Unbiased constrained sampling with Self-Concordant Barrier Hamiltonian Monte Carlo

Maxence Noble[1] Valentin de Bortoli[2] Alain Durmus[1]

[1]Centre de Mathématiques Appliquées, Ecole Polytechnique
Institut Polytechnique de Paris, France

[2]Département d'Informatique, École Normale Supérieure
CNRS, Université PSL, Paris, France

Outline

Outline

1. Motivations and background
   - General setting
   - RMHMC: basics and challenges
   - Summary of the motivations and assumptions
2. Description of BHMC
3. Results
4. Conclusion

Constrained sampling

- Consider a **subset** $\mathsf{M} \subset \mathbb{R}^d$.
- Our goal: sample from a **target distribution** π supported on M and known up to a normalising constant Z

$$\mathrm{d}\pi(x)/\mathrm{d}x = \exp[-V(x)]/Z \quad , \quad V \in \mathrm{C}^2(\mathsf{M}, \mathbb{R}) \; .$$

Constrained sampling

- Consider a **subset** $\mathsf{M} \subset \mathbb{R}^d$.
- Our goal: sample from a **target distribution** π supported on M and known up to a normalising constant Z

$$\mathrm{d}\pi(x)/\mathrm{d}x = \exp[-V(x)]/Z \quad , \quad V \in \mathrm{C}^2(\mathsf{M}, \mathbb{R}) \ .$$

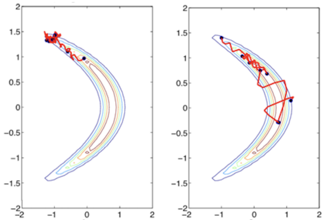
➔ When $\mathsf{M} = \mathbb{R}^d$, **gradient-based Markov Chain Monte Carlo** (MCMC) methods are very popular and come with some theoretical guarantees under assumptions on V (Duane et al., 1987; Roberts and Tweedie, 1996).

➔ However, their extension to ***constrained*** **sampling** still faces challenges (Gelfand et al., 1992; Pakman and Paninski, 2014; Lan and Shahbaba, 2015).

About MCMC methods

Traditional MCMC approaches for constrained sampling suffer from **poor mixing times** if M or V have a *sharp* geometry, including

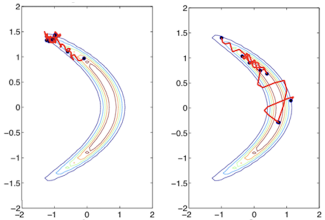
- Hit-and-Run (Lovász and Vempala, 2004),
- Ball Walk (Lee and Vempala, 2017b) (*left*),
- Hamiltonian Monte Carlo (HMC) (Duane et al., 1987) (*right*).



About MCMC methods

Traditional MCMC approaches for constrained sampling suffer from **poor mixing times** if M or V have a *sharp* geometry, including

- Hit-and-Run (Lovász and Vempala, 2004),
- Ball Walk (Lee and Vempala, 2017b) (*left*),
- Hamiltonian Monte Carlo (HMC) (Duane et al., 1987) (*right*).



This fact motivates to directly **incorporate the geometric constraints** into the sampling algorithms.

- If $\mathrm{M} = \{x \in \mathbb{R}^d : c(x) = 0\}$: HMC + RATTLE integrator (Leimkuhler and Skeel, 1994; Brubaker et al., 2012).
- If $(\mathrm{M}, \mathfrak{g})$ is Riemannian submanifold: **Riemannian Manifold HMC** (RMHMC) (Girolami and Calderhead, 2011).

Basics of RMHMC

Assume M (e.g., open) is a d-dimensional submanifold of $\mathbb{R}^d$, endowed with some **Riemannian metric** $\mathfrak{g}$.

Results on $(\mathrm{M}, \mathfrak{g})$ (see Lee (2006) and Mok (1977) for details):

Basics of RMHMC

Assume M (e.g., open) is a d-dimensional submanifold of $\mathbb{R}^d$, endowed with some **Riemannian metric** $\mathfrak{g}$.

Results on $(\mathsf{M}, \mathfrak{g})$ (see Lee (2006) and Mok (1977) for details):

- **Volume element** on M: $\mathrm{dvol}_{\mathsf{M}}(x) = \sqrt{\det \mathfrak{g}(x)}\mathrm{d}x$.

Basics of RMHMC

Assume M (e.g., open) is a d-dimensional submanifold of $\mathbb{R}^d$, endowed with some **Riemannian metric** $\mathfrak{g}$.

Results on $(\mathsf{M}, \mathfrak{g})$ (see Lee (2006) and Mok (1977) for details):

- **Volume element** on M: $\mathrm{dvol}_{\mathsf{M}}(x) = \sqrt{\det \mathfrak{g}(x)}\mathrm{d}x$.
- $\mathrm{T}^\star_x\mathsf{M}$: **cotangent space** at $x \in \mathsf{M}$
  $\rightarrow \mathrm{T}^\star_x\mathsf{M} \equiv \mathbb{R}^d$, endowed with the scalar product $\langle \cdot, \cdot \rangle_{\mathfrak{g}(x)^{-1}}$
  $\rightarrow$ Standard Gaussian distr. w.r.t. $\| \cdot \|_{\mathfrak{g}(x)^{-1}}$: $\mathrm{N}_x(0, \mathrm{I}_d) \equiv \mathrm{N}(0, \mathfrak{g}(x))$

Basics of RMHMC

Assume M (e.g., open) is a d-dimensional submanifold of $\mathbb{R}^d$, endowed with some **Riemannian metric** $\mathfrak{g}$.

Results on $(\mathsf{M}, \mathfrak{g})$ (see Lee (2006) and Mok (1977) for details):

- **Volume element** on M: $\mathrm{dvol}_{\mathsf{M}}(x) = \sqrt{\det \mathfrak{g}(x)}\mathrm{d}x$.
- $\mathrm{T}_x^\star \mathsf{M}$: **cotangent space** at $x \in \mathsf{M}$
  $\to \mathrm{T}_x^\star \mathsf{M} \equiv \mathbb{R}^d$, endowed with the scalar product $\langle \cdot, \cdot \rangle_{\mathfrak{g}(x)^{-1}}$
  $\to$ Standard Gaussian distr. w.r.t. $\| \cdot \|_{\mathfrak{g}(x)^{-1}}$: $\mathrm{N}_x(0, \mathrm{I}_d) \equiv \mathrm{N}(0, \mathfrak{g}(x))$
- $\mathrm{T}^\star \mathsf{M}$: **cotangent bundle** of M, defined by $\mathrm{T}^\star \mathsf{M} = \sqcup_{x \in \mathsf{M}} \{x\} \cup \mathrm{T}_x^\star \mathsf{M}$
  $\to$ $2d$-dim. submanifold which may be endowed with a specific metric $\mathfrak{g}^\star$ inherited from $\mathfrak{g}$ which verifies

$$\mathrm{dvol}_{\mathrm{T}^\star \mathsf{M}}(x, p) = \sqrt{\det \mathfrak{g}^\star(x, p)}\mathrm{d}x\mathrm{d}p = \mathbf{d}\boldsymbol{x}\mathbf{d}\boldsymbol{p}$$

Basics of RMHMC

Assume M (e.g., open) is a d-dimensional submanifold of $\mathbb{R}^d$, endowed with some **Riemannian metric** $\mathfrak{g}$.

➔ Results on $(\mathsf{M}, \mathfrak{g})$ (see Lee (2006) and Mok (1977) for details):

- **Volume element** on M: $\mathrm{dvol}_{\mathsf{M}}(x) = \sqrt{\det \mathfrak{g}(x)}\mathrm{d}x$.
- **Volume element** on $\mathrm{T}^\star\mathsf{M}$: $\mathrm{dvol}_{\mathrm{T}^\star\mathsf{M}}(x,p) = \sqrt{\det \mathfrak{g}^\star(x,p)}\mathrm{d}x\mathrm{d}p = \boldsymbol{\mathrm{d}x\mathrm{d}p}$.

Basics of RMHMC

Assume M (e.g., open) is a d-dimensional submanifold of $\mathbb{R}^d$, endowed with some **Riemannian metric** $\mathfrak{g}$.

➔ Results on $(\mathsf{M}, \mathfrak{g})$ (see Lee (2006) and Mok (1977) for details):

- **Volume element** on M: $\mathrm{dvol}_{\mathsf{M}}(x) = \sqrt{\det \mathfrak{g}(x)}\mathrm{d}x$.
- **Volume element** on $\mathrm{T}^\star\mathsf{M}$: $\mathrm{dvol}_{\mathrm{T}^\star\mathsf{M}}(x, p) = \sqrt{\det \mathfrak{g}^\star(x, p)}\mathrm{d}x\mathrm{d}p = \mathbf{d}\boldsymbol{x}\mathbf{d}\boldsymbol{p}$.

➔ In terms of probability distributions:

- **Target measure**: $\mathrm{d}\pi(x)/\mathrm{dvol}_{\mathsf{M}}(x) = \exp[-V(x) - \frac{1}{2}\log(\det \mathfrak{g}(x))]/Z$.
- **Hamiltonian** on $\mathrm{T}^\star\mathsf{M}$: $H(x, p) \stackrel{\text{def}}{=} V(x) + \frac{1}{2}\log\left(\det \mathfrak{g}(x)\right) + \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}}$.

Basics of RMHMC

Assume M (e.g., open) is a d-dimensional submanifold of $\mathbb{R}^d$, endowed with some **Riemannian metric** $\mathfrak{g}$.

➔ Results on $(\mathsf{M}, \mathfrak{g})$ (see Lee (2006) and Mok (1977) for details):

- **Volume element** on M: $\mathrm{dvol}_{\mathsf{M}}(x) = \sqrt{\det \mathfrak{g}(x)}\mathrm{d}x$.
- **Volume element** on $\mathrm{T}^\star\mathsf{M}$: $\mathrm{dvol}_{\mathrm{T}^\star\mathsf{M}}(x,p) = \sqrt{\det \mathfrak{g}^\star(x,p)}\mathrm{d}x\mathrm{d}p = \mathbf{d}\boldsymbol{x}\mathbf{d}\boldsymbol{p}$.

➔ In terms of probability distributions:

- **Target measure**: $\mathrm{d}\pi(x)/\mathrm{dvol}_{\mathsf{M}}(x) = \exp[-V(x) - \frac{1}{2}\log(\det \mathfrak{g}(x))]/Z$.
- **Hamiltonian** on $\mathrm{T}^\star\mathsf{M}$: $H(x,p) \overset{\text{def}}{=} V(x) + \frac{1}{2}\log\left(\det \mathfrak{g}(x)\right) + \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}}$.

RMHMC aims at sampling from the **augmented** target distribution $\bar{\pi}$ on $\mathrm{T}^\star\mathsf{M}$

$$\mathrm{d}\bar{\pi}(x,p) \overset{\text{def}}{=} \mathrm{d}\pi(x)\mathrm{N}_x(p;0,\mathrm{I}_d)\mathrm{d}p \propto \exp[-H(x,p)]\mathrm{dvol}_{\mathrm{T}^\star\mathsf{M}}(x,p) \; .$$

If $\mathfrak{g}(x) = \mathrm{I}_d$, we recover the setting of "Euclidean" HMC !

Definition of the Hamiltonian dynamics

HMC Ham. on $\mathbb{R}^{2d}$: $H(x,p) = V(x) + \frac{1}{2}\|p\|_2^2$.

RMHMC Ham. on $\mathrm{T}^\star \mathrm{M}$: $H(x,p) = V(x) + \frac{1}{2}\log\left(\det \mathfrak{g}(x)\right) + \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}}$.

Definition of the Hamiltonian dynamics

HMC Ham. on $\mathbb{R}^{2d}$: $H(x,p) = V(x) + \frac{1}{2}\|p\|_2^2$.

RMHMC Ham. on $\mathrm{T}^\star\mathrm{M}$: $H(x,p) = V(x) + \frac{1}{2}\log(\det \mathfrak{g}(x)) + \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}}$.

The **Hamiltonian dynamics** associated with H is given by the following ODEs

$$\dot{x}_t = \partial_p H(x_t, p_t) \;, \qquad \dot{p}_t = -\partial_x H(x_t, p_t) \;. \tag{1}$$

The corresponding **flow** is denoted by $\Psi : t, (x_0, p_0) \mapsto (x_t, p_t)$.

Definition of the Hamiltonian dynamics

HMC Ham. on $\mathbb{R}^{2d}$: $H(x,p) = V(x) + \frac{1}{2}\|p\|_2^2$.
RMHMC Ham. on $\mathrm{T}^\star\mathrm{M}$: $H(x,p) = V(x) + \frac{1}{2}\log\left(\det \mathfrak{g}(x)\right) + \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}}$.

The **Hamiltonian dynamics** associated with H is given by the following ODEs

$$\dot{x}_t = \partial_p H(x_t, p_t)\ , \qquad \dot{p}_t = -\partial_x H(x_t, p_t)\ . \tag{1}$$

The corresponding **flow** is denoted by $\Psi : t, (x_0, p_0) \mapsto (x_t, p_t)$.

- ➔ **Conservation** of the Hamiltonian through time.
- ➔ **Volume preservation** through time: the flow Ψ is *symplectic*.
- ➔ **Time-reversibility**: $\Psi_t^{-1} = s \circ \Psi_t \circ s$ where $s(x,p) = (x, -p)$.

In particular, $s \circ \Psi_t$ is an involution.

Definition of the Hamiltonian dynamics

HMC Ham. on $\mathbb{R}^{2d}$: $H(x,p) = V(x) + \frac{1}{2}\|p\|_2^2$.

RMHMC Ham. on $\mathrm{T}^\star\mathrm{M}$: $H(x,p) = V(x) + \frac{1}{2}\log\left(\det \mathfrak{g}(x)\right) + \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}}$.

The **Hamiltonian dynamics** associated with H is given by the following ODEs

$$\dot{x}_t = \partial_p H(x_t, p_t) \ , \qquad \dot{p}_t = -\partial_x H(x_t, p_t) \ . \tag{1}$$

The corresponding **flow** is denoted by $\Psi : t, (x_0, p_0) \mapsto (x_t, p_t)$.

- ➔ **Conservation** of the Hamiltonian through time.
- ➔ **Volume preservation** through time: the flow Ψ is *symplectic*.
- ➔ **Time-reversibility**: $\Psi_t^{-1} = s \circ \Psi_t \circ s$ where $s(x,p) = (x, -p)$.

In particular, $s \circ \Psi_t$ is an involution.

$$\begin{aligned} &\textbf{HMC}: && \dot{x} = p_t \ , && \dot{p}_t = -\nabla V(x_t). \\ &\textbf{RMHMC}: && \dot{x}_t = \mathfrak{g}(x_t)^{-1} p_t \ , && \dot{p}_t = -\nabla V(x_t) + L(x_t) \ . \end{aligned}$$

where $L(x) = -\frac{1}{2}\mathfrak{g}(x)^{-1} : D\mathfrak{g}(x) + \frac{1}{2}D\mathfrak{g}(x)[\dot{x}, \dot{x}]$. The Riemannian metric $\mathfrak{g}$ is incorporated into the dynamics, but this dynamics is more **complex to solve**...

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on $\mathrm{T}^\star\mathrm{M}$ is given by

$$\mathrm{d}\bar{\pi}(x,p) \stackrel{\text{def}}{=} \mathrm{d}\pi(x)\mathrm{N}_x(p;0,\mathrm{I}_d)\mathrm{d}p \propto \exp[-H(x,p)]\mathrm{dvol}_{\mathrm{T}^\star\mathrm{M}}(x,p) \ .$$

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on $\mathrm{T}^\star\mathsf{M}$ is given by

$$\mathrm{d}\bar{\pi}(x,p) \stackrel{\text{def}}{=} \mathrm{d}\pi(x)\mathrm{N}_x(p;0,\mathrm{I}_d)\mathrm{d}p \propto \exp[-H(x,p)]\mathrm{d}\mathrm{vol}_{\mathrm{T}^\star\mathsf{M}}(x,p)\ .$$

RMHMC builds a **Markov chain** $(x_n, p_n)_{n\in N}$ via a Gibbs-based scheme. For any $n \geq 1$, given $(x_{n-1}, p_{n-1}) \in \mathrm{T}^\star\mathsf{M}$, we sample

$$\begin{aligned}
&(1) \quad p_n \sim \mathrm{N}_{x_{n-1}}(0,\mathrm{I}_d)\mathrm{d}p\ ,\\
&(2) \quad x_n \sim \mathrm{d}\bar{\pi}(x_n|p_n) \propto \exp[-H(x_n,p_n)]\mathrm{d}\mathrm{vol}_{\mathsf{M}}(x_n)\ .
\end{aligned}$$

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on $\mathrm{T}^\star \mathrm{M}$ is given by

$$\mathrm{d}\bar{\pi}(x,p) \overset{\text{def}}{=} \mathrm{d}\pi(x)\mathrm{N}_x(p;0,\mathrm{I}_d)\mathrm{d}p \propto \exp[-H(x,p)]\mathrm{dvol}_{\mathrm{T}^\star\mathrm{M}}(x,p) \ .$$

RMHMC builds a **Markov chain** $(x_n, p_n)_{n\in N}$ via a Gibbs-based scheme. For any $n \geq 1$, given $(x_{n-1}, p_{n-1}) \in \mathrm{T}^\star\mathrm{M}$, we sample

$$\begin{aligned} &(1) \quad p_n \sim \mathrm{N}_{x_{n-1}}(0,\mathrm{I}_d)\mathrm{d}p \ , \\ &(2) \quad x_n \sim \mathrm{d}\bar{\pi}(x_n|p_n) \propto \exp[-H(x_n,p_n)]\mathrm{dvol}_{\mathrm{M}}(x_n) \ . \end{aligned}$$

Step (2) in theory: we want to compute a **Markov kernel** that leaves $\bar{\pi}(\cdot|p_n)$ invariant on M. Denote $z = (x_{n-1}, p_n)$ and define

- A proposal distribution $\mathrm{d}q_z(z')$ which preserves volume.
- The acceptance ratio $a(z \to z') = \min\left(1, \frac{\bar{\pi}(z')q_{z'}(z)}{\bar{\pi}(z)q_z(z')}\right)$.

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on $\mathrm{T}^\star\mathsf{M}$ is given by

$$\mathrm{d}\bar{\pi}(x,p) \overset{\text{def}}{=} \mathrm{d}\pi(x)\mathrm{N}_x(p;0,\mathrm{I}_d)\mathrm{d}p \propto \exp[-H(x,p)]\mathrm{dvol}_{\mathrm{T}^\star\mathsf{M}}(x,p) \ .$$

RMHMC builds a **Markov chain** $(x_n,p_n)_{n\in N}$ via a Gibbs-based scheme.
For any $n\geq 1$, given $(x_{n-1},p_{n-1})\in \mathrm{T}^\star\mathsf{M}$, we sample

$$\begin{aligned} &(1) \quad p_n \sim \mathrm{N}_{x_{n-1}}(0,\mathrm{I}_d)\mathrm{d}p \ , \\ &(2) \quad x_n \sim \mathrm{d}\bar{\pi}(x_n|p_n) \propto \exp[-H(x_n,p_n)]\mathrm{dvol}_{\mathsf{M}}(x_n) \ . \end{aligned}$$

Step (2) in theory: we want to compute a **Markov kernel** that leaves $\bar{\pi}(\cdot|p_n)$ invariant on M. Denote $z=(x_{n-1},p_n)$ and define

- A proposal distribution $\mathrm{d}q_z(z')$ which preserves volume.
- The acceptance ratio $a(z\to z') = \min\left(1, \frac{\bar{\pi}(z')q_{z'}(z)}{\bar{\pi}(z)q_z(z')}\right)$.

Step (2) in practice:

(a) Sample $z'\sim q_z$ and $U\sim U[0,1]$.
(b) If $U\leq a$, set $z^\star = z'$ (*accepted*); otherwise, set $z^\star = z$ (*rejected*).

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on $\mathrm{T}^\star\mathrm{M}$ is given by

$$\mathrm{d}\bar{\pi}(x,p) \overset{\text{def}}{=} \mathrm{d}\pi(x)\mathrm{N}_x(p;0,\mathrm{I}_d)\mathrm{d}p \propto \exp[-H(x,p)]\mathrm{dvol}_{\mathrm{T}^\star\mathrm{M}}(x,p)\ .$$

RMHMC builds a **Markov chain** $(x_n,p_n)_{n\in N}$ via a Gibbs-based scheme.
For any $n\geq 1$, given $(x_{n-1},p_{n-1})\in\mathrm{T}^\star\mathrm{M}$, we sample

$$\begin{aligned} &(1) \quad p_n \sim \mathrm{N}_{x_{n-1}}(0,\mathrm{I}_d)\mathrm{d}p\ ,\\ &(2) \quad x_n \sim \mathrm{d}\bar{\pi}(x_n|p_n) \propto \exp[-H(x_n,p_n)]\mathrm{dvol}_{\mathrm{M}}(x_n)\ . \end{aligned}$$

Step (2) in theory: we want to compute a **Markov kernel** that leaves $\bar{\pi}(\cdot|p_n)$ invariant on M. Denote $z=(x_{n-1},p_n)$ and define

- A proposal distribution $\mathrm{d}q_z(z')$ which preserves volume.
- The acceptance ratio $a(z\to z') = \min\left(1, \frac{\bar{\pi}(z')q_{z'}(z)}{\bar{\pi}(z)q_z(z')}\right)$.

Step (2) in practice:

(a) Sample $z'\sim q_z$ and $U\sim U[0,1]$.

(b) If $U\leq a$, set $z^\star = z'$ (*accepted*); otherwise, set $z^\star = z$ (*rejected*).

This is a Metropolis-Hastings Markov kernel $\mathrm{d}Q_z(z^\star)$: $\bar{\pi}$-reversible.
$\implies$ $(\mathrm{proj}_x)_\# Q_z$ leaves $\bar{\pi}(\cdot|p_n)$ invariant !

Description of RMHMC (Girolami and Calderhead, 2011)

In most cases, q_z is chosen *stochastic*. In RMHMC, q_z is deterministic, **defined by a map** $\mathrm{F} : \mathrm{T}^\star\mathrm{M} \to \mathrm{T}^\star\mathrm{M}$, i.e., $\mathrm{d}q_z(z') = \mathrm{d}\delta_{\mathrm{F}(z)}(z')$.

Description of RMHMC (Girolami and Calderhead, 2011)

In most cases, q_z is chosen *stochastic*. In RMHMC, q_z is deterministic, **defined by a map** $\mathrm{F} : \mathrm{T}^\star\mathsf{M} \to \mathrm{T}^\star\mathsf{M}$, i.e., $\mathrm{d}q_z(z') = \mathrm{d}\delta_{\mathrm{F}(z)}(z')$.

Ideal setting: we choose $\mathrm{F} = s \circ \Psi_t$ for some $t > 0$. Let $z' = \mathrm{F}(z)$.

- F is symplectic, $q_z(z') = 1$ and $q_{z'}(z) = \delta_{(\mathrm{F}\circ\mathrm{F})(z)}(z) = 1$.
- $\pi(z') = \pi(z)$ by conservation of the Hamiltonian.

Description of RMHMC (Girolami and Calderhead, 2011)

In most cases, q_z is chosen *stochastic*. In RMHMC, q_z is deterministic, **defined by a map** $\mathrm{F} : \mathrm{T}^\star\mathrm{M} \to \mathrm{T}^\star\mathrm{M}$, i.e., $\mathrm{d}q_z(z') = \mathrm{d}\delta_{\mathrm{F}(z)}(z')$.

Ideal setting: we choose $\mathrm{F} = s \circ \Psi_t$ for some $t > 0$. Let $z' = \mathrm{F}(z)$.

- F is symplectic, $q_z(z') = 1$ and $q_{z'}(z) = \delta_{(\mathrm{F}\circ\mathrm{F})(z)}(z) = 1$.
- $\pi(z') = \pi(z)$ by conservation of the Hamiltonian.

In this case, $a(z \to z') = 1$; we just have to follow the Hamiltonian flow !
However, F cannot be computed exactly...

Description of RMHMC (Girolami and Calderhead, 2011)

In most cases, q_z is chosen *stochastic*. In RMHMC, q_z is deterministic, **defined by a map** $\mathrm{F} : \mathrm{T}^\star\mathsf{M} \to \mathrm{T}^\star\mathsf{M}$, i.e., $\mathrm{d}q_z(z') = \mathrm{d}\delta_{\mathrm{F}(z)}(z')$.

Ideal setting: we choose $\mathrm{F} = s \circ \Psi_t$ for some $t > 0$. Let $z' = \mathrm{F}(z)$.

- F is symplectic, $q_z(z') = 1$ and $q_{z'}(z) = \delta_{(\mathrm{F}\circ\mathrm{F})(z)}(z) = 1$.
- $\pi(z') = \pi(z)$ by conservation of the Hamiltonian.

In this case, $a(z \to z') = 1$; we just have to follow the Hamiltonian flow !
However, F cannot be computed exactly...

Let $h > 0$ be a step-size. Consider $\mathrm{T}_h \approx \Psi_h$ a numerical integrator such that T_h is **symplectic** and $s \circ \mathrm{T}_h$ is **involutive** (T_h is said to be reversible).

Description of RMHMC (Girolami and Calderhead, 2011)

In most cases, q_z is chosen *stochastic*. In RMHMC, q_z is deterministic, **defined by a map** $\mathrm{F} : \mathrm{T}^\star\mathrm{M} \to \mathrm{T}^\star\mathrm{M}$, i.e., $\mathrm{d}q_z(z') = \mathrm{d}\delta_{\mathrm{F}(z)}(z')$.

Ideal setting: we choose $\mathrm{F} = s \circ \Psi_t$ for some $t > 0$. Let $z' = \mathrm{F}(z)$.

- F is symplectic, $q_z(z') = 1$ and $q_{z'}(z) = \delta_{(\mathrm{F}\circ\mathrm{F})(z)}(z) = 1$.
- $\pi(z') = \pi(z)$ by conservation of the Hamiltonian.

In this case, $a(z \to z') = 1$; we just have to follow the Hamiltonian flow ! However, F cannot be computed exactly...

Let $h > 0$ be a step-size. Consider $\mathrm{T}_h \approx \Psi_h$ a numerical integrator such that T_h is **symplectic** and $s \circ \mathrm{T}_h$ is **involutive** (T_h is said to be reversible).

Realistic setting: we choose $\mathrm{F} = s \circ \mathrm{T}_h$. Let $z' = \mathrm{F}(z)$.

- F is symplectic, $q_z(z') = 1$ and $q_{z'}(z) = \delta_{(\mathrm{F}\circ\mathrm{F})(z)}(z) = 1$.
- However, $\pi(z') \neq \pi(z)$ due to ODE integration error.

Description of RMHMC (Girolami and Calderhead, 2011)

In most cases, q_z is chosen *stochastic*. In RMHMC, q_z is deterministic, **defined by a map** $\mathrm{F} : \mathrm{T}^\star\mathrm{M} \to \mathrm{T}^\star\mathrm{M}$, i.e., $\mathrm{d}q_z(z') = \mathrm{d}\delta_{\mathrm{F}(z)}(z')$.

Ideal setting: we choose $\mathrm{F} = s \circ \Psi_t$ for some $t > 0$. Let $z' = \mathrm{F}(z)$.

- F is symplectic, $q_z(z') = 1$ and $q_{z'}(z) = \delta_{(\mathrm{F}\circ\mathrm{F})(z)}(z) = 1$.
- $\pi(z') = \pi(z)$ by conservation of the Hamiltonian.

In this case, $a(z \to z') = 1$; we just have to follow the Hamiltonian flow !
However, F cannot be computed exactly...

Let $h > 0$ be a step-size. Consider $\mathrm{T}_h \approx \Psi_h$ a numerical integrator such that T_h is **symplectic** and $s \circ \mathrm{T}_h$ is **involutive** (T_h is said to be reversible).

Realistic setting: we choose $\mathrm{F} = s \circ \mathrm{T}_h$. Let $z' = \mathrm{F}(z)$.

- F is symplectic, $q_z(z') = 1$ and $q_{z'}(z) = \delta_{(\mathrm{F}\circ\mathrm{F})(z)}(z) = 1$.
- However, $\pi(z') \neq \pi(z)$ due to ODE integration error.

In this case, the **acceptance ratio** simplifies as

$$a(z \to z') = \min\left(1, \frac{\exp(-H(z'))}{\exp(H(z))}\right).$$

Description of RMHMC (Girolami and Calderhead, 2011)

Algorithm 1: RMHMC (Girolami and Calderhead, 2011)

HMC Input: $(x_0, p_0) \in \mathrm{T}^\star \mathrm{M}$, $\beta \in (0,1]$, $N \in \mathbb{N}^*$

ODE Input: $h > 0$, $K \in \mathbb{N}^*$, numerical integrator $\mathrm{T}_h : \mathrm{T}^\star \mathrm{M} \to \mathrm{T}^\star \mathrm{M}$

Output: $(x_n, p_n)_{n \in [N]}$

for $n = 1, \ldots, N$ **do**

    Step 1: momentum sampling with refresh
    $\tilde{p} \sim \mathrm{N}(0, \mathfrak{g}(x_{n-1}))$, $p_{n-1} \leftarrow \sqrt{1-\beta} p_{n-1} + \sqrt{\beta} \tilde{p}$
    Step 2: performing K steps of discretized ODE (1) with T_h
    $(x', p') \leftarrow (\mathrm{T}_h)^K (x_{n-1}, p_{n-1}) \quad \Longrightarrow \quad (x', p') \approx \Psi_{Kh}(x_{n-1}, p_{n-1})$
    Step 3: applying the Metropolis-Hastings (MH) acceptance filter
    $a \leftarrow \min(1, \exp[-H(x', p') + H(x_{n-1}, p_{n-1})])$, $\quad u \sim U[0,1]$
    if $u \leq a$ **then** $\bar{x}_n, \bar{p}_n \leftarrow x', p'$;
    else $\bar{x}_n, \bar{p}_n \leftarrow x_{n-1}, p_{n-1}$;
    Step 4: flipping the sign of the momentum
    $x_n, p_n \leftarrow \bar{x}_n, -\bar{p}_n$ (guarantees reversibility and exploration)

Results on RMHMC

Originally, Girolami and Calderhead (2011) considered posterior distributions from **Bayesian models** and chose:

- $\mathfrak{g}$ as the **Fisher-Rao** metric,
- T_h as the **Leapfrog** integrator (Hairer et al., 2006) with fixed-point steps.

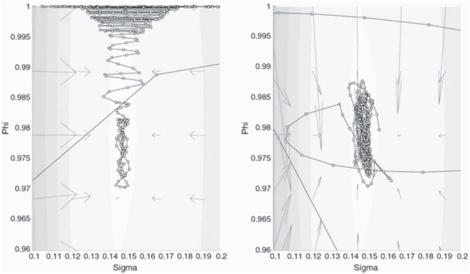


Figure: Figure 4 in Girolami and Calderhead (2011): HMC (left) vs RMHMC (right).

Results on RMHMC

Originally, Girolami and Calderhead (2011) considered posterior distributions from **Bayesian models** and chose:

- $\mathfrak{g}$ as the **Fisher-Rao** metric,
- T_h as the **Leapfrog** integrator (Hairer et al., 2006) with fixed-point steps.

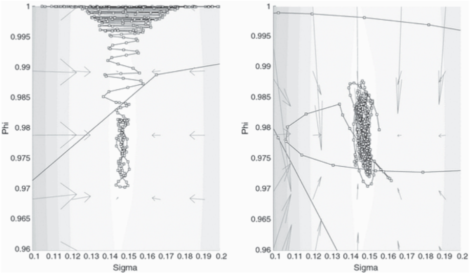


Figure: Figure 4 in Girolami and Calderhead (2011): HMC (left) vs RMHMC (right).

Q1: Can we enlarge the design of $\mathfrak{g}$ to geometric constraints ?

Q2: Can we derive theoretical results from the properties of M, $\mathfrak{g}$ and T_h ?

Introducing self-concordance

If M is convex, one can design a ***self-concordant barrier*** ϕ on M (Nesterov and Nemirovskii, 1994) and set $\mathfrak{g} = D^2\phi$, as done by Kook et al. (2022).

Definition 1 (Self-concordance, Nesterov and Nemirovskii (1994))

Let U be a non-empty open convex domain in $\mathbb{R}^d$. A function ϕ : U $\to \mathbb{R}$ is said to be a ν-**self-concordant** (s.-c.) **barrier** (with $\nu \geq 1$) on U if it satisfies:

(a) $\phi \in \mathrm{C}^3(\mathsf{U}, \mathbb{R})$ and ϕ is convex,

(b) $\phi(x) \to +\infty$ as $x \to \partial$U,

(c) Other technical conditions on $D^3\phi, D^2\phi, D\phi$:

$$\approx D^2\phi \text{ is 2-Lipschitz and } \approx \phi \text{ is } \nu\text{-Lipschitz .}$$

Introducing self-concordance

If M is convex, one can design a ***self-concordant barrier*** ϕ on M (Nesterov and Nemirovskii, 1994) and set $\mathfrak{g} = D^2\phi$, as done by Kook et al. (2022).

➔ S.-c. barriers are well-suited for their minimization by the Newton method.

➔ The analysis of the convergence of Newton methods based on s.-c. is based on the metric $\mathfrak{g}(x) = D^2\phi(x)$.

➔ Balls for $\|\cdot\|_{\mathfrak{g}(x)}$ (Dikin ellipsoids) are central for the study of s.-c.

Self-concordance on a polytope

Assume that M is the **polytope** $\mathrm{M} = \{x \in \mathbb{R}^d \ : \ \mathrm{A}x < b\}$, $\mathrm{A} \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$.

The **logarithmic barrier** on M is given by

$$\phi(x) = -\sum_{i=1}^{m} \ln\left(b_i - \mathrm{A}_i^\top x\right) \ ,$$

and verifies (Nesterov and Nemirovski, 1998):

- ϕ is a m-**self-concordant barrier**.

Self-concordance on a polytope

Assume that M is the **polytope** $\mathsf{M} = \{x \in \mathbb{R}^d \ : \ \mathrm{A}x < b\}$, $\mathrm{A} \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$.

The **logarithmic barrier** on M is given by

$$\phi(x) = -\sum_{i=1}^{m} \ln\left(b_i - \mathrm{A}_i^\top x\right) \ ,$$

and verifies (Nesterov and Nemirovski, 1998):

- ϕ is a m-**self-concordant barrier**.

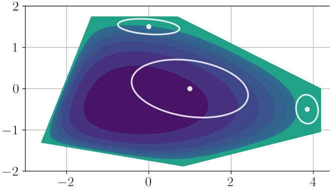


Figure: A self-concordant (logarithmic) barrier for a polytope $\mathsf{M} \subset \mathbb{R}^2$ with three Dikin ellipsoids $\{y \in \mathbb{R}^d \ : \ y^\top \mathfrak{g}(x) y < 1\}$ centered at $x = (0,0), (0,1.5), (1.5,-1)$.

Self-concordance in sampling algorithms

Self-concordant barriers provide theoretical guarantees for **polytope sampling**:

- **Dikin Walk** (Kannan and Narayanan, 2009) $\rightarrow$ *no experiment*
- **Geodesic Walk** (Lee and Vempala, 2017a) $\rightarrow$ *no experiment*
- RMHMC with a metric derived from a s.-c. barrier:

[1] `https://github.com/ConstrainedSampler/PolytopeSamplerMatlab`

Self-concordance in sampling algorithms

Self-concordant barriers provide theoretical guarantees for **polytope sampling**:

- **Dikin Walk** (Kannan and Narayanan, 2009) $\rightarrow$ *no experiment*
- **Geodesic Walk** (Lee and Vempala, 2017a) $\rightarrow$ *no experiment*
- RMHMC with a metric derived from a s.-c. barrier:

  RHMC (Lee and Vempala, 2018) $\rightarrow$ *no experiment*:
  - Consider the *time-continuous Hamiltonian dynamics*.
  - Assume that it exists for all time and is unique: hard to verify!
  - This assumption is not verified in the paper.

[1] `https://github.com/ConstrainedSampler/PolytopeSamplerMatlab`

Self-concordance in sampling algorithms

Self-concordant barriers provide theoretical guarantees for **polytope sampling**:

- **Dikin Walk** (Kannan and Narayanan, 2009) $\rightarrow$ *no experiment*
- **Geodesic Walk** (Lee and Vempala, 2017a) $\rightarrow$ *no experiment*
- RMHMC with a metric derived from a s.-c. barrier:

  RHMC (Lee and Vempala, 2018) $\rightarrow$ *no experiment*:

  - Consider the *time-continuous Hamiltonian dynamics*.
  - Assume that it exists for all time and is unique: hard to verify!
  - This assumption is not verified in the paper.

  CRHMC (Kook et al., 2022):

  - Here, T_h computes *exact* solutions of an **implicit** scheme Φ_h, that is proved to *symplectic* and *reversible*.
  - Assume that Φ_h admits a unique solution for any initial point $\implies$ this guarantees reversibility as explained before.
  - However, this assumption is not verified in practice.
  - Their experiments[1] highlight **asymptotic bias**!

[1] https://github.com/ConstrainedSampler/PolytopeSamplerMatlab

Self-concordance in sampling algorithms

Self-concordant barriers provide theoretical guarantees for **polytope sampling**:

- **Dikin Walk** (Kannan and Narayanan, 2009) $\rightarrow$ *no experiment*
- **Geodesic Walk** (Lee and Vempala, 2017a) $\rightarrow$ *no experiment*
- RMHMC with a metric derived from a s.-c. barrier:

  RHMC (Lee and Vempala, 2018) $\rightarrow$ *no experiment*:
  - Consider the *time-continuous Hamiltonian dynamics*.
  - Assume that it exists for all time and is unique: hard to verify!
  - This assumption is not verified in the paper.

  CRHMC (Kook et al., 2022):
  - Here, T_h computes *exact* solutions of an **implicit** scheme Φ_h, that is proved to *symplectic* and *reversible*.
  - Assume that Φ_h admits a unique solution for any initial point $\Longrightarrow$ this guarantees reversibility as explained before.
  - However, this assumption is not verified in practice.
  - Their experiments[1] highlight **asymptotic bias**!

Can we have better practical and theoretical guarantees ?

[1] https://github.com/ConstrainedSampler/PolytopeSamplerMatlab

Self-concordance in sampling algorithms

Self-concordant barriers provide theoretical guarantees for **polytope sampling**:

- RMHMC with a metric derived from a s.-c. barrier:

RHMC (Lee and Vempala, 2018) $\rightarrow$ *no experiment*:

- Consider the *time-continuous Hamiltonian dynamics*.
- Assume that it exists for all time and is unique: hard to verify!
- This assumption is not verified in the paper.

CRHMC (Kook et al., 2022):

- Here, T_h computes *exact* solutions of an implicit scheme Φ_h, that is proved to *symplectic* and *reversible*.
- Assume that Φ_h admits a unique solution for any initial point $\Longrightarrow$ this guarantees reversibility as explained before.
- However, this assumption is not verified in practice.
- Their experiments[2] highlight **asymptotic bias**!

Can we have better practical and theoretical guarantees ?

Zappa et al. (2018) tackle a similar bias for Ball Walk by **enforcing the reversibility** of the Markov kernel with an "involution checking".

[2] https://github.com/ConstrainedSampler/PolytopeSamplerMatlab

What is at stake ?

- ➔ Traditional MCMC approaches are **not efficient**.
- ➔ RMHMC implemented with $\mathfrak{g}$ may work but comes with **asymptotic bias**.
- ➔ This bias could be tackled by an "**involution checking step**" (**ICS**).

Can we implement this check in RMHMC and derive satisfying theoretical and numerical results ? $\rightarrow$ M.N., Valentin de Bortoli, Alain Durmus (NeurIPS, 2023). **Unbiased constrained sampling with Self-Concordant Barrier Hamiltonian Monte Carlo (BHMC)**.

Notation and reminders

- **Momentum reversal** operator: $s(x,p) = (x,-p)$.

Definition 2 (Reversibility up to momentum reversal.)

Let $\mathrm{Q} : \mathrm{T}^\star\mathrm{M} \times \mathcal{B}(\mathrm{T}^\star\mathrm{M}) \to [0,1]$ be a transition probability kernel and let $\bar{\pi}$ be a probability distribution on $\mathrm{T}^\star\mathrm{M}$. Then, Q is said to be reversible up to momentum reversal with respect to $\bar{\pi}$ if for any $f \in \mathrm{C}(\mathrm{T}^\star\mathrm{M} \times \mathrm{T}^\star\mathrm{M}, \mathbb{R})$ with compact support

$$\int_{\mathrm{T}^\star\mathrm{M}\times\mathrm{T}^\star\mathrm{M}} f(z,z')\mathrm{Q}(z,\mathrm{d}z')\bar{\pi}(\mathrm{d}z) = \int_{\mathrm{T}^\star\mathrm{M}\times\mathrm{T}^\star\mathrm{M}} f(s(z'),s(z))\mathrm{Q}(z,\mathrm{d}z')\bar{\pi}(\mathrm{d}z) \; .$$

Notation and reminders

- **Momentum reversal** operator: $s(x, p) = (x, -p)$.
- **Pushforward**: $\varphi_{\#}\mu \in \mathscr{P}(\mathsf{Y})$ is the pushforward of $\mu \in \mathscr{P}(\mathsf{X})$ by $\varphi : \mathsf{X} \to \mathsf{Y}$.

Lemma 3 (Preservation of measure.)

Let $\mathrm{Q} : \mathrm{T}^{\star}\mathrm{M} \times \mathcal{B}(\mathrm{T}^{\star}\mathrm{M}) \to [0, 1]$ be a transition probability kernel and let $\bar{\pi}$ be a probability distribution on $\mathrm{T}^{\star}\mathrm{M}$. Assume that $s_{\#}\bar{\pi} = \bar{\pi}$ and that Q is reversible up to momentum reversal with respect to $\bar{\pi}$. Then $\bar{\pi}$ is an invariant measure for Q.

Outline

Explicit and implicit integrators

How to integrate the ODEs given by the Riemannian Hamiltonian H ?

As done by Shahbaba et al. (2014), we **split** the Hamiltonian $H = H_1 + H_2$,

$$H_1(x,p) = V(x) + \frac{1}{2}\log(\det \mathfrak{g}(x)) \;, \qquad \text{(separable)}$$

$$H_2(x,p) = \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}} \;. \qquad \text{(non separable)}$$

Explicit and implicit integrators

How to integrate the ODEs given by the Riemannian Hamiltonian H ?

As done by Shahbaba et al. (2014), we **split** the Hamiltonian $H = H_1 + H_2$,

$$H_1(x,p) = V(x) + \tfrac{1}{2}\log(\det \mathfrak{g}(x)) \ , \qquad \text{(separable)}$$
$$H_2(x,p) = \tfrac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}} \ . \qquad \text{(non separable)}$$

➔ **Explicit integrator of H_1**
$\mathrm{S}_{h/2} : \mathrm{T}^\star\mathrm{M} \to \mathrm{T}^\star\mathrm{M}$ is the map defined by $\mathrm{S}_{h/2}(x,p) = (x, p - \frac{h}{2}\partial_x H_1(x,p))$
$\implies$ $\mathrm{S}_{h/2}$ is **symplectic** and **reversible**.

- $\mathrm{S}_{h/2}$ approximates the dynamics of H_1 on a step-size $h/2$.
- $s \circ \mathrm{S}_{h/2}$ is an involution.

Explicit and implicit integrators

How to integrate the ODEs given by the Riemannian Hamiltonian H ?

As done by Shahbaba et al. (2014), we **split** the Hamiltonian $H = H_1 + H_2$,

$$H_2(x, p) = \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}} \; . \qquad \text{(non separable)}$$

Explicit and implicit integrators

How to integrate the ODEs given by the Riemannian Hamiltonian H ?

As done by Shahbaba et al. (2014), we **split** the Hamiltonian $H = H_1 + H_2$,

$$H_2(x,p) = \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}} \, . \qquad \text{(non separable)}$$

➔ **Implicit integrator of H_2**

G_h is the Leapfrog integrator of H_2 with step-size h: for any $z^{(0)} \in \mathrm{T}^\star\mathrm{M}$, $\mathrm{G}_h(z^{(0)}) \subset \mathrm{T}^\star\mathrm{M}$ consists of points $z^{(1)} = (x^{(1)}, p^{(1)})$ that solve

$$\begin{aligned} p^{(1/2)} &= p^{(0)} - \tfrac{h}{2}\partial_x H_2(x^{(0)}, p^{(1/2)}) \, , \\ x^{(1)} &= x^{(0)} + \tfrac{h}{2}[\partial_p H_2(x^{(0)}, p^{(1/2)}) + \partial_p H_2(x^{(1)}, p^{(1/2)})] \, , \\ p^{(1)} &= p^{(1/2)} - \tfrac{h}{2}\partial_x H_2(x^{(1)}, p^{(1/2)}) \, . \end{aligned}$$

Explicit and implicit integrators

How to integrate the ODEs given by the Riemannian Hamiltonian H ?

As done by Shahbaba et al. (2014), we **split** the Hamiltonian $H = H_1 + H_2$,

$$H_2(x,p) = \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}} \; . \qquad \text{(non separable)}$$

➔ **Implicit integrator of H_2**

G_h is the Leapfrog integrator of H_2 with step-size h: for any $z^{(0)} \in \mathrm{T}^\star\mathrm{M}$, $\mathrm{G}_h(z^{(0)}) \subset \mathrm{T}^\star\mathrm{M}$ consists of points $z^{(1)} = (x^{(1)}, p^{(1)})$ that solve

$$\begin{aligned} p^{(1/2)} &= p^{(0)} - \frac{h}{2}\partial_x H_2(x^{(0)}, p^{(1/2)}) \; , \\ x^{(1)} &= x^{(0)} + \frac{h}{2}[\partial_p H_2(x^{(0)}, p^{(1/2)}) + \partial_p H_2(x^{(1)}, p^{(1/2)})] \; , \\ p^{(1)} &= p^{(1/2)} - \frac{h}{2}\partial_x H_2(x^{(1)}, p^{(1/2)}) \; . \end{aligned}$$

$\Longrightarrow$ G_h is **symplectic** and **reversible** (Hairer et al., 2006)

$\Longrightarrow$ above, there may be either 0, 1, 2,... solutions !

- $\mathrm{F}_h = \mathrm{G}_h \circ s$ is a *set-valued* map.
- $\mathrm{F}_h \circ s$ approximates the dynamics of H_2 on a step-size h.
- If $|\mathrm{F}_h(z)| > 0$ then $z \in (\mathrm{F}_h \circ \mathrm{F}_h)(z)$ (almost an involution).

Explicit and implicit integrators

How to integrate the ODEs given by the Riemannian Hamiltonian H ?

As done by Shahbaba et al. (2014), we **split** the Hamiltonian $H = H_1 + H_2$,

$$H_1(x,p) = V(x) + \frac{1}{2}\log(\det \mathfrak{g}(x)) \ , \qquad \text{(separable)}$$

$$H_2(x,p) = \frac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}} \ . \qquad \text{(non separable)}$$

➔ **Explicit integrator of H_1**

- Map $\mathrm{S}_{h/2} : \mathrm{T}^\star\mathrm{M} \to \mathrm{T}^\star\mathrm{M}$ defined by $\mathrm{S}_{h/2}(x,p) = (x, p - \frac{h}{2}\partial_x H_1(x,p))$.

➔ ***Implicit* integrator of H_2**

- Set-valued map $\mathrm{F}_h = \mathrm{G}_h \circ s$, where $\mathrm{G}_h : \mathrm{T}^\star\mathrm{M} \to 2^{\mathrm{T}^\star\mathrm{M}}$ is the Leapfrog integrator of H_2 with step-size h (none, one or multiple solutions).

Explicit and implicit integrators

How to integrate the ODEs given by the Riemannian Hamiltonian H ?

As done by Shahbaba et al. (2014), we **split** the Hamiltonian $H = H_1 + H_2$,

$$H_1(x,p) = V(x) + \tfrac{1}{2}\log(\det \mathfrak{g}(x)) \ , \qquad \text{(separable)}$$

$$H_2(x,p) = \tfrac{1}{2}\|p\|^2_{\mathfrak{g}(x)^{-1}} \ . \qquad \text{(non separable)}$$

➔ **Explicit integrator of H_1**

- Map $\mathrm{S}_{h/2} : \mathrm{T}^\star\mathrm{M} \to \mathrm{T}^\star\mathrm{M}$ defined by $\mathrm{S}_{h/2}(x,p) = (x, p - \frac{h}{2}\partial_x H_1(x,p))$.

➔ ***Implicit* integrator of H_2**

- Set-valued map $\mathrm{F}_h = \mathrm{G}_h \circ s$, where $\mathrm{G}_h : \mathrm{T}^\star\mathrm{M} \to 2^{\mathrm{T}^\star\mathrm{M}}$ is the Leapfrog integrator of H_2 with step-size h (none, one or multiple solutions).

➔ ***Implicit* integrator of H**

- Set-valued map $\mathrm{R}_h = (s \circ \mathrm{S}_{h/2}) \circ \mathrm{F}_h \circ (s \circ \mathrm{S}_{h/2})$.
- By composition, R_h is **symplectic** and **reversible**.
- $s \circ \mathrm{R}_h$ approximates the dynamics of H on a step-size h.

Numerical integrators

In practice, we do not have access to F_h but approximate it with a **numerical map** Φ_h, defined on a domain $\mathrm{dom}_{\Phi_h} \subset \mathrm{T}^\star\mathsf{M}$ with $\Phi_h(\mathrm{dom}_{\Phi_h}) \subset \mathrm{T}^\star\mathsf{M}$.

We also define the **numerical map** $\mathrm{R}_h^\Phi : (s \circ \mathrm{S}_{h/2})(\mathrm{dom}_{\Phi_h}) \subset \mathrm{T}^\star\mathsf{M} \to \mathrm{T}^\star\mathsf{M}$

$$\mathrm{R}_h^\Phi = (s \circ \mathrm{S}_{h/2}) \circ \Phi_h \circ (s \circ \mathrm{S}_{h/2}) \ .$$

Similarly to R_h, $s \circ \mathrm{R}_h^\Phi$ approximates the dynamics of H on a step-size h.

How to implement Φ_h ?

- Fixed-point solver (ours, Kook et al. (2022)).
- Newton's solver (Brofos and Lederman, 2021a,b).

Numerical integrators

In practice, we do not have access to F_h but approximate it with a **numerical map** Φ_h, defined on a domain $\mathrm{dom}_{\Phi_h} \subset \mathrm{T}^\star\mathrm{M}$ with $\Phi_h(\mathrm{dom}_{\Phi_h}) \subset \mathrm{T}^\star\mathrm{M}$.

We also define the **numerical map** $\mathrm{R}_h^\Phi : (s \circ \mathrm{S}_{h/2})(\mathrm{dom}_{\Phi_h}) \subset \mathrm{T}^\star\mathrm{M} \to \mathrm{T}^\star\mathrm{M}$

$$\mathrm{R}_h^\Phi = (s \circ \mathrm{S}_{h/2}) \circ \Phi_h \circ (s \circ \mathrm{S}_{h/2}) \; .$$

Similarly to R_h, $s \circ \mathrm{R}_h^\Phi$ approximates the dynamics of H on a step-size h.

How to implement Φ_h ?

- Fixed-point solver (ours, Kook et al. (2022)).
- Newton's solver (Brofos and Lederman, 2021a,b).

For any $z \in \mathrm{T}^\star\mathrm{M}$, we define on $\mathrm{T}^\star\mathrm{M}$ the norm $\|\cdot\|_z$ by

$$\|z'\|_z = \|x'\|_{\mathfrak{g}(x)} + \|p'\|_{\mathfrak{g}(x)^{-1}} \; , z' = (x', p') \; .$$

- On x': the Dikin norm.
- On p': the “natural“ norm induced by $\mathfrak{g}(x)^{-1}$.

BHMC: CRHMC with “involution checking step”

Algorithm 2: Barrier HMC (BHMC)

HMC Input: $(x_0, p_0) \in \mathrm{T}^\star\mathsf{M}$, $\beta \in (0,1]$, $N \in \mathbb{N}$
ODE Input: $h > 0$, $\eta > 0$, numerical integrator Φ_h with domain dom_{Φ_h}
Output: $(x_n, p_n)_{n \in [N]}$

for $n = 1, \ldots, N$ **do**
    `Step 1:` $\tilde{p} \sim \mathrm{N}(0, \mathfrak{g}(x_{n-1}))$, $p_{n-1} \leftarrow \sqrt{1-\beta}\, p_{n-1} + \sqrt{\beta}\, \tilde{p}$
    `Step 2: solving discretized ODE (1) with` Φ_h
    $x', p' \leftarrow x_{n-1}, p_{n-1}$, $x^{(0)}, p^{(0)} \leftarrow (s \circ \mathrm{S}_{h/2})(x_{n-1}, p_{n-1})$
    if $z^{(0)} \in \mathrm{dom}_{\Phi_h}$ **then**
        $z^{(1)} = \Phi_h(z^{(0)})$, $\mathrm{err} = \|z^{(0)} - \Phi_h(z^{(1)})\|_{z^{(0)}} + \|z^{(0)} - \Phi_h(z^{(1)})\|_{\Phi_h(z^{(1)})}$
        if $z^{(1)} \in \mathrm{dom}_{\Phi_h}$ **&** $\mathrm{err} \leq \eta$ **then** $x', p' \leftarrow (s \circ \mathrm{S}_{h/2})(x^{(1)}, p^{(1)})$;
    `Step 3:` $a \leftarrow \min(1, \exp[-H(x', p') + H(x_{n-1}, p_{n-1})])$, $u \sim U[0,1]$
    if $u \leq a$ **then** $\bar{x}_n, \bar{p}_n \leftarrow x', p'$;
    else $\bar{x}_n, \bar{p}_n \leftarrow x_{n-1}, p_{n-1}$;
    `Step 4:` $x_n, p_n \leftarrow \bar{x}_n, -\bar{p}_n$ `(guarantees reversibility and exploration)`

➔ Checking that Φ_h is well defined on the iterates.

➔ **New:** checking that $(\Phi_h \circ \Phi_h)(z^{(0)}) \approx z^{(0)} \rightarrow$ involution checking !

Outline

Assumptions

We aim at sampling from a **target distribution** π supported on M

$$\mathrm{d}\pi(x)/\mathrm{d}x \propto \exp[-V(x)] \quad , \quad V \in \mathrm{C}^2(\mathsf{M}, \mathbb{R}) \; .$$

A1 (Assumption on M.)

M *is an open convex bounded subset of* $\mathbb{R}^d$.

A2 (Assumption on $\mathfrak{g}$.)

There exists ϕ, ν*-s.-c. barrier on* M *such that* $\mathfrak{g} = D^2\phi$.

From implicit to numerical integrators

We prove that F_h can be locally identified with a C^1-diffeomorphism.

Proposition 1 (Result on F_h.)

Assume **A**1, **A**2. *Let* $z^{(0)} \in \mathrm{T}^\star\mathsf{M}$, *then there exists* $h^\star > 0$ *(explicit) such that for any* $h \in (0, h^\star)$, *there exist* $z_h^{(1)} \in \mathrm{F}_h(z^{(0)})$, *a neighborhood* $\mathsf{U} \subset \mathrm{T}^\star\mathsf{M}$ *of* $z^{(0)}$ *and a* C^1*-diffeomorphism* $\gamma_h : \mathsf{U} \to \gamma_h(\mathsf{U}) \subset \mathrm{T}^\star\mathsf{M}$ *with*

(a) $\gamma_h(z^{(0)}) = z_h^{(1)}$ *and* $|\det \mathrm{Jac}(\gamma_h)| = 1$.

(b) $\gamma_h(z)$ *is the only element of* $\mathrm{F}_h(z)$ *in* $\gamma_h(\mathsf{U})$ *for* $z \in \mathsf{U}$.

Following Proposition 1, we derive a **technical assumption** on the corresponding numerical integrator Φ_h, denoted by **A**3: basically, we assume that Φ_h is locally involutive.

Deriving the reversibility in BHMC

- $\mathrm{Q}_0 : \mathrm{T}^\star\mathsf{M} \times \mathcal{B}(\mathrm{T}^\star\mathsf{M}) \to [0,1]$: the transition kernel for Step 1.
  $\to$ Q_0 is reversible up to m.-r. w.r.t. $\bar{\pi}$
- $\mathrm{Q}_1 : \mathrm{T}^\star\mathsf{M} \times \mathcal{B}(\mathrm{T}^\star\mathsf{M}) \to [0,1]$, the transition kernel for Steps 2 to 4.
  $\to$ Q_1 is reversible up to m.-r. w.r.t. $\bar{\pi}$ under **A**1, **A**2 and **A**3.
- $\mathrm{Q} : \mathrm{T}^\star\mathsf{M} \times \mathcal{B}(\mathrm{T}^\star\mathsf{M}) \to [0,1]$, the transition kernel for Steps 1 to 4 such that

$$\mathrm{Q}(z, \mathrm{d}z') = \int_{\mathrm{T}^\star\mathsf{M}\times\mathrm{T}^\star\mathsf{M}} \mathrm{Q}_0(z, \mathrm{d}z_1)\mathrm{Q}_1(z_1, \mathrm{d}z') \,.$$

Theorem 4 (Reversibility of Q.)

Assume **A**1, **A**2, ***A**3. Then,* Q *is* ***reversible up to momentum reversal****.*
In particular, $\bar{\pi}$ *is an invariant measure for* Q*.*

Sampling on the simplex

Parameters:

- ➔ $\mathsf{M} = \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i = 1, x_i \geq 0, \forall i \in \{1, \dots, d\}\}$ with $d \in \{5, 10\}$
- ➔ $\eta = 10$ if $d = 5$ and $\eta = 200$ if $d = 10$

We aim to sample from a **truncated Gaussian distribution**, and display the estimated expectation of a fixed observable.

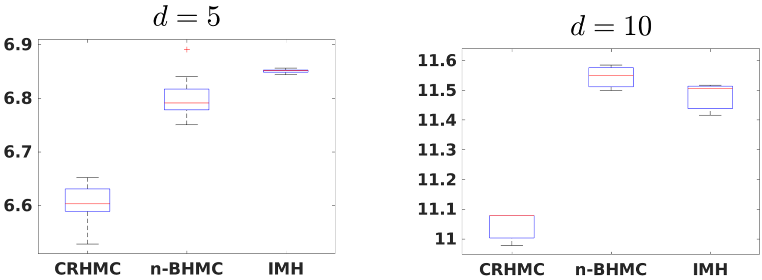


Figure: Comparison between n-BHMC and CRHMC on the simplex.

Outline

1. Motivations and background
2. Description of BHMC
3. Results
4. Conclusion

Conclusion

M.N., Valentin de Bortoli, Alain Durmus (NeurIPS, 2023). **Unbiased constrained sampling with Self-Concordant BHMC.**

➔ We introduced a **novel version of RMHMC**, Barrier HMC, relying on a "involution checking step", to sample from a distribution π over a bounded open convex subset $\mathsf{M} \subset \mathbb{R}^d$ equipped with a self-concordant barrier ϕ.

- **BHMC** approximates the dynamics with a numerical integrator.

➔ We proved that π is **invariant** for BHMC.

➔ We showed that BHMC generates less **asymptotic bias** than the version of RMHMC proposed by Kook et al. (2022) (*see the paper for more details*).

Future work:

$\rightarrow$ Investigate the **"coupled" behaviour** of the hyperparameters h and η.

$\rightarrow$ Study the **irreducibility** of n-BHMC (not easy task).

Unbiased constrained sampling with Self-Concordant Barrier Hamiltonian Monte Carlo

Maxence Noble[1] Valentin de Bortoli[2] Alain Durmus[1]

[1] Centre de Mathématiques Appliquées, Ecole Polytechnique
Institut Polytechnique de Paris, France

[2] Département d'Informatique, École Normale Supérieure
CNRS, Université PSL, Paris, France

James Brofos and Roy R Lederman. Evaluating the implicit midpoint integrator for riemannian hamiltonian monte carlo. In *International Conference on Machine Learning*, pages 1072–1081. PMLR, 2021a.

James A Brofos and Roy R Lederman. On numerical considerations for riemannian manifold hamiltonian monte carlo. *arXiv preprint arXiv:2111.09995*, 2021b.

Marcus Brubaker, Mathieu Salzmann, and Raquel Urtasun. A family of mcmc methods on implicitly defined manifolds. In *Artificial intelligence and statistics*, pages 161–172. PMLR, 2012.

Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

A. E. Gelfand, A. F. Smith, and T.-M. Lee. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, 87(418):523–532, 1992.

Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73 (2):123–214, 2011.

Ernst Hairer, Marlis Hochbruck, Arieh Iserles, and Christian Lubich. Geometric numerical integration. *Oberwolfach Reports*, 3(1):805–882, 2006.

Ravi Kannan and Hariharan Narayanan. Random walks on polytopes and an affine interior point method for linear programming. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 561–570, 2009.

Yunbum Kook, Yin-Tat Lee, Ruoqi Shen, and Santosh Vempala. Sampling with riemannian hamiltonian monte carlo in a constrained space. *Advances in Neural Information Processing Systems*, 35:31684–31696, 2022.

S. Lan and B. Shahbaba. Sampling constrained probability distributions using Spherical Augmentation. *ArXiv e-prints*, June 2015.

John M Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006.

Yin Tat Lee and Santosh S Vempala. Geodesic walks in polytopes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on theory of Computing*, pages 927–940, 2017a.

Yin Tat Lee and Santosh S Vempala. Convergence rate of riemannian hamiltonian monte carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1115–1121, 2018.

Yin Tat Lee and Santosh Srinivas Vempala. Eldan's stochastic localization and the kls hyperplane conjecture: an improved lower bound for expansion. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 998–1007. IEEE, 2017b.

Benedict J Leimkuhler and Robert D Skeel. Symplectic numerical integrators in constrained hamiltonian systems. *Journal of Computational Physics*, 112(1):117–125, 1994.

László Lovász and Santosh Vempala. Hit-and-run from a corner. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 310–314, 2004.

Kam-Ping Mok. Metrics and connections on the cotangent bundle. In *Kodai Mathematical Seminar Reports*, volume 28, pages 226–238. Department of Mathematics, Tokyo Institute of Technology, 1977.

Yu Nesterov and Arkadi Nemirovski. Multi-parameter surfaces of analytic centers and long-step surface-following interior point methods. *Mathematics of operations research*, 23 (1):1–38, 1998.

Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

Ari Pakman and Liam Paninski. Exact hamiltonian monte carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542, 2014.

Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

Babak Shahbaba, Shiwei Lan, Wesley O Johnson, and Radford M Neal. Split hamiltonian monte carlo. *Statistics and Computing*, 24(3):339–349, 2014.

Emilio Zappa, Miranda Holmes-Cerfon, and Jonathan Goodman. Monte carlo on manifolds: sampling densities and integrating functions. *Communications on Pure and Applied Mathematics*, 71(12):2609–2647, 2018.