduction



Variance reduction



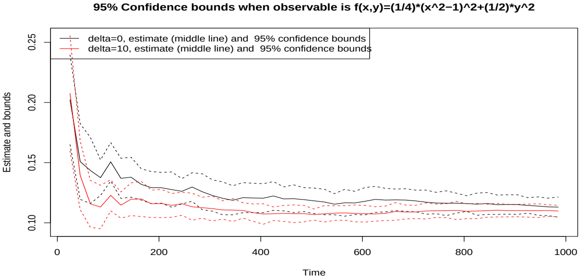
Variance reduction

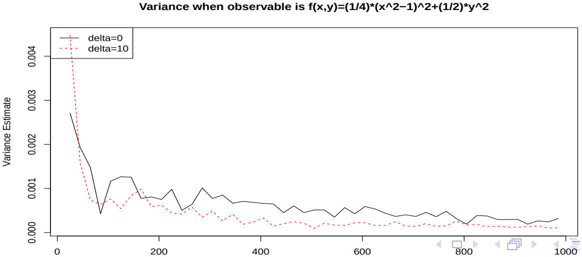
$\delta \mid t$	25	100	160	220	295
0	0.22	0.08	0.038	0.029	0.011
10	0.19	0.01	0.007	0.005	0.002
100	0.09	0.001	$3e-04$	$2.8e-04$	$1.3e-04$

Table: Estimated variance values for different pairs (δ, t).

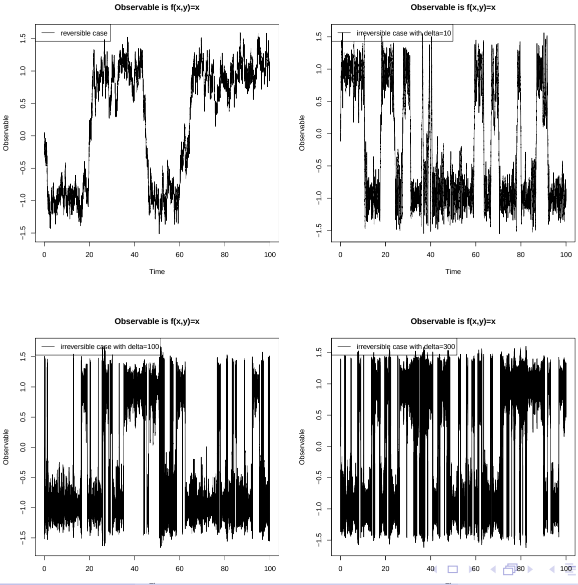
Variance reduction of about two orders of magnitude when $\delta = 100$ versus $\delta = 0$!!

Variance reduction





Effect on metastable behavior



Numerical performance of multiscale integrators

Consider the potential

$$U(x,y) = \frac{1}{4}\left[(x^2-1)^2((y^2-2)^2+1)+2y^2-y/8\right]+e^{-8x^2-4y^2}$$

with inverse temperature $\beta = 0.2$.

- We take the observable $f(x,y) = (x-1)^2 + y^2$.
- Set $\delta = 5e\text{-}3$ and $\tau = 10\delta\epsilon = 0.05\epsilon$. The total simulation time is $T = 2000$ with a burn-in period $T_{\text{burn}} = 20$.
- The true average of the observable is obtained by a discretization of the Gibbs distribution on the phase space with a fine mesh, which gives approximately $f \approx 2.1986$.
- Euler-Maruyama scheme loses stability for ϵ smaller than 0.1.

	ϵ	τ	δ	$\mathbb{E}(\text{Err}_f)$	$sd(\text{Err}_f)$	$\mathbb{E}(\text{AVar}_f)$	$sd(\text{AVar}_f)$
E-M	5		5e-3	5.0964e-1	3.6448e-1	2.4957e00	5.3972e-1
	5e-1		5e-3	3.1799e-1	2.3650e-1	1.8792e00	3.7341e-1
	1e-1		5e-3	1.0730e-1	8.0109e-2	3.4238e-1	1.0624e-1
M-I	1e-2	5e-4	5e-3	1.0347e-1	7.6486e-2	3.0648e-1	9.2405e-2
	1e-3	5e-5	5e-3	1.0255e-1	7.7384e-2	2.9778e-1	9.1258e-2
	1e-4	5e-6	5e-3	1.0108e-1	7.7460e-2	2.9760e-1	8.8149e-2

Table: Comparison of the Euler-Maruyama scheme and the multiscale integrator .

Part IX

Summary and challenges

To summarize, we have developed a systematic approach to the problem.

1. Large deviations action functional is directly related to the empirical measure.
   - Can be used a measure of performance. The bigger it is the better.
2. Asymptotic variance is directly related to the second derivative of the rate function!
3. A natural optimization problem for the choice of the optimal perturbation is being defined.
4. Introducing irreversibility speeds up convergence, and reduces significantly the variance of the estimator.
5. Multiscale integrator numerical methods can help in simulating efficiently.

Challenges.

1. Characterizing optimal perturbations. Known results only in the Gaussian case (Lelievre, Nier and Pavliotis). General case seems to be hard but we have preliminary results...
2. We have developed an appropriate numerical algorithm and studied its theoretical behavior. More work in this area is certainly needed.
3. Using irreversible proposal within a reversible algorithm like MALA. This is a natural question to consider and preprint will be available soon (joint work with Michela Ottobre, Natesh Pillai). It is questionable whether potential benefits balance the extra computational work.

Part X

References

References. I

1. K.A. Athreya, H. Doss and J. Sethuraman, On the convergence of the Markov chain simulation method, *Annals of Statistics*, Vol. 24, (1996), pp. 69-100.
2. C. Barbarosie, Representation of divergence-free vector fields, *Quarterly of Applied Mathematics*, Vol. 69, (2011), pp. 309–316.
3. M. Bedard and J.S. Rosenthal, Optimal Scaling of Metropolis Algorithms: Heading Towards General Target Distributions, *Canadian Jounral of Statistics*, Vol. 36, Issue 4, (2008), pp. 483-503.
4. P. Constantin, A. Kiselev, L. Ryshik and A. Zlatos, Diffusion and mixing in fluid flow *Annals of Mathematics*, Vo. 168 (2008), pp. 643-674

References. II

5. M.D. Donsker and S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large times, I, *Communications Pure in Applied Mathematics*, Vol. 28, (1975), pp. 1-47, II, *Communications on Pure in Applied Mathematics*, Vol. 28, (1975), pp. 279–301, and III, *Communications on Pure in Applied Mathematics*, Vol. 29, (1976), pp. 389-461.

6. P. Dupuis, Y. Liu, N. Plattner, and J. D. Doll, On the Infinite Swapping Limit for Parallel Tempering. *SIAM Multiscale Modeling and Simulation*, Vol. 10, Issue 3, (2012), pp. 986-1022.

7. M.I. Freidlin and A.D. Wentzell, Random Perturbations of Hamiltonian Systems, *Memoirs of the American Mathematical Society*, Vol. 109, No. 523, 1994.

8. B. Franke, C.-R. Hwang, H.-M. Pai, and S.-J. Sheu, The behavior of the spectral gap under growing drift, *Transactions of the American Mathematical Society*, Vol 362, No. 3 (2010), pp. 1325-1350.

References. III

9. A. Frigessi, C.R. Hwang and L. Younes, Optimal spectral structures of reversible stochatic matrices, Monte Carlo methods and the simulation of Markov random fields, *Annals of Applied Probability*, Vol. 2, (1992), pp. 610-628.
10. A. Frigessi, C.R. Hwang, S.J. Sheu and P. Di Stefano, Convergence rates of the Gibbs sampler, the Metropolis algorithm, and their single-site updating dynamics, *Journal of Royal Statistical Society Series B, Statistical Methodology*, Vol. 55, (1993), pp. 205-219.
11. A. Ichiki and M. Ohzeki, Violation of detailed balance condition accelerates relaxation, *Physical Review E*, Vol. 88, (2013), pp. 020101(R).
12. Jürgen Gärtner, On large deviations from the invariant measure, *Theory of probability and its applications*, Vol. XXII, No. 1, (1977), pp. 24-39.
13. W.R. Gilks and G.O. Roberts, Strategies for improving MCMC, *Monte Carlo Markov Chain in practice*, Chapman and Hall, Boca Raton, FL, (1996), pp. 89-114.

References. IV

14. C.R. Hwang, S.Y. Hwang-Ma and S.J. Sheu, Accelerating Gaussian diffusions. *The Annals of Applied Probability* Vol. 3, (1993) pp. 897-913.
15. C.R. Hwang, S.Y. Hwang-Ma and S.J. Sheu, Accelerating diffusions, *The Annals of Applied Probability*, Vol 15, No. 2, (2005), pp. 1433-1444.
16. T. Lelievre, F. Nier and G.A. Pavliotis, Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion, *Journal of Statistical Physics*, 152(2), 237-274, (2013)
17. K.L. Mergessen and R.L. Tweedie, Rates of convergence of the Hastings and Metropolis algorithms, *Annals of Statistics*, Vol. 24, (1996), pp. 101-121.
18. R. Pinsky, The I-function for diffusion processes with boundaries, *The Annals of Probability*, Vol. 13, No. 3, (1985), pp. 676-692.
19. G.O. Roberts and J.S. Rosenthal, General state space Markov Chain and MCMC algorithms, *Probability Surveys*, Vol. 1, (2004), pp. 20-71.

References. V

20. Luc Rey-Bellet and K. Spiliopoulos, Irreversible Langevin samplers and variance reduction: a large deviations approach, Nonlinearity , Vol. 28, (2015), pp. 2081-2103.
21. Luc Rey-Bellet and K. Spiliopoulos, Variance reduction for irreversible Langevin samplers and diffusion on graphs, Electronic Communications in Probability, Vol. 20, (2015), no. 15, pp. 116.
22. Luc Rey-Bellet and K. Spiliopoulos, Improving the convergence of reversible samplers, Journal of Statistical Physics, Vol. 164, Issue 3, (2016), pp. 472-494
23. Jianfeng Lu and K. Spiliopoulos, Multiscale integrators for stochastic differential equations and irreversible Langevin samplers, (2016), submitted.

Thank You!!!!!