









Extreme-scale Mathematically-based Computational Chemistry

Variance reduction for applications in computational statistical physics

IRMAR – Séminaire de probabilités

Urbain Vaes urbain.vaes@inria.fr

MATHERIALS – Inria Paris & CERMICS – École des Ponts ParisTech

January 23, 2023

Mobility estimation for Langevin dynamics using control variates

- Background and problem statement
- Efficient mobility estimation
- Numerical experiments

Optimal importance sampling for overdamped Langevin dynamics

Background and problem statement Minimizing the asymptotic variance for one observable Minimizing the asymptotic variance for a class of observables

Part I: Mobility estimation for Langevin dynamics



Grigorios Pavliotis Imperial College Iondon

Department of Mathematics



Gabriel Stoltz



Reference: G. A. PAVLIOTIS, G. STOLTZ, and U. VAES. Mobility estimation for Langevin dynamics using control variates. arXiv preprint, 2022

Goals of computational statistical physics

- Computation of macroscopic properties from Newtonians atomistic models:
- Static properties, such as
 - the heat capacity and
 - the equations of state $P = P(\rho, T)$.
 - Dynamical properties, such as transport coefficients:
 - the viscosity;
 - the thermal conductivity;
 - the mobility of ions in solution.



- Solid line: experimental measurements^[1].
- Numerical microscope: used in physics, biology, chemistry.

^[1] https://webbook.nist.gov/chemistry/fluid/

Some background material on the Langevin equation

Consider the (one-particle) Langevin equation

$$\begin{cases} \mathrm{d}\mathbf{q}_t = \mathbf{p}_t \,\mathrm{d}t, \\ \mathrm{d}\mathbf{p}_t = -\nabla V(\mathbf{q}_t) \,\mathrm{d}t - \gamma \mathbf{p}_t \,\mathrm{d}t + \sqrt{2\gamma\beta^{-1}} \,\mathrm{d}\mathbf{W}_t, \end{cases} \quad (\mathbf{q}_0, \mathbf{p}_0) \sim \mu,$$

where γ is the friction, V is a periodic potential, and $\beta = \frac{1}{k_{\rm B}T}$.

The invariant probability measure is

$$\mu(\mathbf{q}, \mathbf{p}) = \frac{1}{Z} e^{-\beta H(\mathbf{q}, \mathbf{p})} = \frac{1}{Z} e^{-\beta \left(V(\mathbf{q}) + \frac{|\mathbf{p}|^2}{2}\right)}, \quad \text{on } \mathbf{T}^d \times \mathbf{R}^d.$$

The generator of the associated Markov semigroup

$$\left(\mathrm{e}^{\mathcal{L}t}\,\varphi\right)(\mathbf{q},\mathbf{p}) = \mathbf{E}\big(\varphi(\mathbf{q}_t,\mathbf{p}_t)\big|(\mathbf{q}_0,\mathbf{p}_0) = (\mathbf{q},\mathbf{p})\big)$$

is the following operator:

$$\mathcal{L} = (\mathbf{p} \cdot \nabla_{\mathbf{q}} - \nabla V(q) \cdot \nabla_{\mathbf{p}}) + \gamma \left(-\mathbf{p} \nabla_{\mathbf{p}} + \beta^{-1} \Delta_{\mathbf{p}} \right) =: \mathcal{L}_{\text{ham}} + \gamma \, \mathcal{L}_{\text{FD}}.$$

We denote by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ the norm and inner product of $L^2(\mu),$ and

$$L_0^2(\mu) = \left\{ \varphi \in L^2(\mu) : \langle \varphi, 1 \rangle = \mathbf{E}_{\mu} \varphi = 0 \right\}.$$

Consider Langevin dynamics with additional forcing in a direction e:

$$\begin{cases} \mathrm{d}\mathbf{q}_t = \mathbf{p}_t \,\mathrm{d}t, \\ \mathrm{d}\mathbf{p}_t = -\nabla V(\mathbf{q}_t) \,\mathrm{d}t + \eta \mathbf{e} \,\mathrm{d}t - \gamma \mathbf{p}_t \,\mathrm{d}t + \sqrt{2\gamma\beta^{-1}} \,\mathrm{d}\mathbf{W}_t \end{cases}$$

This dynamics admits a unique invariant probability distribution $\mu_{\eta} \in \mathcal{P}(\mathbf{T}^d \times \mathbf{R}^d)$.

Definition (Mobility)

The mobility in direction $\ensuremath{\mathbf{e}}$ is defined mathematically as

$$M_{\mathbf{e}} = \lim_{\boldsymbol{\eta} \to 0} \frac{1}{\boldsymbol{\eta}} \mathbf{E}_{\mu_{\boldsymbol{\eta}}} [\mathbf{e}^{\mathsf{T}} \mathbf{p}]$$

pprox factor relating the mean momentum to the strength of the inducing force.

• There is a symmetric mobility tensor \mathbf{M} such that $M_{\mathbf{e}} = \mathbf{e}^{\mathsf{T}} \mathbf{M} \mathbf{e}$.

Einstein's relation: $D = \beta^{-1}M$, with D the effective diffusion coefficient.

Effective diffusion

In

It is possible to show a functional central limit theorem for the Langevin dynamics:

$$\begin{split} & \varepsilon \widetilde{\mathbf{q}}_{s/\varepsilon^2} \xrightarrow[\varepsilon \to 0]{} \sqrt{2\mathbf{D}} \, \mathbf{W}_s \qquad \text{weakly on } C([0,\infty)), \qquad \widetilde{\mathbf{q}}_t := \mathbf{q}_0 + \int_0^t \mathbf{p}_s \, \mathrm{d}s \in \mathbf{R}^d. \end{split}$$
particular, $\widetilde{\mathbf{q}}_t / \sqrt{t} \xrightarrow[t \to \infty]{} \mathcal{N}(0, 2\mathbf{D})$ weakly.



Figure: Histogram of q_t/\sqrt{t} . The potential $V(q) = -\cos(q)/2$ is illustrated in the background.

Mathematical expression for the effective diffusion (dimension 1)

Expression of D in terms of the solution to a Poisson equation

The effective diffusion coefficient is given by where $D = \langle \phi, p \rangle$ and ϕ is the solution to

$$-\mathcal{L}\phi=p, \qquad \phi\in L^2_0(\mu):=ig\{u\in L^2(\mu): \langle u,1
angle=0ig\}.$$

Key idea of the proof: Apply Itô's formula to ϕ

0

$$d\phi(q_s, p_s) = -p_s \,ds + \sqrt{2\gamma\beta^{-1}} \,\frac{\partial\phi}{\partial p}(q_s, p_s) \,dW_s$$

and then rearrange:

$$\begin{split} \varepsilon(\widetilde{q}_{t/\varepsilon^2} - \widetilde{q}_0) &= \varepsilon \int_0^{t/\varepsilon^2} p_s \, \mathrm{d}s \\ &= \underbrace{\varepsilon(\phi(q_0, p_0) - \phi(q_{t/\varepsilon^2}, p_{t/\varepsilon^2}))}_{\to 0} + \underbrace{\sqrt{2\gamma\beta^{-1}\varepsilon} \int_0^{t/\varepsilon^2} \frac{\partial \phi}{\partial p}(q_s, p_s) \, \mathrm{d}W_s}_{\to \sqrt{2D}W_t} \, \mathrm{weakly \ by \ MCLT} \, . \end{split}$$

In the multidimensional setting, $D_{\mathbf{e}} = \left\langle \phi_{\mathbf{e}}, \mathbf{e}^{\mathsf{T}} \mathbf{p} \right\rangle$ with $-\mathcal{L}\phi_{\mathbf{e}} = \mathbf{e}^{\mathsf{T}} \mathbf{p}$

Langevin dynamics: underdamped and overdamped regimes^[2]



Figure: Langevin dynamics with friction $\gamma = 0.1$ (left) and $\gamma = 10$ (right)

• The underdamped limit as $\gamma \to 0$ is well understood in dimension 1 but not in the multi-dimensional setting. In dimension 1, it holds that

$$\phi = -\mathcal{L}^{-1}p = \gamma^{-1}\phi_{\text{und}} + \mathcal{O}(\gamma^{-1/2}).$$

• Overdamped limit: as $\gamma \to \infty$, the rescaled process $t \mapsto q_{\gamma t}$ converges weakly to the solution of the overdamped Langevin equation:

$$\dot{\mathbf{q}} = -\nabla V(q) + \sqrt{2\,\beta^{-1}}\,\dot{\mathbf{W}}.$$

[2] M. HAIRER and G. A. PAVLIOTIS. From ballistic to diffusive behavior in periodic potentials. J. Stat. Phys., 2008.

Scaling of the effective diffusion coefficient for Langevin dynamics^[3]

In dimension 1, $\lim_{\gamma \to 0} \gamma D^{\gamma} = D_{und} := \langle \phi_{und}, p \rangle$ and $\lim_{\gamma \to \infty} \gamma D^{\gamma} = D_{ovd}$.



[3] G. A. PAVLIOTIS and A. VOGIANNOU. Diffusive transport in periodic potentials: underdamped dynamics. Fluct. Noise Lett., 2008.

Mobility estimation for Langevin dynamics using control variates - Background and problem statement

Open question: surface diffusion when $\gamma \ll 1^{[4]}$

Applications:

- integrated circuits;
- catalysis.

 •
 •
 •
 •
 •
 •

 •
 •
 •
 •
 •
 •
 •

 •
 •
 •
 •
 •
 •
 •
 •

 •
 •
 •
 •
 •
 •
 •
 •
 •

 •
 •
 •
 •
 •
 •
 •
 •
 •

 •
 •
 •
 •
 •
 •
 •
 •
 •

 •
 •
 •
 •
 •
 •
 •
 •
 •

In dimension > 1, it does not hold that $\gamma D_{\mathbf{e}}^{\gamma} \xrightarrow[\gamma \to 0]{} D_{\mathrm{und}}$ when V is non-separable, e.g.

$$V(\mathbf{q}) = -\frac{1}{2} \left(\cos(q_1) + \cos(q_2) \right) - \frac{\delta}{\delta} \cos(q_1) \cos(q_2)$$

Open question: behavior of the effective diffusion coefficient when $\gamma \ll 1$?

$$D_{\mathbf{e}}^{\gamma} = \lim_{t \to \infty} \frac{\mathbf{E} \left[\left| \mathbf{e}^{\mathsf{T}} \mathbf{q}_t \right|^2 \right]}{2t} \sim \gamma^{-\sigma}, \qquad \sigma = ???$$

[4] Source of the video: https://en.wikipedia.org/wiki/Surface_diffusion

Mobility estimation for Langevin dynamics using control variates - Background and problem statement

Brief literature review

Open question:

How does $D_{\mathbf{e}}^{\gamma}$ behave when $\gamma \ll 1$ and d = 2?

Various answers are given in the literature:

- $D_{\mathbf{e}}^{\gamma} \propto \gamma^{-1/2}$ for specific potentials^[5];
- $D_{\mathbf{e}}^{\gamma} \propto \gamma^{-1/3}$ for specific potentials^[6];
- $D_{\mathbf{e}}^{\gamma} \propto \gamma^{-\sigma}$ with σ depending on the potential^[7].

Difficulty with $\gamma \ll 1$:

- Deterministic methods for the Poisson equation $-\mathcal{L}\phi_{\mathbf{e}} = \mathbf{e}^{\mathsf{T}}\mathbf{p}$ are ill-conditioned.
- Probabilistic methods are very slow to converge.

^[5] L. Y. CHEN, M. R. BALDAN, and S. C. YING. Surface diffusion in the low-friction limit: Occurrence of long jumps. Phys. Rev. B, 1996.

^[6] O. M. BRAUN and R. FERRANDO. Role of long jumps in surface diffusion. Phys. Rev. E, 2002.

^[7] J. ROUSSEL. Theoretical and Numerical Analysis of Non-Reversible Dynamics in Computational Statistical Physics. PhD thesis, Université Paris-Est, 2018.

Mobility estimation for Langevin dynamics using control variates

Background and problem statement

Efficient mobility estimation

Numerical experiments

Optimal importance sampling for overdamped Langevin dynamics

Background and problem statement

Minimizing the asymptotic variance for one observable

Minimizing the asymptotic variance for a class of observables

Numerical approaches for calculating the effective diffusion coefficient

Linear response approach:

$$D_{\mathbf{e}} = \lim_{\eta \to 0} \frac{1}{\beta \eta} \mathbf{E}_{\mu_{\eta}} (\mathbf{e}^{\mathsf{T}} \mathbf{p}).$$

where μ_{η} is the invariant distribution of the system with external forcing.

Einstein's relation:

$$D_{\mathbf{e}} = \lim_{t \to \infty} \frac{1}{2t} \mathbf{E}_{\mu} \Big[\left| \mathbf{e}^{\mathsf{T}} (\widetilde{\mathbf{q}}_{t} - \mathbf{q}_{0}) \right|^{2} \Big].$$

Deterministic method, e.g. Fourier/Hermite Galerkin, for the Poisson equation

$$-\mathcal{L}\phi_{\mathbf{e}} = \mathbf{e}^{\mathsf{T}}\mathbf{p}, \qquad D_{\mathbf{e}} = \langle \phi_{\mathbf{e}}, p \rangle.$$

• Green–Kubo formula: Since $-\mathcal{L}^{-1} = \int_0^\infty e^{t\mathcal{L}} dt$,

$$D_{\mathbf{e}} = \int -\mathcal{L}^{-1}(\mathbf{e}^{\mathsf{T}}\mathbf{p}) (\mathbf{e}^{\mathsf{T}}\mathbf{p}) d\mu = \int_{0}^{\infty} \int e^{t\mathcal{L}}(\mathbf{e}^{\mathsf{T}}\mathbf{p})(\mathbf{e}^{\mathsf{T}}\mathbf{p}) d\mu dt$$
$$= \int_{0}^{\infty} \mathbf{E}_{\mu} ((\mathbf{e}^{\mathsf{T}}\mathbf{p}_{0})(\mathbf{e}^{\mathsf{T}}\mathbf{p}_{t})) dt.$$

Mobility estimation for Langevin dynamics using control variates - Efficient mobility estimation

Consider the following estimator of the effective diffusion coefficient D_e :

$$u(T) = \frac{\left|\mathbf{e}^{\mathsf{T}}(\widetilde{\mathbf{q}}_{T} - \widetilde{\mathbf{q}}_{0})\right|^{2}}{2T}, \qquad (\mathbf{q}_{0}, \mathbf{p}_{0}) \sim \mu.$$

Bias of this estimator:

$$\mathbf{E}[u(T)] = D_{\mathbf{e}} - \int_0^\infty \left\langle e^{t\mathcal{L}}(\mathbf{e}^{\mathsf{T}}\mathbf{p}), \mathbf{e}^{\mathsf{T}}\mathbf{p} \right\rangle \min\left\{1, \frac{t}{T}\right\} \, \mathrm{d}t.$$

Using the decay estimate for the semigroup^[8]

$$\left\| \mathrm{e}^{t\mathcal{L}} \right\|_{\mathcal{B}\left(L_0^2(\mu) \right)} \leq L \, \mathrm{e}^{-\ell \min\{\gamma, \gamma^{-1}\}t},$$

we deduce

$$|\mathbf{E}[u(T)] - D_{\mathbf{e}}| \le \frac{C\max\{\gamma^2, \gamma^{-2}\}}{T}.$$

^[8] J. ROUSSEL and G. STOLTZ. Spectral methods for Langevin dynamics and associated error estimates. ESAIM: Math. Model. Numer. Anal., 2018.

Variance of the estimator u(T) for large T

For $T\gg 1,$ it holds approximately that

$$\frac{\mathbf{e}^{\mathsf{T}}(\widetilde{\mathbf{q}}_{T}-\widetilde{\mathbf{q}}_{0})}{\sqrt{2T}}\sim\mathcal{N}(0,D_{\mathbf{e}})\qquad \rightsquigarrow\qquad \frac{u(T)}{D_{\mathbf{e}}}=\frac{\left|\mathbf{e}^{\mathsf{T}}(\widetilde{\mathbf{q}}_{T}-\widetilde{\mathbf{q}}_{0})\right|^{2}}{2D_{\mathbf{e}}T}\sim\chi^{2}(1).$$

Therefore, we deduce

$$\lim_{T \to \infty} \mathbf{V} \big[u(T) \big] = 2D_{\mathbf{e}}^2.$$

The relative standard deviation (asymptotically as $T o \infty$) is therefore

$$\lim_{T \to \infty} \frac{\sqrt{\mathbf{V}\big[u(T)\big]}}{\mathbf{E}\big[u(T)\big]} = \sqrt{2} \qquad \rightsquigarrow \text{ independent of } \gamma.$$

Scaling of the mean square error when using J realizations

Assuming an asymptotic scaling as $\gamma^{-\sigma}$ of $D_{\mathbf{e}},$ we have

$$\forall \gamma \in (0,1), \qquad \frac{\text{MSE}}{D_{\mathbf{e}}^2} \le \frac{C}{\gamma^{4-2\sigma}T^2} + \frac{2}{J}$$

Let $\phi_{\mathbf{e}}$ denote the solution to the Poisson equation,

$$-\mathcal{L}\phi_{\mathbf{e}}(\mathbf{q},\mathbf{p}) = \mathbf{e}^{\mathsf{T}}\mathbf{p}, \qquad \phi_{\mathbf{e}} \in L_0^2(\mu)$$

By Itô's formula, we obtain

$$\phi_{\mathbf{e}}(\mathbf{q}_T, \mathbf{p}_T) - \phi_{\mathbf{e}}(\mathbf{q}_0, \mathbf{p}_0) = -\int_0^T \mathbf{e}^\mathsf{T} \mathbf{p}_t \, \mathrm{d}t + \sqrt{2\gamma\beta^{-1}} \int_0^T \nabla_{\mathbf{p}} \phi_{\mathbf{e}}(\mathbf{q}_t, \mathbf{p}_t) \cdot \mathrm{d}\mathbf{W}_t.$$

Therefore if $\psi_{\mathbf{e}}$ denote an approximation of $\phi_{\mathbf{e}},$ then

$$\mathbf{e}^{\mathsf{T}}(\widetilde{\mathbf{q}}_{T} - \widetilde{\mathbf{q}}_{0}) = \int_{0}^{T} \mathbf{e}^{\mathsf{T}} \mathbf{p}_{t} \, \mathrm{d}t$$
$$\approx -\psi_{\mathbf{e}}(\mathbf{q}_{T}, \mathbf{p}_{T}) + \psi_{\mathbf{e}}(\mathbf{q}_{0}, \mathbf{p}_{0}) + \sqrt{2\gamma\beta^{-1}} \int_{0}^{T} \nabla_{\mathbf{p}} \psi_{\mathbf{e}}(\mathbf{q}_{t}, \mathbf{p}_{t}) \cdot \mathrm{d}\mathbf{W}_{t} =: \xi_{T}.$$

which suggests the improved estimator

$$v(T) = \frac{\left|\mathbf{e}^{\mathsf{T}}(\widetilde{\mathbf{q}}_{T} - \widetilde{\mathbf{q}}_{0})\right|^{2}}{2T} - \left(\frac{\left|\xi_{T}\right|^{2}}{2T} - \lim_{T \to \infty} \mathbf{E}\left[\frac{\left|\xi_{T}\right|^{2}}{2T}\right]\right).$$

Smaller bias if $-\mathcal{L}\psi_{\mathbf{e}} \approx \mathbf{e}^{\mathsf{T}}\mathbf{p}$:

$$\left|\mathbf{E}[v(T)] - D_{\mathbf{e}}^{\gamma}\right| \leq \frac{L \max\{\gamma^{2}, \gamma^{-2}\}}{T\ell^{2}} \left\|\mathbf{e}^{\mathsf{T}}\mathbf{p} + \mathcal{L}\psi_{\mathbf{e}}\right\| \left(\beta^{-1/2} + \left\|\mathcal{L}\psi_{\mathbf{e}}\right\|\right).$$

Smaller variance:

$$\mathbf{V}[v(T)] \leq C \left(T^{-1} \|\boldsymbol{\phi}_{\mathbf{e}} - \boldsymbol{\psi}_{\mathbf{e}}\|_{L^{4}(\mu)}^{2} + \gamma \|\nabla_{\mathbf{p}}\boldsymbol{\phi}_{\mathbf{e}} - \nabla_{\mathbf{p}}\boldsymbol{\psi}_{\mathbf{e}}\|_{L^{4}(\mu)}^{2}\right) \\ \times \left(T^{-1} \|\boldsymbol{\phi}_{\mathbf{e}} + \boldsymbol{\psi}_{\mathbf{e}}\|_{L^{4}(\mu)}^{2} + \gamma \|\nabla_{\mathbf{p}}\boldsymbol{\phi}_{\mathbf{e}} + \nabla_{\mathbf{p}}\boldsymbol{\psi}_{\mathbf{e}}\|_{L^{4}(\mu)}^{2}\right).$$

Construction of ψ_{e} in the one-dimensional setting. We consider two approaches:

- Approximate the solution to the Poisson equation by a Galerkin method.
- Use asymptotic result for the Poisson equation:

$$\gamma \phi \xrightarrow[\gamma \to 0]{L^2(\mu)} \phi_{\mathrm{und}},$$

which suggests letting $\psi = \phi_{\rm und} / \gamma$.

We consider the potential

$$V(\mathbf{q}) = -\frac{1}{2} \Big(\cos(q_1) + \cos(q_2) \Big) - \frac{\delta}{\delta} \cos(q_1) \cos(q_2).$$

For this potential, \mathbf{D} is isotropic \rightsquigarrow sufficient to consider $\mathbf{e} = (1, 0)$,

$$D_{(1,0)} = \langle \phi_{(1,0)}, p_1 \rangle, \qquad -\mathcal{L}\phi_{(1,0)}(\mathbf{q}, \mathbf{p}) = p_1.$$

If $\delta = 0$, then the solution is $\phi_{(1,0)}(\mathbf{q}, \mathbf{p}) = \phi_{1\mathrm{D}}(q_1, p_1)$, where $\phi_{1\mathrm{D}}$ solves

$$-\mathcal{L}_{1D}\phi_{1D}(q,p) = p, \qquad V_{1D}(q) = \frac{1}{2}\cos(q).$$

• We take $\psi_{(1,0)}(\mathbf{q},\mathbf{p}) = \psi_{1\mathrm{D}}(q_1,p_1)$, where $\psi_{1\mathrm{D}} \approx \phi_{1\mathrm{D}}$.

Mobility estimation for Langevin dynamics using control variates

- Background and problem statement
- Efficient mobility estimation
- Numerical experiments

Optimal importance sampling for overdamped Langevin dynamics

- Background and problem statement
- Minimizing the asymptotic variance for one observable
- Minimizing the asymptotic variance for a class of observables



Numerical experiments for the one-dimensional case (2/2)



Figure: Evolution of the sample mean and standard deviation, estimated from J = 5000 realizations for $\gamma = 10^{-3}$.



• Variance reduction is possible if $|\delta| / \gamma \ll 1$;

Control variates are not very useful as $\gamma \to 0$ and δ is fixed...



In this part, we presented

- a variance reduction approach for efficiently estimating the mobility;
- numerical results showing that the scaling of the mobility is not universal.

Perspectives for future work:

- Use alternative methods (PINNs, Gaussian processes) to solve the Poisson equation;
- Study and improve variance reduction approaches for other transport coefficients.

Part II: Optimal importance sampling for overdamped Langevin dynamics



Martin Chak



Sorbonne Université



Tony Lelièvre



CERMICS & Inria



Gabriel Stoltz



CERMICS & Inria

Mobility estimation for Langevin dynamics using control variates

- Background and problem statement
- Efficient mobility estimation
- Numerical experiments

Optimal importance sampling for overdamped Langevin dynamics

Background and problem statement

Minimizing the asymptotic variance for one observable

Minimizing the asymptotic variance for a class of observables

The sampling problem

Objective of the sampling problem

Calculate averages of the form

$$\mu(f) := \int_{\mathbf{T}^d} f \,\mathrm{d}\mu, \qquad \mu := \frac{\mathrm{e}^{-V}}{Z}, \qquad Z := \int_{\mathbf{T}^d} \mathrm{e}^{-V} \,.$$

Often in applications:

- The dimension *d* is large;
- The normalization constant Z is unknown;
- We cannot generate i.i.d. samples from μ .

Markov chain Monte Carlo (MCMC) approach:

$$I := \mu(f) \approx \mu^T(f) := \frac{1}{T} \int_0^T f(Y_t) \,\mathrm{d}t$$

for a Markov process $(Y_t)_{t\geq 0}$ that is ergodic with respect to μ .

Example: overdamped Langevin dynamics

$$\mathrm{d}Y_t = -\nabla V(Y_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \qquad Y_0 = y_0.$$

Importance sampling in the MCMC context

If $(X_t)_{t\geq 0}$ is a Markov process ergodic with respect to

$$\mu_U = \frac{\mathrm{e}^{-V-U}}{Z_U}, \qquad Z_U = \int_{\mathbf{T}^d} \mathrm{e}^{-V-U},$$

then $I = \mu(f)$ may be approximated by

$$\mu_U^T(f) := \frac{\frac{1}{T} \int_0^T (f e^U)(X_t) dt}{\frac{1}{T} \int_0^T (e^U)(X_t) dt}.$$

Markov process: overdamped Langevin dynamics

$$\mathrm{d}X_t = -\nabla (V+U)(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \qquad X_0 = x_0.$$

Asymptotic variance: Under appropriate conditions, it holds that

$$\sqrt{T}\left(\mu_U^T(f) - I\right) \xrightarrow[T \to \infty]{\text{Law}} \mathcal{N}\left(0, \sigma_f^2[U]\right).$$

Objective

Find U such that the asymptotic variance $\sigma_f^2[U]$ is minimized.

Background: importance sampling in the i.i.d. setting (1/2)

Given i.i.d. samples $\{X^1, X^2, \dots\}$ from μ_U , we define

$$\mu_U^N(f) := \frac{\sum_{n=1}^N (f e^U)(X^n)}{\sum_{n=1}^N (e^U)(X^n)} = I + \frac{\frac{1}{N} \sum_{n=1}^N ((f-I) e^U)(X^n)}{\frac{1}{N} \sum_{n=1}^N (e^U)(X^n)},$$

Numerator: by the central limit theorem,

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left((f-I) e^{U} \right) (X^{n}) \xrightarrow[N \to \infty]{\text{Law}} \mathcal{N} \left(0, \int_{\mathbf{T}^{d}} \left| (f-I) e^{U} \right|^{2} d\mu_{U} \right)$$

Denominator: by the strong law of large numbers,

$$\frac{1}{N} \sum_{n=1}^{N} \left(\mathbf{e}^{U} \right) (X^{n}) \xrightarrow[N \to \infty]{\text{a.s.}} \frac{Z}{Z_{U}}.$$

Therefore, by Slutsky's theorem,

$$\sqrt{N} \left(\mu_U^N(f) - I \right) \xrightarrow[T \to \infty]{\text{Law}} \mathcal{N} \left(0, s_f^2[U] \right), \qquad s_f^2[U] := \frac{Z_U^2}{Z^2} \int_{\mathbf{T}^n} \left| (f - I) \, \mathrm{e}^U \right|^2 \mathrm{d}\mu_U.$$

Optimal importance sampling for overdamped Langevin dynamics - Background and problem statement

By the Cauchy-Schwarz inequality, it holds that

$$s_{f}^{2}[U] \geq \frac{Z_{U}^{2}}{Z^{2}} \left(\int_{\mathbf{T}^{d}} |f - I| e^{U} d\mu_{U} \right)^{2} = \frac{1}{Z^{2}} \left(\int_{\mathbf{T}^{d}} |f - I| e^{-V} \right)^{2},$$

with equality when $|f - I| e^U$ is constant.

Optimal importance distribution

The optimal μ_U in the i.i.d. setting is

$$\mu_U \propto |f - I| e^{-V}$$

Objectives:

- Is there a counterpart of this formula in the MCMC setting?
- If not, can we approximate the optimal distribution numerically?

Mobility estimation for Langevin dynamics using control variates

- Background and problem statement
- Efficient mobility estimation
- Numerical experiments

Optimal importance sampling for overdamped Langevin dynamics

Background and problem statement

Minimizing the asymptotic variance for one observable

Minimizing the asymptotic variance for a class of observables

Let \mathcal{L}_U denote the generator of the Markov semigroup associated to the modified potential;

$$\mathcal{L}_U = -\nabla (V + U) \cdot \nabla + \Delta.$$

Limit theorem

Under appropriate conditions, it holds that

$$\sqrt{T}\left(\mu_U^T(f) - I\right) \xrightarrow[T \to \infty]{\text{Law}} \mathcal{N}\left(0, \sigma_f^2[U]\right).$$

The asymptotic variance is given by

$$\sigma_f^2[U] = \frac{2Z_U^2}{Z^2} \int_{\mathbf{T}^d} \phi_U(f-I) \, \mathrm{e}^U \, \mathrm{d}\mu_U,$$

where ϕ_U is the unique solution in $H^1(\mu_U) \cap L^2_0(\mu_U)$ to

$$-\mathcal{L}_U\phi_U = (f-I)\,\mathrm{e}^U\,.$$

Main ideas of the proof: central limit theorem for martingales, Slutsky's theorem.

In dimension one, it holds that

$$\sigma_f^2[U] \geq \frac{2}{Z^2} \inf_{A \in \mathbf{R}} \left(\int_{\mathbf{T}} \left| F(x) + A \right| \mathrm{d}x \right)^2.$$

where

$$F(x) := \int_0^x (f(\xi) - I) e^{-V(\xi)} d\xi.$$

This inequality (1) is an equality for

$$U(x) = U_*(x) = -V(x) - \ln|F(x) + A_*|,$$

where A_* is the constant achieving the infimum in (1).

- The potential U_* is generally singular: impractical for numerics...
- The lower bound in (1) can be approached by a smooth U.

(1)

Example (1/2)

Assume that V = 0 and $f(x) = \cos(x)$.



\rightsquigarrow The optimal potential "divides" the domain into two parts.

Optimal importance sampling for overdamped Langevin dynamics - Minimizing the asymptotic variance for one observable

Example (2/2)

Assume that $V(x) = 5\cos(2x)$ and $f(x) = \sin(x)$. The target measure is multimodal.



Variance reduction by a factor > 1000!

Optimal importance sampling for overdamped Langevin dynamics - Minimizing the asymptotic variance for one observable

Finding the optimal U in the multidimensional setting

Proposition (Functional derivative of the asymptotic variance)

Let ϕ_U denote the solution to

$$-\mathcal{L}_U\phi_U = (f-I)\,\mathrm{e}^U\,.\tag{2}$$

Under appropriate conditions, it holds for all $\delta U \in C^{\infty}(\mathbf{T}^d)$ that

$$\frac{1}{2} \mathrm{d}\sigma_{f}^{2}[U] \cdot \delta U := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\sigma_{f}^{2}[U + \varepsilon \delta U] - \sigma_{f}^{2}[U] \right) \\
= \frac{Z_{U}^{2}}{Z^{2}} \int_{\mathbf{T}^{d}} \delta U \left(\left| \nabla \phi_{U} \right|^{2} - \int_{\mathbf{T}^{d}} \left| \nabla \phi_{U} \right|^{2} \mathrm{d}\mu_{U} \right) \mathrm{d}\mu_{U}. \tag{3}$$

Steepest descent approach:

- Solve the Poisson equation (2) numerically;
- Construct an ascent direction G for σ_f^2 using (3), e.g. $\delta U = |\nabla \phi_U|^2$;
- Perform a step in this direction: $U \leftarrow U \eta G$;
- Repeat until convergence.

Corollary (No smooth minimizer)

Unless f is constant, there is no perturbation potential $U \in C^{\infty}(\mathbf{T}^n)$ that is a critical point of $\sigma_f^2[U]$.

Proof. Assume by contradiction that U_* is smooth critical point. Then

$$0 = \frac{1}{2} \mathrm{d}\sigma_f^2[U_*] \cdot \delta U = \frac{Z_U^2}{Z^2} \int_{\mathbf{T}^d} \delta U \bigg(|\nabla \phi_{U_*}|^2 - \int_{\mathbf{T}^d} |\nabla \phi_{U_*}|^2 \,\mathrm{d}\mu_{U_*} \bigg) \,\mathrm{d}\mu_{U_*},$$

for all $\delta U \in C^{\infty}(\mathbf{T}^d)$.

- Therefore, it must hold that $|\nabla \phi_{U_*}|^2 = C$ is constant.
- Since ϕ_{U_*} is a smooth function, there is $x \in \mathbf{T}^d$ such that $\nabla \phi_{U_*}(x) = 0$.
- Consequently C = 0 and so $\nabla \phi_{U_*} = 0$: contradiction because then $\mathcal{L}_{U_*} \phi_{U_*} = 0$.

 \rightsquigarrow The optimal perturbation potential is not convenient in practice. . .

Example (1/3)

Assume that V = 0 and $f(x) = \sin(x_1) + \sin(x_2)$.



Figure: Optimal total potential (left) together with the solution to the associated Poisson equation (right).

\rightsquigarrow The domain is again divided into subdomains that suffice for estimating I.

Example (2/3): multimodal target e^{-V}

Assume that $V(x) = 2\cos(x_1) - \cos(x_2)$ and $f(x) = \sin(x_1)$.



Variance reduction by a factor $\approx 6!$

Optimal importance sampling for overdamped Langevin dynamics - Minimizing the asymptotic variance for one observable

In this case we consider that

$$V(x) = \exp\left(\cos(x_1)\sin(x_2) + \frac{1}{5}\cos(3x_1)\right), \qquad f(x) = \sin\left(x_1 + \cos(x_2)\right)^3.$$



Figure: Unperturbed potential V (left) and optimal potential V + U (right).

Mobility estimation for Langevin dynamics using control variates

- Background and problem statement
- Efficient mobility estimation
- Numerical experiments

Optimal importance sampling for overdamped Langevin dynamics

- Background and problem statement
- Minimizing the asymptotic variance for one observable

Minimizing the asymptotic variance for a class of observables

Assume that the observables are well described by a Gaussian random field

$$f = \sum_{j=1}^{J} \sqrt{\lambda_j} u_j f_j, \qquad u_j \sim \mathcal{N}(0, 1), \qquad \lambda_j \in (0, \infty).$$

Question: can we find U such that $\sigma^2[U] := \mathbf{E}(\sigma_f^2[U])$ is minimized?

It holds that

$$\sigma^2[U] = \sum_{j=1}^J \lambda_j \sigma_{f_j}^2.$$

 \blacksquare The functional derivative of $\sigma^2[U]$ is given by

$$\frac{1}{2}\mathrm{d}\sigma^{2}[U]\cdot\delta U = \frac{Z_{U}^{2}}{Z^{2}}\int_{\mathbf{T}^{d}}\left(\delta U - \int_{\mathbf{T}^{d}}\delta U\,\mathrm{d}\mu_{U}\right)\left(\sum_{j=1}^{J}\lambda_{j}|\nabla\phi_{j}|^{2}\right)\,\mathrm{d}\mu_{U}.$$

The steepest descent approach can be employed in this case too!

Example

Here
$$V(x) = 2\cos(2x_1) - \cos(x_2)$$
 and $f \sim \mathcal{N}(0, (\Delta + \mathcal{I})^{-1})$.



Figure: Potential V (left) and optimal potential V + U (right).

Variance reduction in this case: 20%

In this part,

- We studied an importance sampling approach for the overdamped Langevin dynamics.
- We proposed an approach for calculating the optimal perturbation potential.

Perspectives:

- Solving the Poisson equation accurately is not possible in high dimension.
- Application to high-dimensional systems:

 $U(x) = U(\xi(x)), \quad \xi \text{ reaction coordinate.}$

Thank you for your attention!

As $\gamma \rightarrow 0$, the Hamiltonian of the rescaled process

 $\begin{cases} q_{\gamma}(t) = q(t/\gamma), \\ p_{\gamma}(t) = p(t/\gamma), \end{cases}$

converges weakly to a diffusion process on a graph.



Ergodic theorem^[9]: for an observable $\varphi \in L^1(\mu)$,

$$\widehat{\varphi}_t = \frac{1}{t} \int_0^t \varphi(\mathbf{q}_s, \mathbf{p}_s) \,\mathrm{d}s \xrightarrow[t \to \infty]{a.s.} \mathbf{E}_{\mu} \varphi.$$

Central limit theorem^[10]: If the following Poisson equation has a solution $\phi \in L^2(\mu)$,

$$-\mathcal{L}\phi=\varphi-\mathbf{E}_{\mu}\varphi,$$

then a central limit theorem holds:

$$\sqrt{t} \left(\widehat{\varphi}_t - \mathbf{E}_{\mu} \varphi \right) \xrightarrow[t \to \infty]{\text{Law}} \mathcal{N}(0, \sigma_{\varphi}^2), \qquad \sigma_{\varphi}^2 = \langle \phi, \varphi - \mathbf{E}_{\mu} \varphi \rangle.$$

Connection with effective diffusion: Apply this result with $\varphi(\mathbf{q}, \mathbf{p}) = \mathbf{e}^{\mathsf{T}} \mathbf{p}$.

^[9] W. KLIEMANN. Recurrence and invariant measures for degenerate diffusions. Ann. Probab., 1987.

^[10] R. N. BHATTACHARYA. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. Z. Wahrsch. Verw. Gebiete, 1982.