Sampling probability measures on submanifolds

T. Lelièvre

CERMICS - Ecole des Ponts ParisTech & Matherials team - INRIA

Joint work with M. Rousset, G. Stoltz and W. Zhang

École des Ponts ParisTech

MATHerials

INVENTEURS DU MONDE NUMÉRIQUE

Séminaire Phimeca, 18th November 2021

Objective

Let us be given a submanifold of $\mathbb{R}^d$:

$$\mathcal{M} = \left\{ q \in \mathbb{R}^d,\, \xi(q) = 0 \right\}$$

where $\xi : \mathbb{R}^d \to \mathbb{R}^m$ is a given smooth function (with $m < d$) such that

$$G(q) = [\nabla \xi(q)]^T \, \nabla \xi(q) \in \mathbb{R}^{m \times m}$$

is an invertible matrix for all q in a neighborhood of $\mathcal{M}$. The objective is to sample the probability measure:

$$\nu(dq) = Z_\nu^{-1} \, \mathrm{e}^{-V(q)} \, \sigma_{\mathcal{M}}(dq), \qquad Z_\nu = \int_{\mathcal{M}} \mathrm{e}^{-V(q)} \, \sigma_{\mathcal{M}}(dq) < \infty,$$

where $\sigma_{\mathcal{M}}(dq)$ is the Riemannian measure on $\mathcal{M}$ induced by the scalar product $\langle \cdot, \cdot \rangle$ defined in the ambient space $\mathbb{R}^d$.

Motivation

Such problems arise in many contexts: constrained mechanical systems with noise, statistics, ...

One example is computational statistical physics: free energy calculations. If $X \sim \rho$ where $\rho(dq) = Z^{-1} \mathrm{e}^{-V(q)}\, dq$, then $\xi(X) \sim \xi_{\#}\rho$. Let us define $A : \mathbb{R}^m \to \mathbb{R}$ by:

$$\mathrm{e}^{-A(z)} dz = \xi_{\#}\rho(dz).$$

Then, using the co-area formula,

$$\nabla A(0) = \mathbb{E}_{\rho}(f(X)|\xi(X) = 0) = \int_{\mathcal{M}} f(q) \tilde{\nu}(dq)$$

where $f = G^{-1}\nabla\xi \cdot \nabla V - \mathrm{div}(G^{-1}\nabla\xi)$ and

$$\tilde{\nu}(dq) = Z_{\tilde{\nu}}^{-1}\, \mathrm{e}^{-\tilde{V}(q)}\, \sigma_{\mathcal{M}}(dq)$$

where $\tilde{V}(q) = V(q) + \frac{\ln \det G(q)}{2}$.

$\longrightarrow$ Thermodynamic integration.

Plan

The constrained overdamped Langevin dynamics

The constrained Langevin dynamics

The reverse projection check

Beyond reverse projection check

Step 1: the overdamped Langevin dynamics (1/3)

The constrained overdamped Langevin dynamics (W_t is a d-dimensional Brownian motion):

$$\begin{cases} dq_t = -\nabla V(q_t)\,dt + \sqrt{2}dW_t + \nabla\xi(q_t)d\lambda_t \\ d\lambda_t \in \mathbb{R}^m \text{ such that } \xi(q_t) = 0 \end{cases}$$

is ergodic with respect to ν. It can indeed be rewritten as:

$$dq_t = \Pi(q_t) \circ \left(-\nabla V(q_t)\,dt + \sqrt{2}dW_t\right)$$

where $\circ$ denotes the Stratonovitch product and

$$\Pi(q) = \mathrm{Id} - \nabla\xi(q)G^{-1}(q)[\nabla\xi(q)]^T$$

is the orthogonal projector from $\mathbb{R}^d$ to $T_q\mathcal{M}$. One can then use the divergence theorem on manifolds to prove that its unique invariant measure is ν [Ciccotti, TL, Vanden-Einjden, 2008].

Step 1: the overdamped Langevin dynamics (2/3)

Discretization of the constrained overdamped Langevin dynamics:

$$\begin{cases} q^{n+1} = q^n - \nabla V(q^n)\Delta t + \sqrt{2\Delta t}\,G_n + \nabla\xi(q_n)\lambda^n \\ \lambda^n \in \mathbb{R}^m \text{ such that } \xi(q^{n+1}) = 0 \end{cases}$$

where $G_n \sim \mathcal{N}(0, \mathrm{Id})$.

Remark: By choosing $V = \tilde{V}$, an approximation of $\nabla A(0)$ is given by the average of the Lagrange multipliers:

$$\lim_{T\to\infty} \lim_{\Delta t\to 0} \frac{1}{T} \sum_{n=1}^{T/\Delta t} \lambda^n = \nabla A(0).$$

Step 1: the overdamped Langevin dynamics (3/3)

Time discretization implies a bias, which is of order Δt. Let $\nu_{\Delta t}$ be the invariant measure for $(q^n)_{n \geq 0}$, then [Faou, TL, 2010]: for all smooth function $\varphi : \mathcal{M} \to \mathbb{R}$, $\exists C$, for small Δt,

$$\left| \int_{\mathcal{M}} \varphi d\nu_{\Delta t} - \int_{\mathcal{M}} \varphi d\nu \right| \leq C \Delta t.$$

The proof is based on expansions à la Talay-Tubaro.

How to eliminate the bias?

Metropolis-Hastings is not easy to apply since the probability to go from q^n to q^{n+1} does not have a simple analytical expression.

Idea: lift the problem to phase space in order to use the symmetry up to momentum reversal of the constrained Hamiltonian dynamics.

Step 2: the Langevin dynamics (1/8)

Extended measure in phase space:

$$\mu(dq\,dp) = Z_\mu^{-1} \mathrm{e}^{-H(q,p)}\, \sigma_{T^*\mathcal{M}}(dq\,dp)$$

where $H(q,p) = V(q) + \frac{|p|^2}{2}$ and $\sigma_{T^*\mathcal{M}}(dq\,dp)$ is the phase space Liouville measure on

$$T^*\mathcal{M} = \left\{(q,p) \in \mathbb{R}^d \times \mathbb{R}^d,\, \xi(q) = 0 \text{ and } [\nabla\xi(q)]^T p = 0\right\}.$$

The marginal of μ in q is ν. Indeed, the measure μ rewrites:

$$\mu(dq\,dp) = \nu(dq)\,\kappa_q(dp)$$

where

$$\kappa_q(dp) = (2\pi)^{\frac{m-d}{2}} \mathrm{e}^{-\frac{|p|^2}{2}} \sigma_{T_q^*\mathcal{M}}(dp)$$

with $T_q^*\mathcal{M} = \left\{p \in \mathbb{R}^d,\, [\nabla\xi(q)]^T p = 0\right\} \subset \mathbb{R}^d$.

Remark: Here and in the following, we assume for simplicity that the mass tensor $M = \mathrm{Id}$. It is easy to generalize the algorithm and the analysis to the case $M \neq \mathrm{Id}$.

Step 2: the Langevin dynamics (2/8)

The constrained Langevin dynamics ($\gamma > 0$ is the friction parameter)

$$\begin{cases} dq_t = p_t\, dt \\ dp_t = -\nabla V(q_t)\, dt - \gamma p_t\, dt + \sqrt{2\gamma} dW_t + \nabla\xi(q_t)\, d\lambda_t \\ \xi(q_t) = 0 \end{cases}$$

is ergodic with respect to μ. Notice that $[\nabla\xi(q_t)]^T p_t = 0$.

It can be seen as the composition of two dynamics:

- the constrained Hamiltonian dynamics:

$$\begin{cases} dq_t = p_t\, dt \\ dp_t = -\nabla V(q_t)\, dt + \nabla\xi(q_t)\, d\lambda_t \\ \xi(q_t) = 0. \end{cases}$$

- the Ornstein-Uhlenbeck process on momenta:

$$\begin{cases} dq_t = 0 \\ dp_t = -\gamma p_t\, dt + \sqrt{2\gamma} dW_t + \nabla\xi(q_t)\, d\lambda_t \\ [\nabla\xi(q_t)]^T p_t = 0. \end{cases}$$

Step 2: the Langevin dynamics (3/8)

Discretization of the Ornstein-Uhlenbeck process on momenta: midpoint Euler leaves the measure κ_{q^n} and thus μ invariant:

$$\begin{cases} p^{n+1} = p^n - \frac{\Delta t}{2}\gamma\,(p^n + p^{n+1}) + \sqrt{2\gamma\Delta t}\,G^n + \nabla\xi(q^n)\,\lambda^n, \\ \nabla\xi(q^n)^T p^{n+1} = 0, \end{cases}$$

In the following, we denote one step of this dynamics by $\Psi_{\Delta t}^{OU} : T^*\mathcal{M} \to T^*\mathcal{M}$:

$$\Psi_{\Delta t}^{OU}(q^n, p^n) = (q^n, p^{n+1}).$$

Remark: The projection is always well defined, and easy to implement:

$$p^{n+1} = \Pi^*(q^n)\left(\frac{(1-\Delta t\gamma/2)p^n + \sqrt{2\gamma\Delta t}\,G^n}{1+\Delta t\gamma/2}\right)$$

where $\Pi^*(q) = \mathrm{Id} - \nabla\xi(q)G^{-1}(q)[\nabla\xi(q)]^T$ is the orthogonal projector from $\mathbb{R}^d$ to $T_q^*\mathcal{M}$.

Step 2: the Langevin dynamics (4/8)

Discretization of the constrained Hamiltonian dynamics (RATTLE):

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2}\nabla V(q^n) + \nabla\xi(q^n)\,\lambda^{n+1/2}, & \\ q^{n+1} = q^n + \Delta t\, p^{n+1/2}, & \\ \xi(q^{n+1}) = 0, & (C_q) \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}) + \nabla\xi(q^{n+1})\,\lambda^{n+1}, & \\ \left[\nabla\xi(q^{n+1})\right]^T p^{n+1} = 0, & (C_p) \end{cases}$$

where $\lambda^{n+1/2} \in \mathbb{R}^m$ are the Lagrange multipliers associated with the position constraints (C_q), and $\lambda^{n+1} \in \mathbb{R}^m$ are the Lagrange multipliers associated with the velocity constraints (C_p).

In the following, we denote one step of the RATTLE dynamics by $\Psi_{\Delta t}^{RATTLE} : T^*\mathcal{M} \to T^*\mathcal{M}$:

$$\Psi_{\Delta t}^{RATTLE}(q^n, p^n) = (q^{n+1}, p^{n+1}).$$

Step 2: the Langevin dynamics (5/8)

Discretization of the constrained Langevin dynamics (Strang splitting):

$$\begin{cases} (q^n, p^{n+1/4}) = \Psi^{OU}_{\Delta t/2}(q^n, p^n) \\ (q^{n+1}, p^{n+3/4}) = \Psi^{RATTLE}_{\Delta t}(q^n, p^{n+1/4}) \\ (q^{n+1}, p^{n+1}) = \Psi^{OU}_{\Delta t/2}(q^{n+1}, p^{n+3/4}) \end{cases}$$

But there is still a bias due to time discretization...

Step 2: the Langevin dynamics (6/8)

Let us denote by

$$S(q,p) = (q,-p)$$

the momentum reversal map and

$$\Psi_{\Delta t}(q,p) = S\left(\Psi_{\Delta t}^{RATTLE}(q,p)\right).$$

Fundamental properties of RATTLE: for Δt small enough,

- $\Psi_{\Delta t}\left(\Psi_{\Delta t}(q,p)\right) = (q,p)$
- $\Psi_{\Delta t}$ is a symplectic map, which thus preserves $\sigma_{T^*\mathcal{M}}$

[Hairer, Lubich, Wanner, 2006] [Leimkuhler, Reich, 2004] [Leimkuhler, Skeel, 1994].

One can thus add a Metropolis Hastings rejection step to get unbiased samples: if $(q',p') = \Psi_{\Delta t}(q,p)$, the MH ratio writes:

$$\frac{\delta_{\Psi_{\Delta t}(q',p')}(dq\,dp)\,\mathrm{e}^{-H(q',p')}\,\sigma_{T^*\mathcal{M}}(dq'\,dp')}{\delta_{\Psi_{\Delta t}(q,p)}(dq'\,dp')\,\mathrm{e}^{-H(q,p)}\,\sigma_{T^*\mathcal{M}}(dq\,dp)} = \mathrm{e}^{-H(q',p')+H(q,p)}.$$

Step 2: the Langevin dynamics (7/8)

Constrained Generalized Hybrid Monte Carlo algorithm ([TL, Rousset, Stoltz 2012], constrained version of [Horowitz 1991]):

$$\begin{cases} (q^n, p^{n+1/4}) = \Psi^{OU}_{\Delta t/2}(q^n, p^n) \\ (\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) = \Psi_{\Delta t}(q^n, p^{n+1/4}) \\ \text{If } U^n \leq \mathrm{e}^{-H(\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) + H(q^n, p^{n+1/4})} \\ \qquad \text{accept the proposal: } (q^{n+1}, p^{n+3/4}) = (\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) \\ \qquad \text{else reject the proposal: } (q^{n+1}, p^{n+3/4}) = (q^n, p^{n+1/4}) \\ \tilde{p}^{n+1} = -p^{n+3/4} \\ (q^{n+1}, p^{n+1}) = \Psi^{OU}_{\Delta t/2}(q^{n+1}, \tilde{p}^{n+1}) \end{cases}$$

where $U^n \sim \mathcal{U}(0,1)$.

Remark: If $\Delta t \gamma/4 = 1$, then $p^{n+1/4} = \Pi^*(q^n)(G^n) \sim \kappa_{q^n}$. One thus obtains a constrained HMC algorithm, consistent with the constrained overdamped Langevin (constrained MALA).

Step 2: the Langevin dynamics (8/8)

Problem: RATTLE is only well defined and reversible for locally small timesteps. Three possible difficulties:

- $\Psi_{\Delta t}(q,p)$ may not be defined;
- If $\Psi_{\Delta t}(q,p)$ is well defined, $\Psi_{\Delta t}\left(\Psi_{\Delta t}(q,p)\right)$ may not be defined;
- If $\Psi_{\Delta t}(q,p)$ and $\Psi_{\Delta t}\left(\Psi_{\Delta t}(q,p)\right)$ are well defined, one may have $\Psi_{\Delta t}\left(\Psi_{\Delta t}(q,p)\right) \neq (q,p)$.

Step 3: the reverse projection check (1/9)

In order to introduce the the set of positions and momenta from which RATTLE is well defined, let us rewrite the RATTLE dynamics as follows:

$$\begin{cases} q^{n+1} = q^n + \Delta t \left[p^n - \dfrac{\Delta t}{2} \nabla V(q^n) \right] + \Delta t \nabla \xi(q^n)\, \lambda^{n+1/2} \\ p^{n+1} = \Pi^*(q^n) \left(p^n - \dfrac{\Delta t}{2} \left(\nabla V(q^n) + \nabla V(q^{n+1}) \right) + \nabla \xi(q^n)\, \lambda^{n+1/2} \right) \end{cases}$$

where

$$\Delta t \lambda^{n+1/2} = \Lambda \left(q^n, q^n + \Delta t \left[p^n - \frac{\Delta t}{2} \nabla V(q^n) \right] \right).$$

The function $\Lambda : \mathcal{D} \to \mathbb{R}^m$, where $\mathcal{D}$ is an open set of $\mathcal{M} \times \mathbb{R}^d$ is the Lagrange multiplier function which satisfies:

$$\forall (q, \tilde{q}) \in \mathcal{D},\ \tilde{q} + \nabla \xi(q) \Lambda(q, \tilde{q}) \in \mathcal{M}.$$

We will discuss later how to rigorously build such a Lagrange multiplier function.

Step 3: the reverse projection check (2/9)

The function Λ is only defined on $\mathcal{D}$ and thus $\Psi_{\Delta t}^{RATTLE}$ is only defined on the open set:

$$A = \left\{ (q,p) \in T^*\mathcal{M}, \ \left(q, q + \Delta t\, M^{-1} \left[p - \frac{\Delta t}{2} \nabla V(q) \right] \right) \in \mathcal{D} \right\}$$

and likewise, $\Psi_{\Delta t} = S \circ \Psi_{\Delta t}^{RATTLE}$ is defined on A.

Proposition ([TL, Rousset, Stoltz 2018])

If Λ is C^1, then $\Psi_{\Delta t} : A \to T^*\mathcal{M}$ is a C^1 local diffeomorphism, locally preserving the phase-space measure $\sigma_{T^*\mathcal{M}}(dq\,dp)$.

Step 3: the reverse projection check (3/9)

Let us now introduce the RATTLE dynamics with momentum reversal and reverse projection check: for any $(q, p) \in T^*\mathcal{M}$,

$$\Psi^{\mathrm{rev}}_{\Delta t}(q, p) = \Psi_{\Delta t}(q, p)1_{\{(q,p)\in B\}} + (q, p)1_{\{(q,p)\notin B\}}$$

where the set $B \subset A \subset T^*\mathcal{M}$ is defined by

$$B = \Big\{(q, p) \in A,\ \Psi_{\Delta t}(q, p) \in A \text{ and } \Psi_{\Delta t} \circ \Psi_{\Delta t}(q, p) = (q, p)\Big\}.$$

Proposition ([TL, Rousset, Stoltz 2018])

Let us assume that Λ is C^1 and satisfies the non-tangential condition: $\forall (q, \tilde{q}) \in \mathcal{D}$,

$$\left[\nabla\xi\left(\tilde{q} + \nabla\xi(q)\Lambda(q, \tilde{q})\right)\right]^T \nabla\xi(q) \in \mathbb{R}^{m\times m} \text{ is invertible.}$$

Then, the set B is the union of path connected components of the open set $A \cap \Psi^{-1}_{\Delta t}(A)$. It is thus an open set of $T^*\mathcal{M}$. Moreover, $\Psi^{\mathrm{rev}}_{\Delta t} : T^*\mathcal{M} \to T^*\mathcal{M}$ is globally well defined, preserves globally the measure $\sigma_{T^*\mathcal{M}}(dq\, dp)$ and satisfies $\Psi^{\mathrm{rev}}_{\Delta t} \circ \Psi^{\mathrm{rev}}_{\Delta t} = \mathrm{Id}$.

Step 3: the reverse projection check (4/9)

Practically, $\Psi^{\mathrm{rev}}_{\Delta t}(q,p)$ is obtained from $(q,p) \in T^*\mathcal{M}$ by the following procedure:

(1) check if (q,p) is in A; if not return (q,p);

(2) when $(q,p) \in A$, compute the configuration (q^1,p^1) obtained by one step of the RATTLE scheme;

(3) check if $(q^1,-p^1)$ is in A; if not, return (q,p);

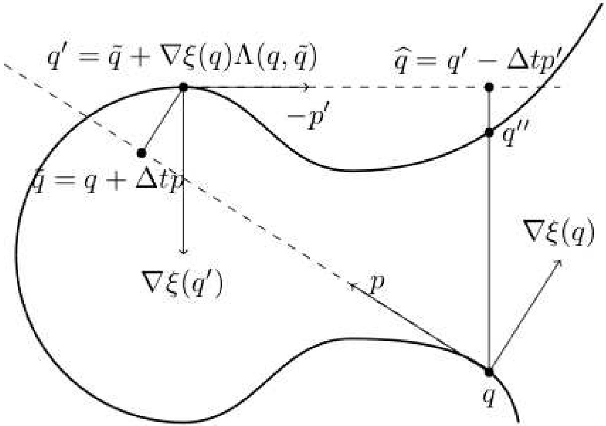
(4) compute the configuration $(q^2,-p^2)$ obtained by one step of the RATTLE scheme starting from $(q^1,-p^1)$;

(5) if $(q^2,p^2)=(q,p)$, return $(q^1,-p^1)$; otherwise return (q,p).

The steps (3)-(4)-(5) correspond to the *reverse projection check* [Goodman, Holmes-Cerfon, Zappa, 2017].

Step 3: the reverse projection check (5/9)

The reverse projection check is useful!



Here, $V = 0$ and the projection is defined as the closest point to $\mathcal{M}$. Notice that $q'' \neq q$!

Step 3: the reverse projection check (6/9)

Assume for simplicity that $\exists \alpha > 0$, $\{q \in \mathbb{R}^d, \|\xi(q)\| \leq \alpha\}$ is compact. How to build admissible Lagrange multiplier functions?

Theoretically, one can use the implicit function theorem to build a function $\Lambda : \mathcal{D}_{\mathrm{imp}} \to \mathbb{R}^m$ where $\mathcal{D}_{\mathrm{imp}}$ is an open subset of $\mathcal{M} \times \mathbb{R}^d$ such that

$$\{(q, \tilde{q}) \in \mathcal{M} \times \mathcal{M},\ [\nabla\xi(q)]^T \nabla\xi(\tilde{q}) \text{ is invertible}\} \subset \mathcal{D}_{\mathrm{imp}}$$

This defines $\Lambda(q, \tilde{q})$ for $\tilde{q}$ in a neighborhood of $\mathcal{M}$.

Remarks:

- There exists a non increasing function δ such that:

$$\|\xi(\tilde{q})\| < \delta\left(\|([\nabla\xi(q)]^T \nabla\xi(\tilde{q}))^{-1}\|\right) \Rightarrow (q, \tilde{q}) \in \mathcal{D}_{\mathrm{imp}}.$$

- For a fixed $(q, p) \in T^*\mathcal{M}$, for a sufficiently small timestep Δt, $\Psi_{\Delta t}(q, p) \in A_{\mathrm{imp}}$ and $\Psi_{\Delta t} \circ \Psi_{\Delta t}(q, p) = (q, p)$. Thus, B_{imp} is non empty!

Step 3: the reverse projection check (7/9)

Numerically, one can use the Newton algorithm to extend this local construction and compute the Lagrange multipliers for $\tilde{q}$ far from $\mathcal{M}$: perform a fixed number of iterations of the Newton algorithm to solve

$$\text{find } \lambda \in \mathbb{R}^m,\ \xi(\tilde{q} + \nabla\xi(q)\lambda) = 0$$

and check if one gets into $\mathcal{D}_{\mathrm{imp}}$.

In practice, the set $\mathcal{D}_{\mathrm{newt}}$ is defined as the configurations for which convergence to a point in $\mathcal{M}$ is observed (up to numerical error).

Step 3: the reverse projection check (8/9)

The constrained GHMC algorithm writes:

$$\begin{cases} (q^n, p^{n+1/4}) = \Psi^{OU}_{\Delta t/2}(q^n, p^n) \\ (\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) = \Psi^{rev}_{\Delta t}(q^n, p^{n+1/4}) \\ \text{If } U^n \leq \mathrm{e}^{-H(\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) + H(q^n, p^{n+1/4})} \\ \qquad \text{accept the proposal: } (q^{n+1}, p^{n+3/4}) = (\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) \\ \qquad \text{else reject the proposal: } (q^{n+1}, p^{n+3/4}) = (q^n, p^{n+1/4}) \\ \tilde{p}^{n+1} = -p^{n+3/4} \\ (q^{n+1}, p^{n+1}) = \Psi^{OU}_{\Delta t/2}(q^{n+1}, \tilde{p}^{n+1}) \end{cases}$$

where $U^n \sim \mathcal{U}(0,1)$.

Proposition ([TL, Rousset, Stoltz 2018])

The Markov chain $(q^n, p^n)_{n \geq 0}$ admits μ as an invariant measure.

To prove ergodicity, it remains to check irreducibility [Hartmann, 2008].

Step 3: the reverse projection check (9/9)

Remarks:

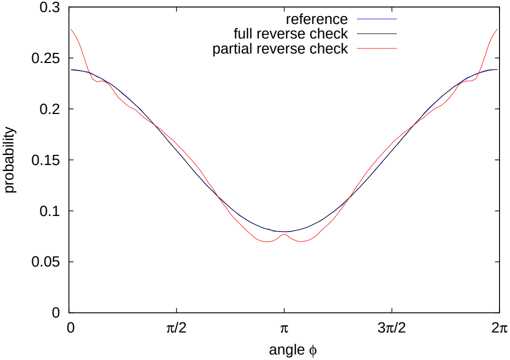
- In $\Psi^{rev}_{\Delta t}$, one can use any potential V! Choosing the potential V of the target measure $d\nu = Z_\nu^{-1} e^{-V} d\sigma_{\mathcal{M}}$ is good to increase the acceptance probability.
- If $\Delta t \gamma/4 = 1$, one obtains a HMC (or MALA) algorithm. If $\Delta t \gamma/4 = 1$ and $V = 0$ in $\Psi^{rev}_{\Delta t}$, this is a constrained random walk MH algorithm [Goodman, Holmes-Cerfon, Zappa, 2017].
- In pratice, one can use K steps of RATTLE within Ψ^{rev} to get less correlated samples. [Bou-Rabee, Sanz Serna]
- If $\Psi^{rev}_{\Delta t}(q^n, p^{n+1/4}) = \Psi_{\Delta t}(q^n, p^{n+1/4})$ (reverse projection check OK), one obtains a consistent discretization of the constrained Langevin dynamics.
- Similar ideas can be used to enforce inequality constraints.
- It may be interesting for numerical purposes to consider non identity mass matrices.

Numerical experiments (1/3)

Let $\mathcal{M}$ be the three-dimensional torus $\mathcal{M} = \{q \in \mathbb{R}^3, \, \xi(q) = 0\}$ where

$$\xi(q) = \left(R - \sqrt{x^2 + y^2}\right)^2 + z^2 - r^2,$$

with $0 < r < R$. Let us consider for $\nu = \sigma_{T^*\mathcal{M}}$ the uniform measure on $\mathcal{M}$.

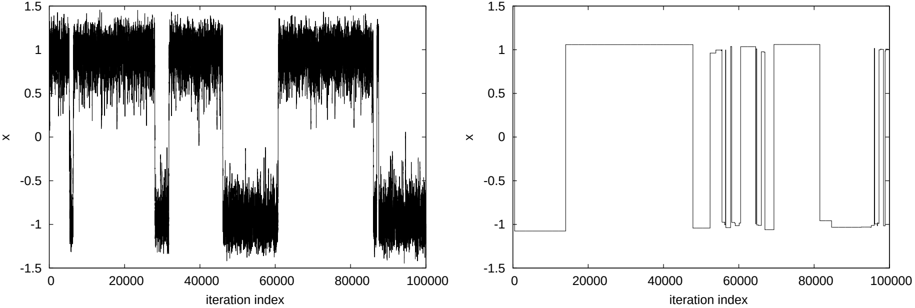


"partial reverse check" = do not check $(\Psi_{\Delta t} \circ \Psi_{\Delta t} = \mathrm{Id}) \Rightarrow$ BIAS!

Numerical experiments (2/3)

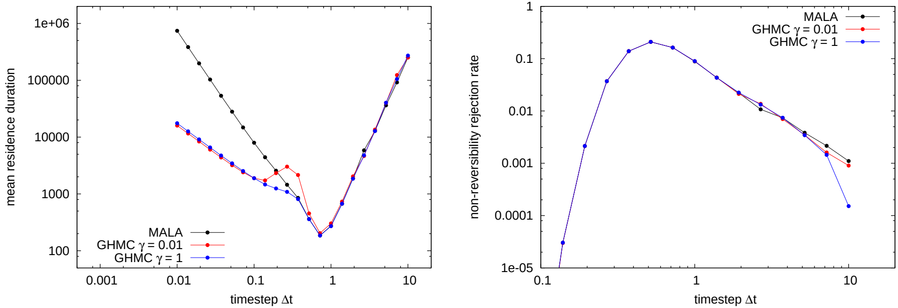
Let us now consider a double well case: $\nu = \mathrm{e}^{-V} \sigma_{T^*\mathcal{M}}$ where $V(x, y, z) = k(x^2 - R^2)^2$.

Typical trajectories for the GHMC dynamics (left $\Delta t = 0.05$, right $\Delta t = 3$):



Numerical experiments (3/3)

Analysis of the efficiency (left mean residence duration, right: non-reversibility rejection rate)



The optimal timestep is of the order of 0.7. For such timesteps, the rejections due to the non reversibility condition $(q^n, p^n) \neq \Psi_{\Delta t} \circ \Psi_{\Delta t}(q^n, p^n)$ are of the order of 15-20%, the total rejection rate being about 90%.

$\rightarrow$ reverse projection check is useful to get efficient algorithms.

Beyond reverse projection check

In collaboration with G. Stoltz and W. Zhang...
... following discussions with P. Breiding

In many situations, one is able to compute if not all, many solutions to the problem: for $(q, \tilde{q}) \in \mathcal{M} \times \mathbb{R}^d$,

$$\text{find } \Lambda(q, \tilde{q}) \in \mathbb{R}^m \text{ s.t. } \tilde{q} + \nabla\xi(q)\Lambda(q, \tilde{q}) \in \mathcal{M}.$$

Let us assume that there exists $\mathcal{D} \subset \mathcal{M} \times \mathbb{R}^d$ such that for all $(q, \tilde{q}) \in \mathcal{D}$, there exists $n(q, \tilde{q}) \in \mathbb{N}$ which is locally constant, and $n(q, \tilde{q})$ C^1 functions $(\Lambda^i(q, \tilde{q}) : \mathcal{D} \to \mathbb{R}^m)_{1 \le i \le n(q,\tilde{q})}$ such that

$$\forall (q, \tilde{q}) \in \mathcal{D},\ \tilde{q} + \nabla\xi(q)\Lambda^i(q, \tilde{q}) \in \mathcal{M}.$$

How to use this additional information ?

Examples

We have two situations in mind:

- All the solutions to $\xi(\tilde{q} + \nabla\xi(q)\lambda) = 0$ can be analytically computed
- Many solutions to $\xi(\tilde{q} + \nabla\xi(q)\lambda) = 0$ can be numerically computed

This is typically the case for algebraic submanifolds, i.e. when ξ is a polynomial function. See for example the Julia packages *PolynomialRoots* [J. Skowron, A. Gould] and *HomotopyContinuation* [P. Breiding, S. Timme].

Generalized algorithm (1/2)

Let us assign a probability $\pi^i(q, \tilde{q})$ to each of the solutions $(\Lambda^i(q, \tilde{q}))_{1 \le i \le n(q,\tilde{q})}$:

$$\pi^i(q, \tilde{q}) \ge 0 \text{ and } \sum_{i=1}^{n(q,\tilde{q})} \pi^i(q, \tilde{q}) = 1.$$

For example, $\pi^i(q, \tilde{q}) = \frac{1}{n(q,\tilde{q})}$.

Then choose one of the solution at random, and adapt the constrained GHMC algorithm.

Generalized algorithm (2/2)

1. Update momenta: $(q^n, p^{n+1/4}) = \Psi^{OU}_{\Delta t/2}(q^n, p^n)$
2. Compute the Lagrange multipliers $(\Lambda^i(q^n, \tilde{q}^n))_{1\leq i\leq n(q^n,\tilde{q}^n)}$, where $\tilde{q}^n = q^n + \Delta t\left[p^{n+1/4} - \frac{\Delta t}{2}\nabla V(q^n)\right]$. Choose index $i^n \in \{1, \ldots, n(q^n, \tilde{q}^n)\}$ with probability $\pi^{i^n}(q^n, \tilde{q}^n)$.
3. Compute the move $(\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) = \Psi_{\Delta t}(q^n, p^{n+1/4})$, where $\Psi_{\Delta t}$ uses in the RATTLE step the Lagrange multiplier Λ^{i^n}.
4. Check if one of the Lagrange multipliers (denoted by Λ^{j^n}) $(\Lambda^j(\tilde{q}^{n+1}, \bar{q}^{n+1}))_{1\leq j\leq n(q^{n+1},\bar{q}^{n+1})}$ brings back to $(q^n, -p^{n+1/4})$, where $\bar{q}^{n+1} = \tilde{q}^{n+1} + \Delta t\left[\tilde{p}^{n+3/4} - \frac{\Delta t}{2}\nabla V(\tilde{q}^{n+1})\right]$. If not, set $(q^{n+1}, p^{n+3/4}) = (q^n, p^{n+1/4})$, and go to Step 6.
5. Accept the move $(q^{n+1}, p^{n+3/4}) = (\tilde{q}^{n+1}, \tilde{p}^{n+3/4})$ with probability
$$1 \wedge \left(\frac{\pi^{j^n}(\tilde{q}^{n+1}, \bar{q}^{n+1})}{\pi^{i^n}(q^n, \tilde{q}^n)}\mathrm{e}^{-H(\tilde{q}^{n+1},\tilde{p}^{n+3/4})+H(q^n,p^{n+1/4})}\right)$$
else reject $(q^{n+1}, p^{n+3/4}) = (q^n, p^{n+1/4})$.
6. $\tilde{p}^{n+1} = -p^{n+3/4}$ and $(q^{n+1}, p^{n+1}) = \Psi^{OU}_{\Delta t/2}(q^{n+1}, \tilde{p}^{n+1})$

No need of a reverse projection check (1/2)

- If all the solutions to $\xi(\tilde{q} + \nabla\xi(q)\lambda) = 0$ can be computed, one knows in Step 4 that one of the Lagrange multipliers $(\Lambda^j(\tilde{q}^{n+1}, \bar{q}^{n+1}))_{1 \le j \le n(q^{n+1}, \bar{q}^{n+1})}$ brings back to $(q^n, -p^{n+1/4})$ $\longrightarrow$ no rejection in Step 4.
- If, in addition, one chooses $\pi^i(q, \tilde{q}) = \frac{1}{n(q,\tilde{q})}$, no need to identify the Lagrange multiplier Λ^{j^n} which brings back to $(q^n, -p^{n+1/4})$ in Setp 4.

Under these two assumptions (all Lagrange multipliers can be computed, and uniform probability on the Lagrange multipliers), the algorithm is the following:

No need of a reverse projection check (2/2)

1. Update momenta: $(q^n, p^{n+1/4}) = \Psi^{OU}_{\Delta t/2}(q^n, p^n)$
2. Compute the Lagrange multipliers $(\Lambda^i(q^n, \tilde{q}^n))_{1\leq i\leq n(q^n,\tilde{q}^n)}$, where $\tilde{q}^n = q^n + \Delta t \left[p^{n+1/4} - \frac{\Delta t}{2}\nabla V(q^n)\right]$. Choose index $i^n \in \{1, \dots, n(q^n, \tilde{q}^n)\}$ with probability $1/n(q^n, \tilde{q}^n)$.
3. Compute the move $(\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) = \Psi_{\Delta t}(q^n, p^{n+1/4})$, where $\Psi_{\Delta t}$ uses in the RATTLE step the Lagrange multiplier Λ^{i^n}.
4. Compute the number of Lagrange multipliers $n(q^{n+1}, \bar{q}^{n+1})$ for the backward move, where $\bar{q}^{n+1} = \tilde{q}^{n+1} + \Delta t \left[\tilde{p}^{n+3/4} - \frac{\Delta t}{2}\nabla V(\tilde{q}^{n+1})\right]$.
5. Accept the move $((q^{n+1}, p^{n+3/4}) = (\tilde{q}^{n+1}, \tilde{p}^{n+3/4}))$ with probability
$$1 \wedge \left(\frac{n(q^n, \tilde{q}^n)}{n(\tilde{q}^{n+1}, \bar{q}^{n+1})} \mathrm{e}^{-H(\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) + H(q^n, p^{n+1/4})}\right)$$
else reject $((q^{n+1}, p^{n+3/4}) = (q^n, p^{n+1/4}))$.
6. $\tilde{p}^{n+1} = -p^{n+3/4}$ and $(q^{n+1}, p^{n+1}) = \Psi^{OU}_{\Delta t/2}(q^{n+1}, \tilde{p}^{n+1})$

Theoretical results

One can adapt the arguments of the work [TL, Rousset, Stoltz 2018] to show that

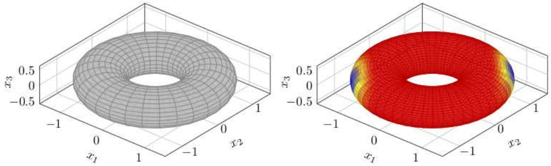
Proposition ([TL, Stoltz, Zhang 2019])

The Markov chain $(q^n, p^n)_{n\geq 0}$ admits μ as an invariant measure.

Same remarks as before apply: change V or the number of steps in $\Psi_{\Delta t}$, MALA version if $\Delta t\gamma/4 = 1$, inequality constraints, change the mass matrix, ...

Numerical experiments

Two measures on the torus.

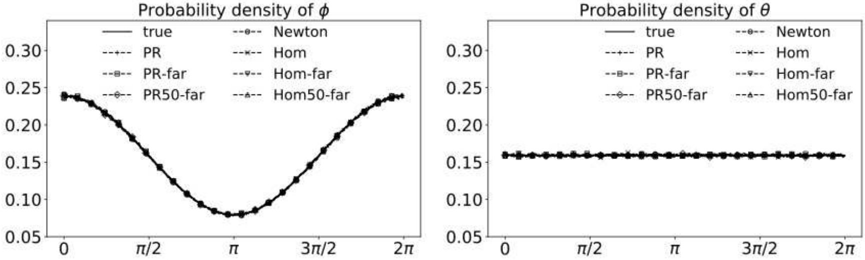


Compare, over 10^7 iterations:

- Newton
- PolynomialRoots (PR) or HomotopyContinuation (Hom) with uniform law on the Lagrange multipliers
- PolynomialRoots (PR-far) or HomotopyContinuation (Hom-far) with non-uniform law on the Lagrange multipliers (favor large jumps)
- Use PR-far and Hom-far every 50 Newton steps: PR50-far and Hom50-far

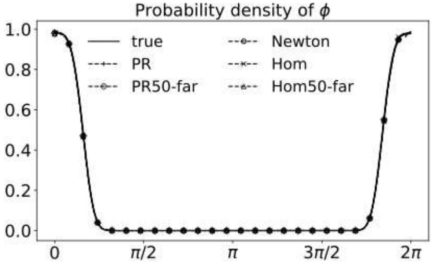
Numerical experiments

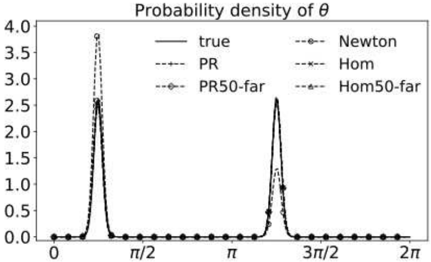
For the uniform law on the torus, results are indeed unbiased:



Numerical experiments

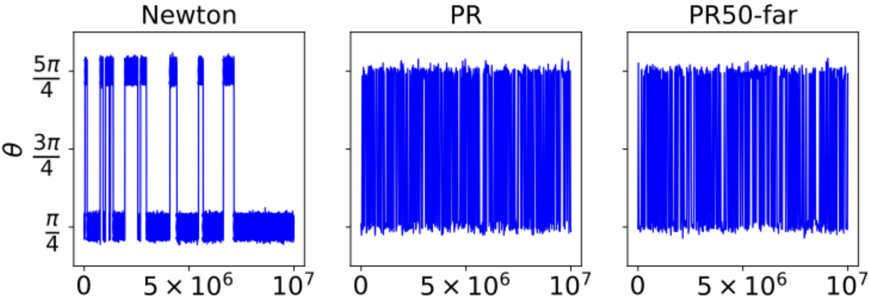
For the bimodal distribution on the torus, Newton exhibits metastability:





Numerical experiments

For the bimodal distribution on the torus, Newton exhibits metastability:



Numerical experiments

Quantitative comparisons for the bimodal distribution on the torus:

Scheme	No. of solutions in forward					No. of solutions in backward				
	0	1	2	3	4	0	1	2	3	4
Newton	7.9%	92.1%	0.0%	0.0%	0.0%	0.1%	99.9%	0.0%	0.0%	0.0%
PR	2.1%	0.0%	51.8%	0.0%	46.1%	0.0%	0.0%	75.2%	0.0%	24.8%
PR50-far	2.2%	95.8%	1.0%	0.0%	0.9%	0.0%	98.0%	1.5%	0.0%	0.5%
Hom	2.1%	0.0%	51.8%	0.0%	46.1%	0.0%	0.0%	75.2%	0.0%	24.8%
Hom50-far	2.2%	95.8%	1.0%	0.0%	0.9%	0.0%	98.0%	1.5%	0.0%	0.5%

Scheme	FSR	BSR	Jump rate	Large jump rate	Time (s)
Newton	0.92	1.00	0.38	1.8×10^{-6}	1.7×10^{3}
PR	0.98	1.00	0.22	4.1×10^{-3}	1.0×10^{4}
PR50-far	0.98	1.00	0.60 (0.18)	7.3×10^{-5}	2.0×10^{3}
Hom	0.98	1.00	0.22	4.1×10^{-3}	1.2×10^{4}
Hom50-far	0.98	1.00	0.60 (0.18)	6.4×10^{-5}	2.1×10^{3}

References

- G. Ciccotti, TL and E. Vanden-Eijnden, *Projection of diffusions on submanifolds: Application to mean force computation*, CPAM 61(3), 2008.
- A.M. Horowitz, *A generalized guided Monte-Carlo algorithm*. Phys. Lett. B 268, 1991.
- TL, M. Rousset and G. Stoltz, *Langevin dynamics with constraints and computation of free energy differences*, Math. Comp. 81, 2012.
- TL, M. Rousset and G. Stoltz, *Hybrid Monte Carlo methods for sampling probability measures on submanifolds*, Numerische Mathematik 143(2), 2019.
- TL, G. Stoltz and W. Zhang, *Multiple projection MCMC algorithms on submanifolds*, `https://arxiv.org/abs/2003.09402`
- E. Zappa, M. Holmes-Cerfon and J. Goodman, *Monte Carlo on manifolds: sampling densities and integrating functions*, CPAM 71(12), 2018.

References

Implementations:

- `https://github.com/zwpku/Constrained-HMC`
- `https://github.com/matt-graham/mici`