Ensuring unbiased sampling of HMC schemes for non separable Hamiltonian systems

Régis SANTET

(CERMICS, École des Ponts & MATHERIALS Team, Inria Paris)

1st year PhD student, Supervisors: G. Stoltz, T. Lelièvre

MCQMC 2022

Diffusion dependent Overdamped Langevin dynamics

• **Aim**: **Unbiased** estimation of $\mathbb{E}_\pi[f] = \int_{\mathcal{X}} f(q)\pi(q)\mathrm{d}q, \quad \pi \propto \mathrm{e}^{-\beta V}$ with the estimator

$$\hat{I}_N := \frac{1}{N}\sum_{i=1}^{N} f(q^i), \qquad q^i \sim \pi$$

Diffusion dependent Overdamped Langevin dynamics

- **Aim**: **Unbiased** estimation of $\mathbb{E}_\pi[f] = \int_{\mathcal{X}} f(q)\pi(q)\mathrm{d}q, \quad \pi \propto \mathrm{e}^{-\beta V}$

with the estimator

$$\hat{I}_N := \frac{1}{N}\sum_{i=1}^{N} f(q^i), \qquad q^i \sim \pi$$

- **Difficulty**: explore **anisotropic** potentials with **multiple minima**

Diffusion dependent Overdamped Langevin dynamics

- **Aim**: **Unbiased** estimation of $\mathbb{E}_\pi[f] = \int_{\mathcal{X}} f(q)\pi(q)\mathrm{d}q, \quad \pi \propto \mathrm{e}^{-\beta V}$

with the estimator

$$\hat{I}_N := \frac{1}{N}\sum_{i=1}^{N} f(q^i), \qquad q^i \sim \pi$$

- **Difficulty**: explore **anisotropic** potentials with **multiple minima**

- **Solution**: Position dependent positive definite symmetric matrix D[1]

$$\boxed{\mathrm{d}q_t = \left(-D(q_t)\nabla V(q_t) + \beta^{-1}\mathrm{div}\ D(q_t)\right)\ \mathrm{d}t + \sqrt{2\beta^{-1}D(q_t)}\,\mathrm{d}W_t}$$

[1] Bou-Rabee/Donev/Vanden-Eijnden (2014)

Diffusion dependent Overdamped Langevin dynamics

- **Aim**: **Unbiased** estimation of $\mathbb{E}_\pi[f] = \int_{\mathcal{X}} f(q)\pi(q)\mathrm{d}q, \quad \pi \propto \mathrm{e}^{-\beta V}$

with the estimator

$$\hat{I}_N := \frac{1}{N}\sum_{i=1}^{N} f(q^i), \qquad q^i \sim \pi$$

- **Difficulty**: explore **anisotropic** potentials with **multiple minima**

- **Solution**: Position dependent positive definite symmetric matrix D[1]

$$\boxed{\mathrm{d}q_t = \left(-D(q_t)\nabla V(q_t) + \beta^{-1}\mathrm{div}\ D(q_t)\right)\ \mathrm{d}t + \sqrt{2\beta^{-1}D(q_t)}\,\mathrm{d}W_t}$$

- **Challenge**: Efficient unbiased numerical integration

[1] Bou-Rabee/Donev/Vanden-Eijnden (2014)

Which diffusion coefficient? Metastable case

- Approach mainly used in Bayesian Inference[2]: $D \equiv (\nabla^2 V)^{-1}$
- Various works[3] suggest $D \propto \mathrm{e}^{\beta V} \mathrm{I}_d$

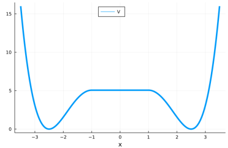
[2] Girolami/Calderhead (2011)

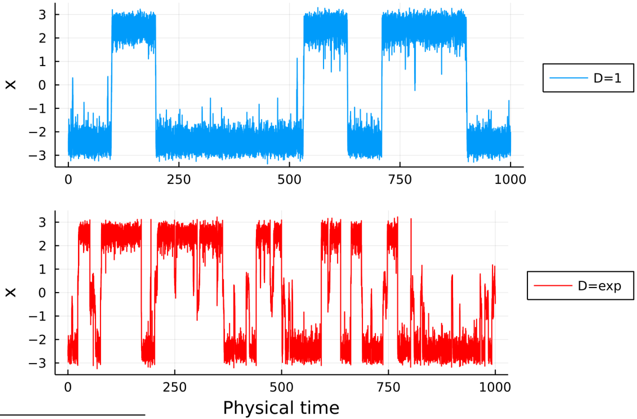
[3] Roberts/Stramer (2002), Lelièvre/Pavliotis/Robin/Stoltz (In prep.)

Which diffusion coefficient? Metastable case

- Approach mainly used in Bayesian Inference[2]: $D \equiv (\nabla^2 V)^{-1}$
- Various works[3] suggest $D \propto e^{\beta V} \mathrm{I}_d$

$\Rightarrow$ Helps to **cross energy barriers**: if $V \uparrow$, then $D \uparrow$





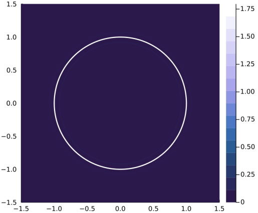
[2] Girolami/Calderhead (2011)

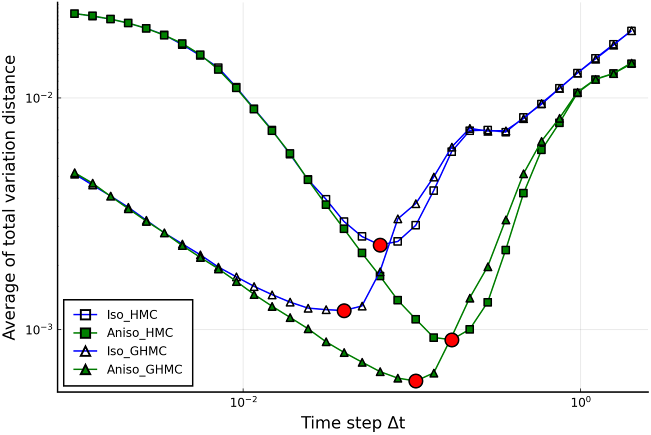
[3] Roberts/Stramer (2002), Lelièvre/Pavliotis/Robin/Stoltz (In prep.)

Which diffusion coefficient? Anisotropic case

- Anisotropic diffusion coefficient $D_{\mathsf{Tan}}(q) = \varepsilon \mathrm{I}_2 + \tilde{q}\tilde{q}^\mathsf{T}/\|q\|^2,\ \tilde{q} = (-y\ x)^\mathsf{T}$
- Isotropic diffusion coefficient $D_{\mathsf{One}} \equiv (1+\varepsilon)\mathrm{I}_2,\ \varepsilon = 0.1$

Computing: after fixed number of iterations, distance to the invariant measure of the angle distribution (uniform on $[0, 2\pi]$)

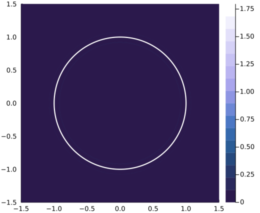


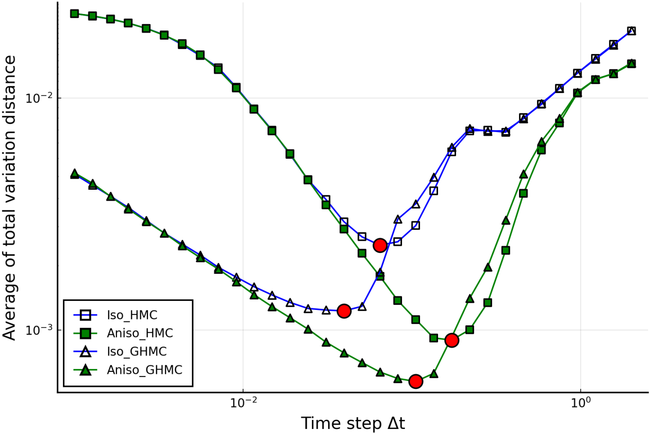


Which diffusion coefficient? Anisotropic case

- Anisotropic diffusion coefficient $D_{\mathsf{Tan}}(q) = \varepsilon \mathrm{I}_2 + \tilde{q}\tilde{q}^{\mathsf{T}}/\|q\|^2,\ \tilde{q} = (-y\ x)^{\mathsf{T}}$
- Isotropic diffusion coefficient $D_{\mathsf{One}} \equiv (1+\varepsilon)\mathrm{I}_2,\ \varepsilon = 0.1$

Computing: after fixed number of iterations, distance to the invariant measure of the angle distribution (uniform on $[0, 2\pi]$)





$\Rightarrow$ Compromise: small/large time steps (exploration vs rejection)

Unbiased sampling with Metropolis schemes

- Metropolis-Hastings: accept/reject with proba $\min\left(1, \frac{\pi(q')T(q', \mathrm{d}q)}{\pi(q)T(q, \mathrm{d}q')}\right)$

Unbiased sampling with Metropolis schemes

- Metropolis-Hastings: accept/reject with proba $\min\left(1, \frac{\pi(q')T(q', \mathrm{d}q)}{\pi(q)T(q, \mathrm{d}q')}\right)$
- Natural candidate: **Large** rejection rates[4] $\mathcal{O}\left(\Delta t^{1/2}\right)$

$$q' = q + \left(-D(q)\nabla V(q) + \beta^{-1}\mathrm{div}\ D(q)\right)\Delta t + \sqrt{2\Delta t \beta^{-1} D(q)} G$$

[4] Rossky/Doll/Friedman (1978), Fathi/Stoltz (2017)

Unbiased sampling with Metropolis schemes

- Metropolis-Hastings: accept/reject with proba $\min\left(1, \dfrac{\pi(q')T(q', \mathrm{d}q)}{\pi(q)T(q, \mathrm{d}q')}\right)$
- Natural candidate: **Large** rejection rates[4] $\mathcal{O}\left(\Delta t^{1/2}\right)$

$$q' = q + \left(-D(q)\nabla V(q) + \beta^{-1}\mathrm{div}\ D(q)\right)\Delta t + \sqrt{2\Delta t \beta^{-1} D(q)} G$$

- **Better choice**: (Generalized) Hamiltonian Monte Carlo[5] based on

$$\begin{cases} \mathrm{d}q_t = \nabla_p H(q_t, p_t)\ \mathrm{d}t \\ \mathrm{d}p_t = -\nabla_q H(q_t, p_t)\ \mathrm{d}t - \gamma \nabla_p H(q_t, p_t)\ \mathrm{d}t + \sqrt{2\gamma\beta^{-1}}\ \mathrm{d}W_t \end{cases}$$

with

$$H(q,p) = V(q) - \frac{1}{2}\ln\left(\det D(q)\right) + \frac{1}{2}p^{\mathsf{T}} D(q) p$$

- $p \sim \mathcal{N}(0, D(q)^{-1})$, marginal in position of $\mathrm{e}^{-\beta H}$ is π
- Consistent approximation of overdamped Langevin dynamics

[4] Rossky/Doll/Friedman (1978), Fathi/Stoltz (2017)

[5] Duane/Kennedy/Pendleton/Roweth (1987), Neal (1993)

Riemann Manifold HMC[7]

i) Sample momenta (Ornstein-Uhlenbeck or direct sampling)

ii) Integrate Hamiltonian dynamics
⇒ Generalized Störmer–Verlet[6] (time-reversible, symplectic but implicit)

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2}\nabla_q H(q^n, p^{n+1/2}) \\ q^{n+1} = q^n + \dfrac{\Delta t}{2}\left(\nabla_p H(q^n, p^{n+1/2}) + \nabla_p H(q^{n+1}, p^{n+1/2})\right) \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla_q H(q^{n+1}, p^{n+1/2}) \end{cases}$$

iii) Apply M–H procedure

[6] Hairer/Lubich/Wanner (2006)
[7] Girolami/Calderhead (2011)

Riemann Manifold HMC[7]

i) Sample momenta (Ornstein-Uhlenbeck or direct sampling)

ii) Integrate Hamiltonian dynamics
$\Rightarrow$ Generalized Störmer–Verlet[6] (time-reversible, symplectic but implicit)

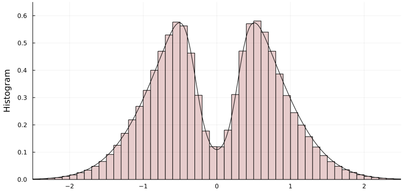
$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2} \nabla_q H(q^n, p^{n+1/2}) \\ q^{n+1} = q^n + \dfrac{\Delta t}{2} \left(\nabla_p H(q^n, p^{n+1/2}) + \nabla_p H(q^{n+1}, p^{n+1/2}) \right) \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2} \nabla_q H(q^{n+1}, p^{n+1/2}) \end{cases}$$

iii) Apply M–H procedure

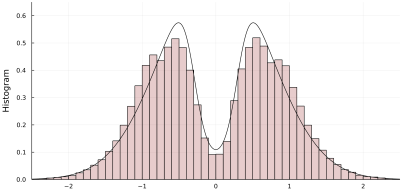
- (Effective) Rejection rates scale as $\mathcal{O}(\Delta t^{3/2})$

[6] Hairer/Lubich/Wanner (2006)
[7] Girolami/Calderhead (2011)

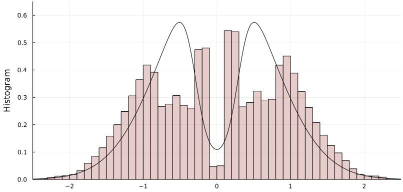
Bias arising with standard RMHMC implementation



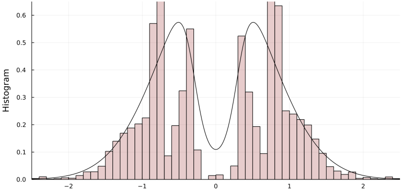
$\Delta t = 0.18$



$\Delta t = 0.69$

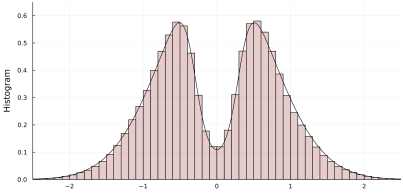


$\Delta t = 0.86$

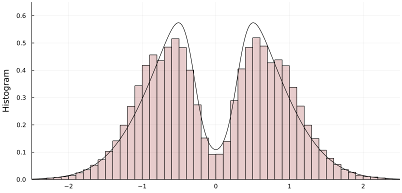


$\Delta t = 1.68$

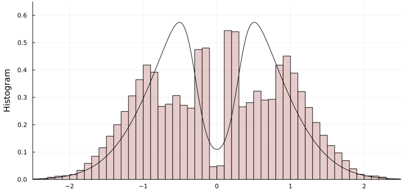
Bias arising with standard RMHMC implementation



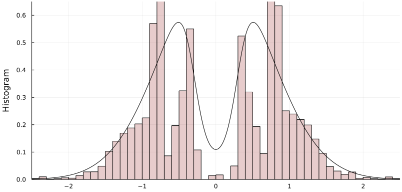
$\Delta t = 0.18$



$\Delta t = 0.69$



$\Delta t = 0.86$

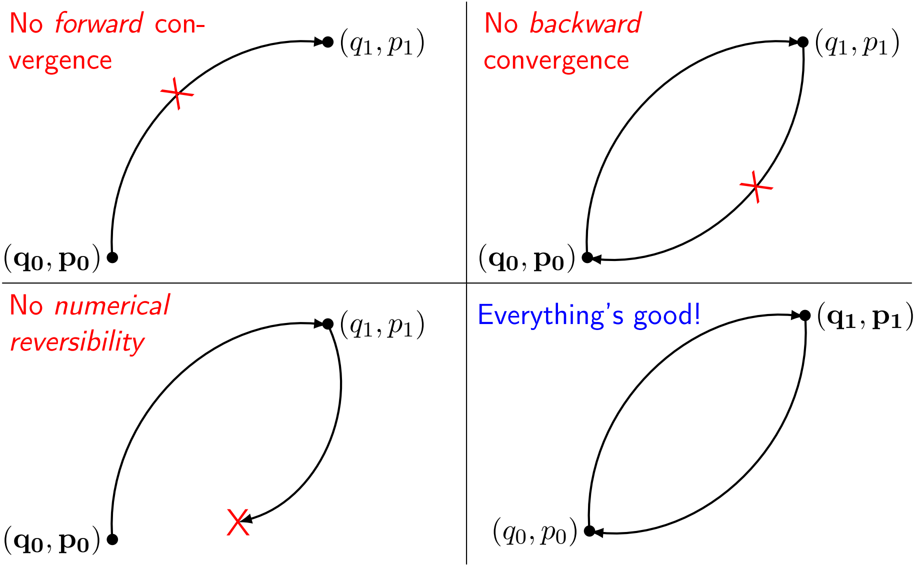


$\Delta t = 1.68$

Implicit methods $\Rightarrow$ **convergence** and **numerical reversibility** issues[8]

[8] Brofos/Lederman (2021)

RMHMC with enforced numerical reversibility[9]



[9] Zappa/Holmes-Cerfon/Goodman (2018)

HMC: canonical measure preservation

Theorem

HMC algorithm preserves the probability measure

$$\mu = \exp\left(-H(q,p)\right)/Z_\mu \,\mathrm{d}q\,\mathrm{d}p$$

Proof

$T_{\Delta t}((q,p),\mathrm{d}q'\,\mathrm{d}p') = r_{\Delta t}\delta_{\varphi_{\Delta t}(q,p)}(\mathrm{d}q'\,\mathrm{d}p') + (1 - r_{\Delta t}(q,p))\delta_{(q,p)}(\mathrm{d}q'\,\mathrm{d}p')$

If $f : \mathbb{R}^d \to \mathbb{R}^d \to \mathbb{R}$ measurable & bounded, [x=(q,p), S(q,p)=(q,-p)]

$$\begin{aligned}
\int r_{\Delta t}(x) f(\varphi_{\Delta t}(x))\mu(\mathrm{d}x) \quad &= \int r_{\Delta t}(\varphi_{\Delta t}^{-1}(y)) f(y) \frac{\mathrm{e}^{-\beta\left[H\circ\varphi_{\Delta t}^{-1}\right](y)}}{Z_\mu}\mathrm{d}y \\
[|\nabla\varphi_{\Delta t}| = 1] \quad &= \int r_{\Delta t}((S\circ\varphi_{\Delta t})(z)) f(z) \frac{\mathrm{e}^{-\beta[H\circ S\circ\varphi_{\Delta t}](z)}}{Z_\mu}\mathrm{d}z \\
[S\circ\varphi_{\Delta t}\circ S = \varphi_{\Delta t}^{-1}] \quad &= \int r_{\Delta t}(z) f(z)\mu(\mathrm{d}z)
\end{aligned}$$

Our contribution

Define $\psi_{\Delta t} = S \circ \varphi_{\Delta t}$ and

$$\psi_{\Delta t}^{\mathrm{REV}} = \psi_{\Delta t} \mathbf{1}_{\mathcal{B}} + \mathrm{id} \mathbf{1}_{\mathcal{B}^c}$$

where

$$\mathcal{B} = \left\{ (q,p) \,\middle|\, \psi_{\Delta t}^2 (q,p) = (q,p) \right\}$$

Proposition

$\psi_{\Delta t}^{\mathrm{REV}}$ is a globally defined measure preserving involution. RMHMC algorithm performed with $\psi_{\Delta t}^{\mathrm{REV}}$ yields an unbiased estimator.

Our contribution

Define $\psi_{\Delta t} = S \circ \varphi_{\Delta t}$ and

$$\psi_{\Delta t}^{\mathrm{REV}} = \psi_{\Delta t} \mathbf{1}_{\mathcal{B}} + \mathrm{id} \mathbf{1}_{\mathcal{B}^c}$$

where

$$\mathcal{B} = \left\{ (q,p) \,\middle|\, \psi_{\Delta t}^2(q,p) = (q,p) \right\}$$

Proposition

$\psi_{\Delta t}^{\mathrm{REV}}$ is a globally defined measure preserving involution. RMHMC algorithm performed with $\psi_{\Delta t}^{\mathrm{REV}}$ yields an unbiased estimator.

Show that $\mathcal{B}$ is measurable (even open)

Our contribution

Define $\psi_{\Delta t} = S \circ \varphi_{\Delta t}$ and

$$\psi_{\Delta t}^{\mathrm{REV}} = \psi_{\Delta t} \mathbf{1}_{\mathcal{B}} + \mathrm{id} \mathbf{1}_{\mathcal{B}^c}$$

where

$$\mathcal{B} = \left\{ (q,p) \,\middle|\, \psi_{\Delta t}^2(q,p) = (q,p) \right\}$$

Proposition

$\psi_{\Delta t}^{\mathrm{REV}}$ is a globally defined measure preserving involution. RMHMC algorithm performed with $\psi_{\Delta t}^{\mathrm{REV}}$ yields an unbiased estimator.

Show that $\mathcal{B}$ is measurable (even open)

- Whole space is a topological manifold, hence locally path connected

Our contribution

Define $\psi_{\Delta t} = S \circ \varphi_{\Delta t}$ and

$$\psi_{\Delta t}^{\mathrm{REV}} = \psi_{\Delta t} \mathbf{1}_{\mathcal{B}} + \mathrm{id} \mathbf{1}_{\mathcal{B}^c}$$

where

$$\mathcal{B} = \left\{ (q,p) \,\middle|\, \psi_{\Delta t}^2(q,p) = (q,p) \right\}$$

Proposition

$\psi_{\Delta t}^{\mathrm{REV}}$ is a globally defined measure preserving involution. RMHMC algorithm performed with $\psi_{\Delta t}^{\mathrm{REV}}$ yields an unbiased estimator.

Show that $\mathcal{B}$ is measurable (even open)

- Whole space is a topological manifold, hence locally path connected
- Show that each path connected component of $\mathcal{B}$ is open

Our contribution

Define $\psi_{\Delta t} = S \circ \varphi_{\Delta t}$ and

$$\psi_{\Delta t}^{\mathrm{REV}} = \psi_{\Delta t} \mathbf{1}_{\mathcal{B}} + \mathrm{id} \mathbf{1}_{\mathcal{B}^c}$$

where

$$\mathcal{B} = \left\{ (q,p) \,\middle|\, \psi_{\Delta t}^2(q,p) = (q,p) \right\}$$

Proposition

$\psi_{\Delta t}^{\mathrm{REV}}$ is a globally defined measure preserving involution. RMHMC algorithm performed with $\psi_{\Delta t}^{\mathrm{REV}}$ yields an unbiased estimator.

Show that $\mathcal{B}$ is measurable (even open)

- Whole space is a topological manifold, hence locally path connected
- Show that each path connected component of $\mathcal{B}$ is open
- Then $\varphi_{\Delta t}^{\mathrm{REV}}$ is a $\mathcal{C}^1$-measure preserving local diffeomorphism on $\mathcal{B}$

Our contribution

Define $\psi_{\Delta t} = S \circ \varphi_{\Delta t}$ and

$$\psi_{\Delta t}^{\mathrm{REV}} = \psi_{\Delta t}\mathbf{1}_{\mathcal{B}} + \mathrm{id}\mathbf{1}_{\mathcal{B}^c}$$

where

$$\mathcal{B} = \left\{(q,p) \,\middle|\, \psi_{\Delta t}^2(q,p) = (q,p)\right\}$$

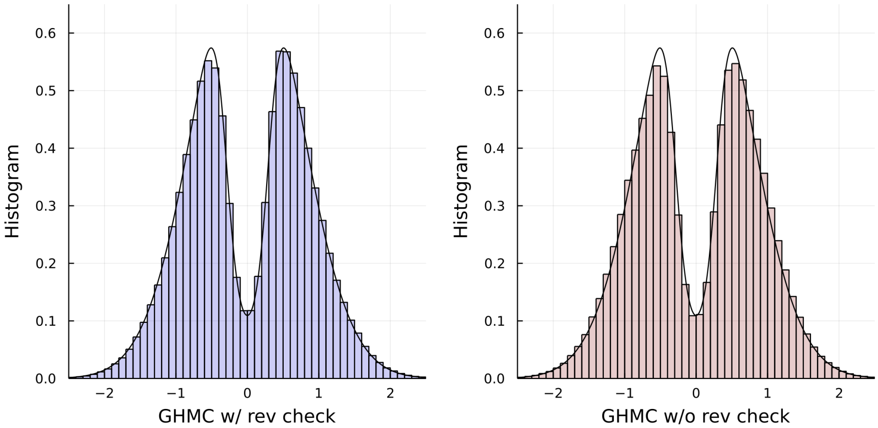
Proposition

$\psi_{\Delta t}^{\mathrm{REV}}$ is a globally defined measure preserving involution. RMHMC algorithm performed with $\psi_{\Delta t}^{\mathrm{REV}}$ yields an unbiased estimator.

Show that $\mathcal{B}$ is measurable (even open)

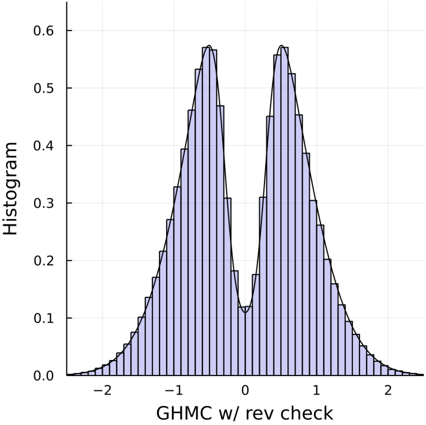
- Whole space is a topological manifold, hence locally path connected
- Show that each path connected component of $\mathcal{B}$ is open
- Then $\varphi_{\Delta t}^{\mathrm{REV}}$ is a $\mathcal{C}^1$-measure preserving local diffeomorphism on $\mathcal{B}$
- Coming from its definition on $\mathcal{B}^c$, it is a involution

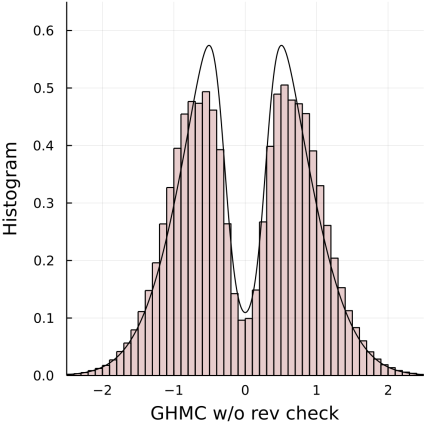
Unbiased sampling with corrected RMHMC



Sampling results with $\Delta t = 0.28$. Left histogram: reversibility checks.

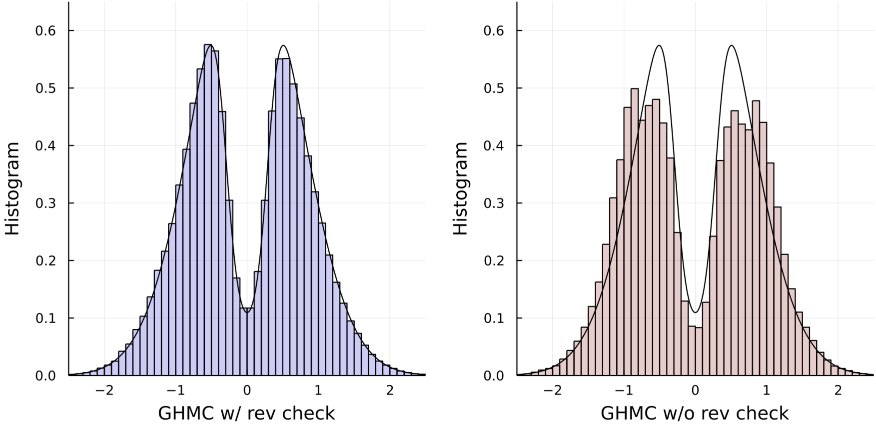
Unbiased sampling with corrected RMHMC





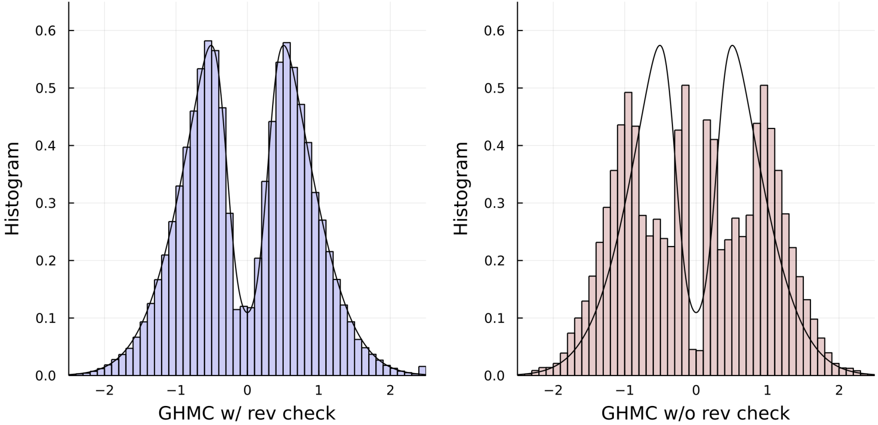
Sampling results with $\Delta t = 0.44$.

Unbiased sampling with corrected RMHMC



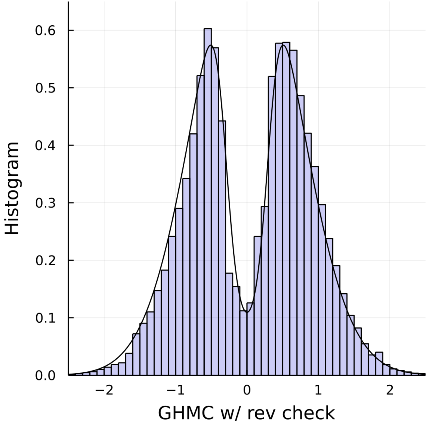
Sampling results with $\Delta t = 0.69$.

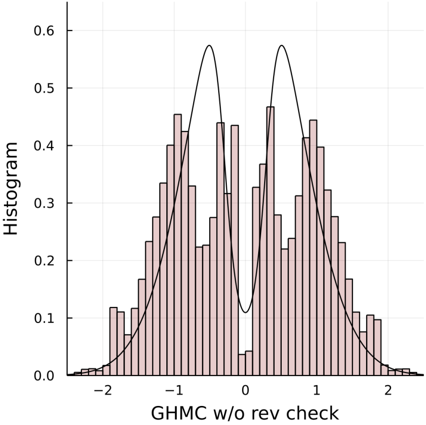
Unbiased sampling with corrected RMHMC



Sampling results with $\Delta t = 0.86$.

Unbiased sampling with corrected RMHMC





Sampling results with $\Delta t = 1.08$.

Conclusion and perspectives

Conclusions

- Overdamped Langevin with position dependent diffusion can dramatically **accelerate convergence**
- Care is required in numerical integration

Conclusions

- Overdamped Langevin with position dependent diffusion can dramatically **accelerate convergence**
- Care is required in numerical integration

Perspectives

- Higher dimension case: **use free energy F**, reaction coordinate ξ

$$D(q) \propto \mathrm{e}^{\beta \mathrm{F}(\xi(q))}$$

- Extension to **non-equilibrium** systems, F non-gradient force

$$\mathrm{d}q_t^\eta = \big(D(q_t)\left[-\nabla V(q_t^\eta) + \eta F(q_t^\eta)\right] + \beta^{-1}\mathrm{div} D(q_t^\eta)\big)\,\mathrm{d}t + \sqrt{2\beta^{-1}D(q_t^\eta)}\,\mathrm{d}W_t$$

Thank you !