Optimal friction for Langevin sampling

Martin Chak
joint work w/ Nikolas Kantas, Tony Lelièvre, Grigorios Pavliotis

14th International conference on Monte Carlo methods and applications
28th June 2023

Introduction

Given

- a smooth potential function $V : \mathbb{R}^n \to \mathbb{R}$ and its gradient $\nabla V : \mathbb{R}^n \to \mathbb{R}^n$, which can be evaluated at any point in $\mathbb{R}^n$ at a cost, and
- an everywhere differentiable observable function $L^2(e^{-V}) \ni f : \mathbb{R}^n \to \mathbb{R}$,

approximate the integral

$$\pi(f) := \int_{\mathbb{R}^n} f\left(\frac{e^{-V}}{\int_{\mathbb{R}^n} e^{-V}}\right).$$

MCMC Solution (for large n): Use stochastic dynamics, e.g. (discretizations of)

$$\begin{aligned} \text{overdamped Langevin:} \qquad & dq_t = -\nabla V(q_t)dt + \sqrt{2}dW_t, \\ \text{or (underdamped) Langevin:} \qquad & dq_t = p_t dt, \\ & dp_t = -\nabla V(q_t)dt - \Gamma p_t dt + \sqrt{2\Gamma}dW_t \end{aligned}$$

for any positive symmetric definite $\Gamma \in \mathbb{R}^{n\times n}$ and set $\pi_T(f) := \frac{1}{T}\int_0^T f(q_t)dt$. Overdamped Langevin dynamics has the invariant measure $\pi(dq) \propto e^{-V(q)}dq$ and underdamped Langevin has the invariant measure $\tilde{\pi}(dqdp) \propto e^{-V(q)-\frac{p^2}{2}}dqdp$.

Introduction

Given

- a smooth potential function $V : \mathbb{R}^n \to \mathbb{R}$ and its gradient $\nabla V : \mathbb{R}^n \to \mathbb{R}^n$, which can be evaluated at any point in $\mathbb{R}^n$ at a cost, and
- an everywhere differentiable observable function $L^2(e^{-V}) \ni f : \mathbb{R}^n \to \mathbb{R}$,

approximate the integral

$$\pi(f) := \int_{\mathbb{R}^n} f\left(\frac{e^{-V}}{\int_{\mathbb{R}^n} e^{-V}}\right).$$

MCMC Solution (for large n): Use stochastic dynamics, e.g. (discretizations of)

$$\begin{aligned}
\text{overdamped Langevin:} \qquad & dq_t = -\nabla V(q_t)dt + \sqrt{2}dW_t, \\
\text{or (underdamped) Langevin:} \qquad & dq_t = p_t dt, \\
& dp_t = -\nabla V(q_t)dt - \Gamma p_t dt + \sqrt{2\Gamma}dW_t
\end{aligned}$$

for any positive symmetric definite $\Gamma \in \mathbb{R}^{n\times n}$ and set $\pi_T(f) := \frac{1}{T}\int_0^T f(q_t)dt$.
Overdamped Langevin dynamics has the invariant measure $\pi(dq) \propto e^{-V(q)}dq$ and underdamped Langevin has the invariant measure $\tilde{\pi}(dqdp) \propto e^{-V(q)-\frac{p^2}{2}}dqdp$.

Underdamped Langevin $dq_t = p_t dt,\ dp_t = -\nabla V(q_t)dt - \Gamma p_t dt + \sqrt{2\Gamma} dW_t$

In order to obtain some measure of 'speed' at which
$\pi_T(f) := \frac{1}{T}\int_0^T f(q_t)dt \to \int_{\mathbb{R}^n} f\left(\frac{e^{-V}}{\int_{\mathbb{R}^n} e^{-V}}\right) = \pi(f)$ occurs as $T \to \infty$, use the fact that $\frac{1}{\sqrt{T}}\int_0^T (f(q_t) - \pi(f))dt \to \mathcal{N}(0,\sigma^2)$ in distribution as $T \to \infty$ for some $\sigma^2 > 0$:

Claim: $\sigma^2 = 2\int_{\mathbb{R}^n} \nabla_p \phi^\top \Gamma \nabla_p \phi d\tilde{\pi}$, where ϕ solves

$$-\mathcal{L}\phi := (-p^\top \nabla_q + \nabla V(q)^\top \nabla_p + \Gamma p^\top \nabla_p - \nabla_p^\top \Gamma \nabla_p)\phi = f - \pi(f).$$

'Proof': Suppose there exists a well-behaved solution $\phi : \mathbb{R}^{2n} \to \mathbb{R}$ and use Itô's lemma to obtain

$$\begin{aligned}\phi(q_t,p_t) &= \phi(q_0,p_0) + \int_0^t \mathcal{L}\phi(q_s,p_s)ds + \int_0^t \nabla_p\phi(q_s,p_s)^\top \sqrt{2\Gamma} dW_s \\ &= \phi(q_0,p_0) - \int_0^t (f(q_s) - \pi(f))ds + \int_0^t \nabla_p\phi(q_s,p_s)^\top \sqrt{2\Gamma} dW_s\end{aligned}$$

so that, using Itô's isometry and ergodicity, the variance of $\frac{1}{\sqrt{t}}\int_0^t (f(q_s) - \pi(f))ds$ as $t \to \infty$ is

$$\sigma^2 = \lim_{t\to\infty} \frac{2}{t}\mathbb{E}\int_0^t \nabla_p\phi^\top(q_s,p_s)\Gamma\nabla_p\phi(q_s,p_s)ds = 2\int_{\mathbb{R}^n} \nabla_p\phi^\top \Gamma \nabla_p \phi d\tilde{\pi}.$$

In order to obtain some measure of 'speed' at which $\pi_T(f) := \frac{1}{T}\int_0^T f(q_t)dt \to \int_{\mathbb{R}^n} f\left(\frac{e^{-V}}{\int_{\mathbb{R}^n} e^{-V}}\right) = \pi(f)$ occurs as $T \to \infty$, use the fact that $\frac{1}{\sqrt{T}}\int_0^T (f(q_t) - \pi(f))dt \to \mathcal{N}(0,\sigma^2)$ in distribution as $T \to \infty$ for some $\sigma^2 > 0$:

Claim: $\sigma^2 = 2\int_{\mathbb{R}^n} \nabla_p\phi^\top \Gamma \nabla_p \phi d\tilde{\pi}$, where ϕ solves

$$-\mathcal{L}\phi := (-p^\top \nabla_q + \nabla V(q)^\top \nabla_p + \Gamma p^\top \nabla_p - \nabla_p^\top \Gamma \nabla_p)\phi = f - \pi(f).$$

'Proof': Suppose there exists a well-behaved solution $\phi : \mathbb{R}^{2n} \to \mathbb{R}$ and use Itô's lemma to obtain

$$\begin{aligned}\phi(q_t, p_t) &= \phi(q_0, p_0) + \int_0^t \mathcal{L}\phi(q_s, p_s)ds + \int_0^t \nabla_p \phi(q_s, p_s)^\top \sqrt{2\Gamma}dW_s \\ &= \phi(q_0, p_0) - \int_0^t (f(q_s) - \pi(f))ds + \int_0^t \nabla_p \phi(q_s, p_s)^\top \sqrt{2\Gamma}dW_s\end{aligned}$$

so that, using Itô's isometry and ergodicity, the variance of $\frac{1}{\sqrt{t}}\int_0^t (f(q_s) - \pi(f))ds$ as $t \to \infty$ is

$$\sigma^2 = \lim_{t\to\infty} \frac{2}{t}\mathbb{E}\int_0^t \nabla_p\phi^\top(q_s,p_s)\Gamma\nabla_p\phi(q_s,p_s)ds = 2\int_{\mathbb{R}^n} \nabla_p\phi^\top\Gamma\nabla_p\phi d\tilde{\pi}.$$

Underdamped Langevin $dq_t = p_t dt,\ dp_t = -\nabla V(q_t)dt - \Gamma p_t dt + \sqrt{2\Gamma}dW_t$

In order to obtain some measure of 'speed' at which $\pi_T(f) := \frac{1}{T}\int_0^T f(q_t)dt \to \int_{\mathbb{R}^n} f\left(\frac{e^{-V}}{\int_{\mathbb{R}^n} e^{-V}}\right) = \pi(f)$ occurs as $T \to \infty$, use the fact that $\frac{1}{\sqrt{T}}\int_0^T (f(q_t) - \pi(f))dt \to \mathcal{N}(0,\sigma^2)$ in distribution as $T \to \infty$ for some $\sigma^2 > 0$:

Claim: $\sigma^2 = 2\int_{\mathbb{R}^n} \nabla_p\phi^\top \Gamma \nabla_p \phi d\tilde{\pi}$, where ϕ solves

$$-\mathcal{L}\phi := (-p^\top\nabla_q + \nabla V(q)^\top \nabla_p + \Gamma p^\top \nabla_p - \nabla_p^\top \Gamma \nabla_p)\phi = f - \pi(f).$$

'Proof': Suppose there exists a well-behaved solution $\phi : \mathbb{R}^{2n} \to \mathbb{R}$ and use Itô's lemma to obtain

$$\begin{aligned}\phi(q_t,p_t) &= \phi(q_0,p_0) + \int_0^t \mathcal{L}\phi(q_s,p_s)ds + \int_0^t \nabla_p\phi(q_s,p_s)^\top\sqrt{2\Gamma}dW_s \\ &= \phi(q_0,p_0) - \int_0^t (f(q_s) - \pi(f))ds + \int_0^t \nabla_p\phi(q_s,p_s)^\top\sqrt{2\Gamma}dW_s\end{aligned}$$

so that, using Itô's isometry and ergodicity, the variance of $\frac{1}{\sqrt{t}}\int_0^t (f(q_s) - \pi(f))ds$ as $t \to \infty$ is

$$\sigma^2 = \lim_{t\to\infty} \frac{2}{t}\mathbb{E}\int_0^t \nabla_p\phi^\top(q_s,p_s)\Gamma\nabla_p\phi(q_s,p_s)ds = 2\int_{\mathbb{R}^n} \nabla_p\phi^\top\Gamma\nabla_p\phi d\tilde{\pi}.$$

Directional derivative of σ^2 with respect to Γ

To summarise: $\frac{1}{\sqrt{T}} \int_0^T (f(q_t) - \pi(f))dt \to \mathcal{N}\left(0, 2\int_{\mathbb{R}^n} \nabla_p \phi^\top \Gamma \nabla_p \phi d\tilde{\pi}\right)$ in distribution,
where $-\mathcal{L}\phi := (-p^\top \nabla_q + \nabla V(q)^\top \nabla_p + \Gamma p^\top \nabla_p - \nabla_p^\top \Gamma \nabla_p)\phi = f - \pi(f)$.

Theorem (Main result for underdamped Langevin dynamics)

$d\sigma^2.\delta\Gamma = -2\int \nabla_p \phi^\top \delta\Gamma \nabla_p \tilde{\phi} d\tilde{\pi}$, *where* $\tilde{\phi}(q,p) = \phi(q,-p)$.

The direction $\delta\Gamma = \int \nabla_p \phi \otimes \nabla_p \tilde{\phi} d\tilde{\pi}$ guarantees a decrease in asymptotic variance!

Monte Carlo expression for solution ϕ

Claim: $\phi(q,p) = \int_0^\infty \mathbb{E}^{(q,p)}[f(q_t) - \pi(f)]dt$. ($\mathbb{E}^{(q,p)}$ denotes $(q_0, p_0) = (q,p)$)
'Proof': By the Feynman-Kac representation formula,

$$-\mathcal{L}\int_0^\infty \mathbb{E}^{(q,p)}[f(q_t) - \pi(f)]dt = -\int_0^\infty \mathcal{L}\mathbb{E}^{(q,p)}[f(q_t) - \pi(f)]dt$$
$$= -\int_0^\infty \partial_t \mathbb{E}^{(q,p)}[f(q_t) - \pi(f)]dt = f(q) - \pi(f).$$

Monte Carlo expression for solution $\nabla_p \phi$

$$\nabla_p \phi(q,p) = \nabla_p \int_0^\infty \mathbb{E}[f(q_t) - \pi(f)]dt = \int_0^\infty \mathbb{E}[\nabla f(q_t)^\top \partial_{p_0} q_t]dt,$$

with initial condition $(q_0, p_0) = (q,p)$ for dynamics (q_t, p_t).

Directional derivative of σ^2 with respect to Γ

To summarise: $\frac{1}{\sqrt{T}} \int_0^T (f(q_t) - \pi(f))dt \to \mathcal{N}\left(0, 2\int_{\mathbb{R}^n} \nabla_p \phi^\top \Gamma \nabla_p \phi d\tilde{\pi}\right)$ in distribution,
where $-\mathcal{L}\phi := (-p^\top \nabla_q + \nabla V(q)^\top \nabla_p + \Gamma p^\top \nabla_p - \nabla_p^\top \Gamma \nabla_p)\phi = f - \pi(f)$.

Theorem (Main result for underdamped Langevin dynamics)

$d\sigma^2.\delta\Gamma = -2\int \nabla_p \phi^\top \delta\Gamma \nabla_p \tilde{\phi} d\tilde{\pi}$, where $\tilde{\phi}(q,p) = \phi(q,-p)$.

The direction $\delta\Gamma = \int \nabla_p \phi \otimes \nabla_p \tilde{\phi} d\tilde{\pi}$ guarantees a decrease in asymptotic variance!

Monte Carlo expression for solution ϕ

Claim: $\phi(q,p) = \int_0^\infty \mathbb{E}^{(q,p)}[f(q_t) - \pi(f)]dt$. ($\mathbb{E}^{(q,p)}$ denotes $(q_0, p_0) = (q,p)$)
'Proof': By the Feynman-Kac representation formula,

$$-\mathcal{L}\int_0^\infty \mathbb{E}^{(q,p)}[f(q_t) - \pi(f)]dt = -\int_0^\infty \mathcal{L}\mathbb{E}^{(q,p)}[f(q_t) - \pi(f)]dt$$
$$= -\int_0^\infty \partial_t \mathbb{E}^{(q,p)}[f(q_t) - \pi(f)]dt = f(q) - \pi(f).$$

Monte Carlo expression for solution $\nabla_p \phi$

$$\nabla_p \phi(q,p) = \nabla_p \int_0^\infty \mathbb{E}[f(q_t) - \pi(f)]dt = \int_0^\infty \mathbb{E}[\nabla f(q_t)^\top \partial_{p_0} q_t]dt,$$

with initial condition $(q_0, p_0) = (q,p)$ for dynamics (q_t, p_t).

Simulate (q_t, p_t) and $(\partial_{p_0} q_t, \partial_{p_0} p_t)$ and increment Γ by

$$\delta\Gamma = \int \nabla_p \phi \otimes \nabla_p \tilde{\phi} d\tilde{\pi}, \qquad \nabla_p \phi(q,p) = \int_0^\infty \mathbb{E}[\nabla f(q_t)^\top \partial_{p_0} q_t] dt,$$

where $(q_0, p_0) = (q, p)$.

1D quadratic potential $V(q) = \frac{5}{2} q^2$:

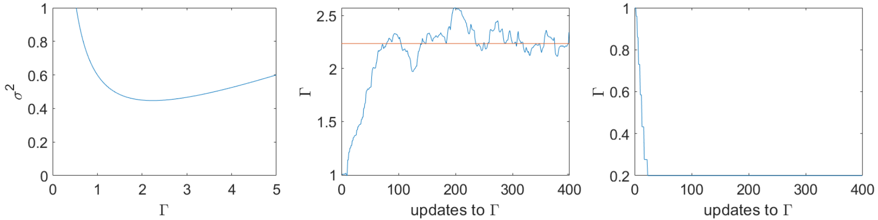

Figure: Left: Known relationship between σ^2 and Γ for $f(q) = \frac{1}{2} q^2$. Middle: changes to Γ for $f(q) = \frac{1}{2} q^2$. Right: for $f(q) = q$.

Simulate (q_t, p_t) and $(\partial_{p_0} q_t, \partial_{p_0} p_t)$ and increment Γ by

$$\delta\Gamma = \int \nabla_p \phi \otimes \nabla_p \tilde{\phi} d\tilde{\pi}, \qquad \nabla_p \phi(q,p) = \int_0^\infty \mathbb{E}[\nabla f(q_t)^\top \partial_{p_0} q_t] dt,$$

where $(q_0, p_0) = (q, p)$.

Bayesian inference for binary regression

- 2359 data points (no. of summands in potential)
- 642 features (no. of dimensions in the SDE)

Problem: to find the posterior mean, that is, $f_i(q) = q_i$.

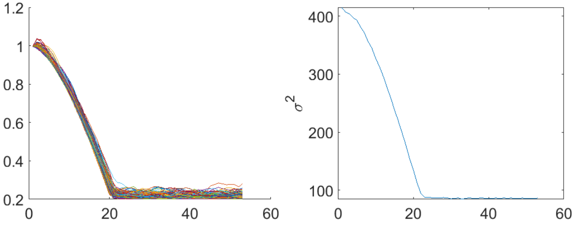

Figure: Left: Diagonal values of Γ over iterations. Right: Sum over i of estimated asymptotic variances for $f_i(q)$.

Optimal importance sampling for overdamped Langevin dynamics

joint work w/ Tony Lelièvre, Gabriel Stoltz, Urbain Vaes,

with acknowledgements to A. Duncan, G. Pavliotis

Optimal importance sampling for i.i.d. samples

Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. r.v.'s drawn from $\pi_U(dx) \propto e^{-V(x)-U(x)}dx$.
To estimate $I = \int f d\pi$, use that

$$\frac{\frac{1}{N}\sum_{i=1}^N f(X_n)e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^N e^{U(X_n)}} \to \frac{\int fe^U d\pi_U}{\int e^U d\pi_U} = I$$

almost surely as $N \to \infty$. The CLT and Slutsky's theorem gives

$$\sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^N f(X_n)e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^N e^{U(X_n)}} - I\right) = \frac{\frac{1}{\sqrt{N}}\sum_{n=1}^N (f(X_n)-I)e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^N e^{U(X_n)}}$$

$$\to \mathcal{N}\left(0, \frac{Z_U^2}{Z^2}\int |(f-I)e^U|^2 d\pi_U\right)$$

in distribution as $\mathbb{N} \to \infty$, where Z_U and Z are the normalization constants for $\pi_U(dx) \propto e^{-V(x)-U(x)}dx$ and $\pi(dx) \propto e^{-V(x)}dx$ respectively.
By Cauchy-Schwarz,

$$\left(\int |f-I|e^{-V}\right)^2 \leq \left(\int |f-I|^2 e^{U-V}\right)\left(\int e^{-U-V}\right) = Z_U^2 \int |(f-I)e^U|^2 d\pi_U,$$

with equality for $U = -\log|f-I|$.

Optimal importance sampling for i.i.d. samples

Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. r.v.'s drawn from $\pi_U(dx) \propto e^{-V(x)-U(x)}dx$.
To estimate $I = \int f d\pi$, use that

$$\frac{\frac{1}{N}\sum_{i=1}^{N} f(X_n)e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^{N} e^{U(X_n)}} \to \frac{\int fe^U d\pi_U}{\int e^U d\pi_U} = I$$

almost surely as $N\to\infty$. The CLT and Slutsky's theorem gives

$$\begin{aligned}\sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} f(X_n)e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^{N} e^{U(X_n)}} - I\right) &= \frac{\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(f(X_n)-I)e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^{N} e^{U(X_n)}} \\ &\to \mathcal{N}\left(0, \frac{Z_U^2}{Z^2}\int |(f-I)e^U|^2 d\pi_U\right)\end{aligned}$$

in distribution as $N\to\infty$, where Z_U and Z are the normalization constants for $\pi_U(dx) \propto e^{-V(x)-U(x)}dx$ and $\pi(dx)\propto e^{-V(x)}dx$ respectively.
By Cauchy-Schwarz,

$$\left(\int |f-I|e^{-V}\right)^2 \leq \left(\int |f-I|^2 e^{U-V}\right)\left(\int e^{-U-V}\right) = Z_U^2\int |(f-I)e^U|^2 d\pi_U,$$

with equality for $U = -\log|f-I|$.

Optimal importance sampling for i.i.d. samples

Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. r.v.'s drawn from $\pi_U(dx) \propto e^{-V(x)-U(x)}dx$.
To estimate $\pi(f) = \int f d\pi$, use that

$$\frac{\frac{1}{N}\sum_{i=1}^N f(X_n)e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^N e^{U(X_n)}} \to \frac{\int fe^U d\pi_U}{\int e^U d\pi_U} = \pi(f)$$

almost surely as $N \to \infty$. The CLT and Slutsky's theorem gives

$$\begin{aligned}\sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^N f(X_n)e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^N e^{U(X_n)}} - \pi(f)\right) &= \frac{\frac{1}{\sqrt{N}}\sum_{n=1}^N (f(X_n)-\pi(f))e^{U(X_n)}}{\frac{1}{N}\sum_{n=1}^N e^{U(X_n)}} \\ &\to \mathcal{N}\left(0, \frac{Z_U^2}{Z^2}\int |(f-\pi(f))e^U|^2 d\pi_U\right)\end{aligned}$$

in distribution as $N \to \infty$, where Z_U and Z are the normalization constants for $\pi_U(dx) \propto e^{-V(x)-U(x)}dx$ and $\pi(dx) \propto e^{-V(x)}dx$ respectively.
By Cauchy-Schwarz,

$$\left(\int |f-\pi(f)|e^{-V}\right)^2 \leq \left(\int |f-\pi(f)|^2 e^{U-V}\right)\left(\int e^{-U-V}\right) = Z_U^2 \int |(f-\pi(f))e^U|^2 d\pi_U,$$

with equality for $U = -\log|f-\pi(f)|$.

Overdamped Langevin $dq_t = -\nabla V(q_t)dt - \nabla U(q_t)dt + \sqrt{2}dW_t$

By the same ideas as before, there is an explicit CLT variance

$$\sqrt{T}\frac{\frac{1}{T}\int_0^T (f(q_t) - \pi(f))e^{U(q_t)}dt}{\frac{1}{T}\int_0^T e^{U(q_t)}dt} \to \mathcal{N}\left(0, \frac{2Z_U^2}{Z^2}\int |\nabla\phi|^2 d\pi_U\right),$$

where ϕ solves $-\mathcal{L}\phi = f - \pi(f)$.

Theorem (Explicit optimal U on $\mathbb{R}$ (when $U = U : \mathbb{R} \to \mathbb{R}$))

Under mild assumptions, the minimum asymptotic variance is achieved by

$$U(x) = -V(x) - \log\left|\int_{-\infty}^{x} (f(y) - \pi(f))e^{-V(y)}dy\right|.$$

Theorem (Directional derivative of $\sigma^2 = \frac{2Z_U^2}{Z^2}\int|\nabla\phi|^2 d\pi_U$ with respect to U)

$$d\sigma^2.\delta U = \frac{2Z_U^2}{Z^2}\int \delta U\left(|\nabla\phi|^2 - \int|\nabla\phi|^2 d\pi_U\right)d\pi_U.$$

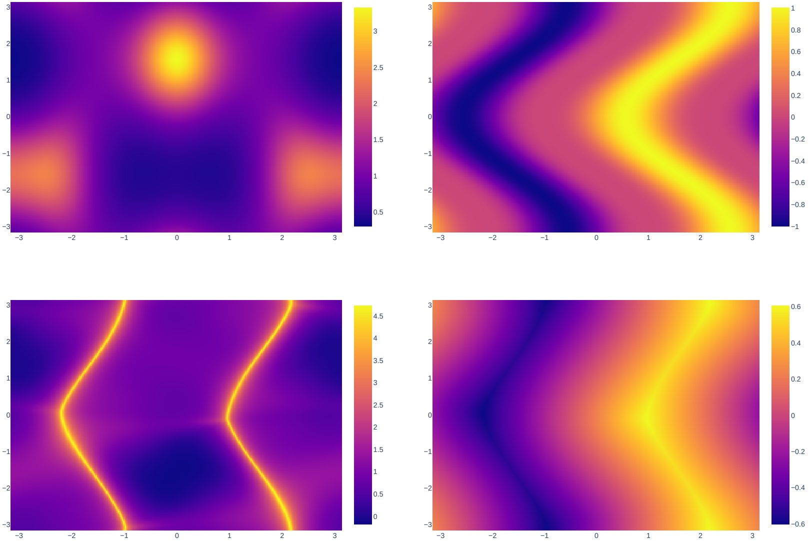

Figure: Unperturbed potential V (top left), observable f (top right), optimal potential $V + U$ (bottom left), and corresponding solution to the Poisson equation ϕ_U (bottom right).

Optimal friction for Langevin sampling

Martin Chak
joint work w/ Nikolas Kantas, Tony Lelièvre, Grigorios Pavliotis

14th International conference on Monte Carlo methods and applications
28th June 2023