

Unbiased constrained sampling with Self-Concordant Barrier Hamiltonian Monte Carlo

Maxence Noble¹ Valentin de Bortoli² Alain Durmus¹

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

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



Outline

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

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 $M \subset \mathbb{R}^d$.

Our goal: sample from a target distribution supported on M and known up to a normalising constant Z

$$p(x) = \frac{\exp[-V(x)]}{Z} \quad ; \quad V \in C^2(M; \mathbb{R}) :$$

Constrained sampling

Consider a subset $M \subseteq \mathbb{R}^d$.

Our goal: sample from a target distribution supported on M and known up to a normalising constant Z

$$p(x) \propto \exp[-V(x)] / Z \quad ; \quad V \in C^2(M; \mathbb{R}) :$$

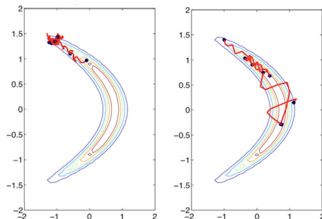
⊙ When $M = \mathbb{R}^d$, gradient-based Markov Chain Monte Carlo (MCMC) methods are very popular and come with some theoretical guarantees under assumptions on V (Duane et al., 1987; Roberts and Tweedie, 1996).

⊙ 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

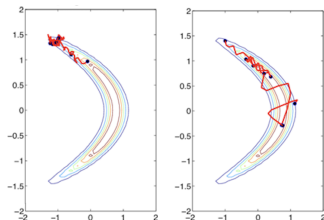
Hit-and-Run (Lovász and Vempala, 2004),
 Ball Walk (Lee and Vempala, 2017b) (*left*),
 Hamiltonian Monte Carlo (HMC) (Duane et al., 1987) (*right*).



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 $M = f_X \mathcal{Z} \mathbb{R}^d : c(x) = 0g$: HMC + RATTLE integrator (Leimkuhler and Skeel, 1994; Brubaker et al., 2012).

If $(M;g)$ 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 g .

Results on $(M; g)$ (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 g .

Results on $(M; g)$ (see [Lee \(2006\)](#) and [Mok \(1977\)](#) for details):

$$\text{Volume element on } M: d\text{vol}_M(x) = \sqrt{\det g(x)} dx.$$

Basics of RMHMC

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

Results on $(M; g)$ (see [Lee \(2006\)](#) and [Mok \(1977\)](#) for details):

Volume element on M : $d\text{vol}_M(x) = \sqrt{\det g(x)} dx$.

T_x^*M : cotangent space at $x \in M$

! $T_x^*M \cong \mathbb{R}^d$, endowed with the scalar product $\langle \cdot, \cdot \rangle_{g(x)}$

! Standard Gaussian distr. w.r.t. $\langle \cdot, \cdot \rangle_{g(x)}$: $N_x(0; I_d) \cong N(0; g(x))$

Basics of RMHMC

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

Results on $(M; g)$ (see [Lee \(2006\)](#) and [Mok \(1977\)](#) for details):

Volume element on M : $d\text{vol}_M(x) = \sqrt{\det g(x)} dx$.

T_x^*M : cotangent space at $x \in M$

! $T_x^*M \cong \mathbb{R}^d$, endowed with the scalar product $\langle \cdot, \cdot \rangle_{g(x)}$

! Standard Gaussian distr. w.r.t. $k_{g(x)} = \mathcal{N}(0; I_d)$ on T_x^*M

T^*M : cotangent bundle of M , defined by $T^*M = \{ (x, p) \mid x \in M, p \in T_x^*M \}$

! $2d$ -dim. submanifold which may be endowed with a specific metric $g^?$ inherited from g which verifies

$$d\text{vol}_{T^*M}(x; p) = \sqrt{\det g^?(x; p)} dx dp = dx dp$$

Basics of RMHMC

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

○ Results on $(M; g)$ (see [Lee \(2006\)](#) and [Mok \(1977\)](#) for details):

Volume element on M : $d\text{vol}_M(x) = \sqrt{\det g(x)} dx$.

Volume element on T^*M : $d\text{vol}_{T^*M}(x; p) = \sqrt{\det g^*(x; p)} dx dp = dx dp$.

Basics of RMHMC

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

○ Results on $(M; g)$ (see [Lee \(2006\)](#) and [Mok \(1977\)](#) for details):

Volume element on M : $d\text{vol}_M(x) = \sqrt{\det g(x)} dx$.

Volume element on $T^?M$: $d\text{vol}_{T^?M}(x; p) = \sqrt{\det g^?(x; p)} dx dp = dx dp$.

○ In terms of probability distributions:

Target measure: $d\pi(x) = d\text{vol}_M(x) \exp[-V(x) - \frac{1}{2} \log(\det g(x))] = Z$.

Hamiltonian on $T^?M$: $H(x; p) \stackrel{\text{def}}{=} V(x) + \frac{1}{2} \log(\det g(x)) + \frac{1}{2} p^T k_{g(x)}^{-1} p$.

Basics of RMHMC

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

○ Results on $(M; g)$ (see [Lee \(2006\)](#) and [Mok \(1977\)](#) for details):

Volume element on M : $d\text{vol}_M(x) = \sqrt{\det g(x)} dx$.

Volume element on $T^?M$: $d\text{vol}_{T^?M}(x; p) = \sqrt{\det g^?(x; p)} dx dp = dx dp$.

○ In terms of probability distributions:

Target measure: $d(x) = d\text{vol}_M(x) = \exp[-V(x) - \frac{1}{2} \log(\det g(x))] dx = Z$.

Hamiltonian on $T^?M$: $H(x; p) \stackrel{\text{def}}{=} V(x) + \frac{1}{2} \log(\det g(x)) + \frac{1}{2} p^T k_{g(x)}^{-1} p$.

RMHMC aims at sampling from the augmented target distribution on $T^?M$

$$d(x; p) \stackrel{\text{def}}{=} d(x) N_x(p; 0; I_d) dp / \exp[-H(x; p)] d\text{vol}_{T^?M}(x; p)$$

If $g(x) = I_d$, we recover the setting of “Euclidean” HMC !

Definition of the Hamiltonian dynamics

HMC Ham. on \mathbb{R}^{2d} : $H(x; p) = V(x) + \frac{1}{2} k p k_2^2$.

RMHMC Ham. on $T^?M$: $H(x; p) = V(x) + \frac{1}{2} \log(\det g(x)) + \frac{1}{2} k p k_{g(x)}^2$ 1.

Definition of the Hamiltonian dynamics

HMC Ham. on \mathbb{R}^{2d} : $H(x; p) = V(x) + \frac{1}{2} k p k_2^2$.

RMHMC Ham. on T^2M : $H(x; p) = V(x) + \frac{1}{2} \log(\det g(x)) + \frac{1}{2} k p k_{g(x)}^2$.

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

$$\dot{x}_t = \partial_p H(x_t; p_t) ; \quad \dot{p}_t = -\partial_x H(x_t; p_t) ; \quad (1)$$

The corresponding flow is denoted by $\mathcal{F}_t : (x_0; p_0) \mapsto (x_t; p_t)$.

Definition of the Hamiltonian dynamics

HMC Ham. on \mathbb{R}^{2d} : $H(x; p) = V(x) + \frac{1}{2} k p k_2^2$.

RMHMC Ham. on T^2M : $H(x; p) = V(x) + \frac{1}{2} \log(\det g(x)) + \frac{1}{2} k p k_{g(x)}^2$.

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

$$\dot{x}_t = \partial_p H(x_t; p_t) ; \quad \dot{p}_t = -\partial_x H(x_t; p_t) ; \quad (1)$$

The corresponding flow is denoted by $\varphi_t : (x_0; p_0) \mapsto (x_t; p_t)$.

⊙ Conservation of the Hamiltonian through time.

⊙ Volume preservation through time: the flow is *symplectic*.

⊙ Time-reversibility: $\varphi_t^{-1} = s \circ \varphi_t \circ s$ where $s(x; p) = (x; -p)$.

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

Definition of the Hamiltonian dynamics

HMC Ham. on \mathbb{R}^{2d} : $H(x; p) = V(x) + \frac{1}{2} k p k_2^2$.

RMHMC Ham. on T^2M : $H(x; p) = V(x) + \frac{1}{2} \log(\det g(x)) + \frac{1}{2} k p k_{g(x)}^2$.

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

$$\dot{x}_t = \partial_p H(x_t; p_t) ; \quad \dot{p}_t = -\partial_x H(x_t; p_t) ; \quad (1)$$

The corresponding flow is denoted by $\varphi_t : (x_0; p_0) \mapsto (x_t; p_t)$.

⊙ Conservation of the Hamiltonian through time.

⊙ Volume preservation through time: the flow is *symplectic*.

⊙ Time-reversibility: $\varphi_t^{-1} = s \circ \varphi_t \circ s$ where $s(x; p) = (x; -p)$.

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

HMC : $\dot{x} = p_t ; \quad \dot{p}_t = -\nabla V(x_t) :$

RMHMC : $\dot{x}_t = g(x_t)^{-1} p_t ; \quad \dot{p}_t = -\nabla V(x_t) - L(x_t) :$

where $L(x) = \frac{1}{2} g(x)^{-1} : Dg(x) + \frac{1}{2} Dg(x)[\dot{x}; \dot{x}]$. The Riemannian metric 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 $T^?M$ is given by

$$d(x; p) \stackrel{\text{def}}{=} d(x) N_x(p; 0; I_d) dp / \exp[-H(x; p)] d\text{vol}_{T^?M}(x; p) :$$

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on $T^?M$ is given by

$$d(x; p) \stackrel{\text{def}}{=} d(x) N_x(p; 0; I_d) dp / \exp[-H(x; p)] d\text{vol}_{T^?M}(x; p) :$$

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

- (1) $p_n \sim N_{x_{n-1}}(0; I_d) dp$;
- (2) $x_n \sim d(x_n | p_n) / \exp[-H(x_n; p_n)] d\text{vol}_M(x_n) :$

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on T^2M is given by

$$d(x; p) \stackrel{\text{def}}{=} d(x) N_x(p; 0; I_d) dp / \exp[-H(x; p)] d\text{vol}_{T^2M}(x; p) :$$

RMHMC builds a Markov chain $(x_n; p_n)_{n \geq N}$ via a Gibbs-based scheme.
 For any $n \geq 1$, given $(x_{n-1}; p_{n-1}) \in T^2M$, we sample

- (1) $p_n \sim N_{x_{n-1}}(0; I_d) dp$;
- (2) $x_n \sim d(x_n | p_n) / \exp[-H(x_n; p_n)] d\text{vol}_M(x_n)$:

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

A proposal distribution $dq_z(z^\theta)$ which preserves volume.

The acceptance ratio $a(z \rightarrow z^\theta) = \min\left(1; \frac{(z^\theta) q_z(z)}{(z) q_z(z^\theta)}\right)$.

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on T^2M is given by

$$d(x; p) \stackrel{\text{def}}{=} d(x) N_x(p; 0; I_d) dp / \exp[-H(x; p)] d\text{vol}_{T^2M}(x; p) :$$

RMHMC builds a Markov chain $(x_n; p_n)_{n \geq N}$ via a Gibbs-based scheme.
 For any $n \geq 1$, given $(x_{n-1}; p_{n-1}) \in T^2M$, we sample

- (1) $p_n \sim N_{x_{n-1}}(0; I_d) dp$;
- (2) $x_n \sim d(x_n | p_n) / \exp[-H(x_n; p_n)] d\text{vol}_M(x_n)$:

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

A proposal distribution $dq_z(z^\theta)$ which preserves volume.

The acceptance ratio $a(z^\theta | z) = \min\{1; \frac{(z^\theta) q_z(z)}{(z) q_z(z^\theta)}\}$.

Step (2) in practice:

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

Description of RMHMC (Girolami and Calderhead, 2011)

We recall that the target distribution on $T^?M$ is given by

$$d(x; p) \stackrel{\text{def}}{=} d(x) N_x(p; 0; I_d) dp / \exp[-H(x; p)] d\text{vol}_{T^?M}(x; p) :$$

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

- (1) $p_n \sim N_{x_{n-1}}(0; I_d) dp$;
- (2) $x_n \sim d(x_n | p_n) / \exp[-H(x_n; p_n)] d\text{vol}_M(x_n)$:

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

A proposal distribution $dq_z(z^j)$ which preserves volume.

The acceptance ratio $a(z^j | z) = \min\{1; \frac{(z^j) q_z(z)}{(z) q_z(z^j)}\}$.

Step (2) in practice:

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

This is a Metropolis-Hastings Markov kernel $dQ_z(z^?)$: -reversible.

$\Rightarrow (\text{proj}_x)_\# Q_z$ leaves $(\cdot | p_n)$ 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 $F : T^2M \rightarrow T^2M$, i.e., $dq_z(z^0) = d_{F(z)}(z^0)$.

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 $F : T^2M \rightarrow T^2M$, i.e., $dq_z(z^0) = d_{F(z)}(z)$.

Ideal setting: we choose $F = S_{-t}$ for some $t > 0$. Let $z^0 = F(z)$.

F is symplectic, $q_z(z^0) = 1$ and $q_{z^0}(z) = (F^{-1})_{(z)}(z^0) = 1$.

$(z^0) = (z)$ 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 $F : T^2M \rightarrow T^2M$, i.e., $dq_z(z^0) = d_{F(z)}(z)$.

Ideal setting: we choose $F = S_t$ for some $t > 0$. Let $z^0 = F(z)$.

F is symplectic, $q_z(z^0) = 1$ and $q_{z^0}(z) = (F^{-1})_{(z)}(z^0) = 1$.

$z^0 = z$ by conservation of the Hamiltonian.

In this case, $a(z, z^0) = 1$; 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 $F : T^2M \rightarrow T^2M$, i.e., $dq_z(z^0) = d_{F(z)}(z^0)$.

Ideal setting: we choose $F = s_{-t}$ for some $t > 0$. Let $z^0 = F(z)$.

F is symplectic, $q_z(z^0) = 1$ and $q_{z^0}(z) = (F^{-1})_{(z)}(z) = 1$.

$z^0 = z$ by conservation of the Hamiltonian.

In this case, $a(z, z^0) = 1$; we just have to follow the Hamiltonian flow !
 However, F cannot be computed exactly...

Let $h > 0$ be a step-size. Consider T_h a numerical integrator such that T_h is symplectic and $s \circ T_h$ 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 $F : T^2M \rightarrow T^2M$, i.e., $dq_z(z^0) = d_{F(z)}(z^0)$.

Ideal setting: we choose $F = s_{-t}$ for some $t > 0$. Let $z^0 = F(z)$.

F is symplectic, $q_z(z^0) = 1$ and $q_{z^0}(z) = (F^{-1})_{(z)}(z) = 1$.

$z^0 = s_{-t}(z)$ by conservation of the Hamiltonian.

In this case, $a(z^0, z) = 1$; we just have to follow the Hamiltonian flow!
 However, F cannot be computed exactly...

Let $h > 0$ be a step-size. Consider $T_h = s_{-h}$ a numerical integrator such that T_h is symplectic and $s \circ T_h$ is involutive (T_h is said to be reversible).

Realistic setting: we choose $F = s \circ T_h$. Let $z^0 = F(z)$.

F is symplectic, $q_z(z^0) = 1$ and $q_{z^0}(z) = (F^{-1})_{(z)}(z) = 1$.

However, $z^0 \neq s_{-t}(z)$ 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 $F : T^2M \rightarrow T^2M$, i.e., $dq_z(z^0) = d_{F(z)}(z)$.

Ideal setting: we choose $F = s_{-t}$ for some $t > 0$. Let $z^0 = F(z)$.

F is symplectic, $q_z(z^0) = 1$ and $q_{z^0}(z) = (F^{-1})_{(z)}(z) = 1$.

$(z^0) = (z)$ by conservation of the Hamiltonian.

In this case, $a(z \rightarrow z^0) = 1$; we just have to follow the Hamiltonian flow!

However, F cannot be computed exactly...

Let $h > 0$ be a step-size. Consider T_h a numerical integrator such that T_h is symplectic and s_{-T_h} is involutive (T_h is said to be reversible).

Realistic setting: we choose $F = s_{-T_h}$. Let $z^0 = F(z)$.

F is symplectic, $q_z(z^0) = 1$ and $q_{z^0}(z) = (F^{-1})_{(z)}(z) = 1$.

However, $(z^0) \neq (z)$ due to ODE integration error.

In this case, the acceptance ratio simplifies as

$$a(z \rightarrow z^0) = \min \left(1, \frac{\exp(-H(z^0))}{\exp(-H(z))} \right)$$

Description of RMHMC (Girolami and Calderhead, 2011)

Algorithm 1: **RMHMC** (Girolami and Calderhead, 2011)

HMC Input: $(x_0; p_0) \in \mathbb{T}^2 M$, $\beta \in (0; 1]$, $N \in \mathbb{N}$

ODE Input: $h > 0$, $K \in \mathbb{N}$, numerical integrator $T_h: \mathbb{T}^2 M \rightarrow \mathbb{T}^2 M$

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

```

1 for  $n = 1; \dots; N$  do
2   Step 1: momentum sampling with refresh
3    $p \sim N(0; g(x_{n-1}))$ ;  $p_n = \beta^{-1} p_{n-1} + \beta p$ 
4   Step 2: performing  $K$  steps of discretized ODE (1) with  $T_h$ 
5    $(x^0; p^0) = (T_h)^K(x_{n-1}; p_{n-1}) = (x^0; p^0)_{\kappa h(x_{n-1}; p_{n-1})}$ 
6   Step 3: applying the Metropolis-Hastings (MH) acceptance filter
7    $a = \min(1; \exp[-H(x^0; p^0) + H(x_{n-1}; p_{n-1})])$ ;  $u \sim U[0; 1]$ 
8   if  $u < a$  then  $x_n; p_n = x^0; p^0$ ;
9   else  $x_n; p_n = x_{n-1}; p_{n-1}$ ;
10  Step 4: flipping the sign of the momentum
11   $x_n; p_n = x_n; -p_n$  (guarantees reversibility and exploration)

```

Results on RMHMC

Originally, [Girolami and Calderhead \(2011\)](#) considered posterior distributions from Bayesian models and chose:

g as the Fisher-Rao metric,

T_h as the Leapfrog integrator ([Hairer et al., 2006](#)) with fixed-point steps.

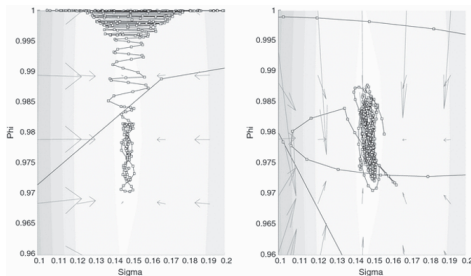


Figure: Figure 4 in [Girolami and Calderhead \(2011\)](#): HMC (left) vs RMHMC (right).

Results on RMHMC

Originally, [Girolami and Calderhead \(2011\)](#) considered posterior distributions from Bayesian models and chose:

g as the Fisher-Rao metric,

T_h as the Leapfrog integrator ([Hairer et al., 2006](#)) with fixed-point steps.

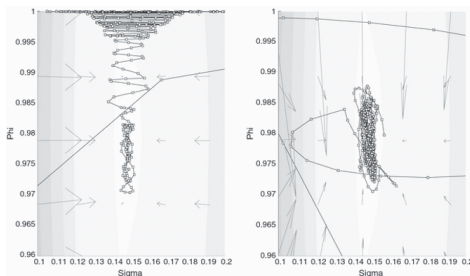


Figure: Figure 4 in [Girolami and Calderhead \(2011\)](#): HMC (left) vs RMHMC (right).

Q1: Can we enlarge the design of g to geometric constraints ?

Q2: Can we derive theoretical results from the properties of M , g and T_h ?

Introducing self-concordance

If M is **convex**, one can design a *self-concordant barrier* on M (Nesterov and Nemirovskii, 1994) and set $g = D^2$, 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 $\psi : U \rightarrow \mathbb{R}$ is said to be a ν -self-concordant (s.-c.) barrier (with $\nu \geq 1$) on U if it satisfies:

- (a) $\psi \in C^3(U; \mathbb{R})$ and ψ is **convex**,
- (b) $\psi(x) \rightarrow +\infty$ as $x \rightarrow \partial U$,
- (c) Other technical conditions on $D^3 \psi; D^2 \psi; D \psi$:

$D^2 \psi$ is 2-Lipschitz and $D \psi$ is ν -Lipschitz :

Introducing self-concordance

If M is **convex**, one can design a *self-concordant barrier* on M (Nesterov and Nemirovskii, 1994) and set $g = D^2$, 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 $g(x) = D^2(x)$.

⊙ Balls for $k_{g(x)}$ (**Dikin ellipsoids**) are central for the study of s.-c.

Self-concordance on a polytope

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

The logarithmic barrier on M is given by

$$\phi(x) = -\sum_{i=1}^m \ln(b_i - A_i^\top x) ;$$

and verifies (Nesterov and Nemirovski, 1998):

is a m -self-concordant barrier.

Self-concordance on a polytope

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

The logarithmic barrier on M is given by

$$g(x) = \sum_{i=1}^m \ln b_i - A_i^T x ;$$

and verifies (Nesterov and Nemirovski, 1998):

is a m -self-concordant barrier.

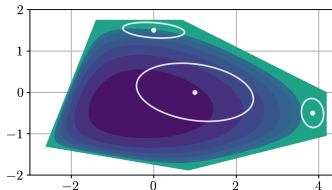


Figure: A self-concordant (logarithmic) barrier for a polytope $M \subset \mathbb{R}^2$ with three Dikin ellipsoids $\{y \in \mathbb{R}^d : y^T g(x) y < 1\}$ centered at $x = (0; 0); (0; 1.5); (1.5; 1)$.

Self-concordant barriers provide theoretical guarantees for polytope sampling

Dikin Walk ([Kannan and Narayanan, 2009](#)) no experiment

Geodesic Walk ([Lee and Vempala, 2017a](#)) no experiment

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

¹<https://github.com/ConstrainedSampler/PolytopeSamplerMatlab>

Self-concordant barriers provide theoretical guarantees for polytope sampling

Dikin Walk (Kannan and Narayanan, 2009) no experiment

Geodesic Walk (Lee and Vempala, 2017a) no experiment

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

RHMC (Lee and Vempala, 2018) 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.

Self-concordant barriers provide theoretical guarantees for polytope sampling

Dikin Walk (Kannan and Narayanan, 2009) no experiment

Geodesic Walk (Lee and Vempala, 2017a) no experiment

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

RHMC (Lee and Vempala, 2018) 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 h , that is proved to be symplectic and reversible
- ^ Assume that h admits a unique solution for any initial point
=) this guarantees reversibility as explained before.
- ^ However, **this assumption is not verified in practice.**
- ^ Their experiments¹ highlight asymptotic bias

¹ <https://github.com/ConstrainedSampler/PolytopeSamplerMatlab>

Self-concordant barriers provide theoretical guarantees for polytope sampling

Dikin Walk (Kannan and Narayanan, 2009) no experiment

Geodesic Walk (Lee and Vempala, 2017a) no experiment

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

RHMC (Lee and Vempala, 2018) 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 h , that is proved to be symplectic and reversible
- ^ Assume that h admits a unique solution for any initial point
=) this guarantees reversibility as explained before.
- ^ However, **this assumption is not verified in practice.**
- ^ Their experiments¹ highlight asymptotic bias

Can we have better practical and theoretical guarantees ?

¹<https://github.com/ConstrainedSampler/PolytopeSamplerMatlab>

Self-concordant barriers provide theoretical guarantees for polytope sampling

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

RHMC (Lee and Vempala, 2018) 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 h , that is proved to be symplectic and reversible
- ^ Assume that T_h admits a unique solution for any initial point
=> this guarantees reversibility as explained before.
- ^ However, **this assumption is not verified in practice.**
- ^ Their experiments² 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 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 ? ! [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 $Q : T^?M \times B(T^?M) \rightarrow [0;1]$ be a transition probability kernel and let μ be a probability distribution on $T^?M$. Then, Q is said to be **reversible up to momentum reversal** with respect to μ if for any $f \in C(T^?M \times T^?M; \mathbb{R})$ with compact support

$$\int_{T^?M} \int_{T^?M} f(z; z^0) Q(z; dz^0) \mu(z) = \int_{T^?M} \int_{T^?M} f(s(z^0); s(z)) Q(z; dz^0) \mu(z) :$$

Notation and reminders

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

Pushforward: $\mu_{\#} \in \mathcal{P}(Y)$ is the pushforward of $\mu \in \mathcal{P}(X)$ by $\gamma : X \rightarrow Y$.

Lemma 3 (Preservation of measure.)

Let $Q : T^2M \rightarrow B(T^2M) \times [0;1]$ be a transition probability kernel and let μ be a probability distribution on T^2M . Assume that $\mu_{\#} = \mu$ and that Q is *reversible up to momentum reversal* with respect to μ . Then μ is an invariant measure for Q .

- 1 Motivations and background
- 2 **Description of BHMC**
 - Hamiltonian integrators of BHMC
 - Introducing BHMC
- 3 Results
- 4 Conclusion

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

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

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

$$H_2(x; p) = \frac{1}{2} p^T g(x)^{-1} p ; \quad (\text{non separable})$$

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

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

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

$$H_2(x; p) = \frac{1}{2} p^T g(x)^{-1} p ; \quad (\text{non separable})$$

\hat{O} Explicit integrator of H_1

$S_{h=2} : T^*M \rightarrow T^*M$ is the map defined by $S_{h=2}(x; p) = (x; p - \frac{h}{2} \nabla_x H_1(x; p))$

=) $S_{h=2}$ is symplectic and reversible

$S_{h=2}$ approximates the dynamics of H_1 on a step-size $h=2$.

s $S_{h=2}$ is an involution.

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

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

$$H_2(\mathbf{x}; \mathbf{p}) = \frac{1}{2} \mathbf{p} \mathbf{k}_{g(\mathbf{x})}^{-1} \mathbf{p} \quad : \quad (\text{non separable})$$

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

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

$$H_2(x; p) = \frac{1}{2} k p^2 g(x)^{-1} : \quad (\text{non separable})$$

\hat{O} **Implicit** integrator of H_2

G_h is the **Leapfrog integrator** of H_2 with step-size: for any $z^{(0)} \in T^2M$, $G_h(z^{(0)}) \in T^2M$ consists of points $z^{(1)} = (x^{(1)}; p^{(1)})$ that solve

$$p^{(1=2)} = p^{(0)} - \frac{h}{2} \nabla_x H_2(x^{(0)}; p^{(1=2)}) ;$$

$$x^{(1)} = x^{(0)} + \frac{h}{2} [\nabla_x H_2(x^{(0)}; p^{(1=2)}) + \nabla_x H_2(x^{(1)}; p^{(1=2)})] ;$$

$$p^{(1)} = p^{(1=2)} - \frac{h}{2} \nabla_x H_2(x^{(1)}; p^{(1=2)}) ;$$

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

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

$$H_2(x; p) = \frac{1}{2} k p^2 g(x)^{-1} : \quad (\text{non separable})$$

\hat{O} **Implicit** integrator of H_2

G_h is the **Leapfrog integrator** of H_2 with step-size: for any $z^{(0)} \in T^2M$, $G_h(z^{(0)}) \subset T^2M$ consists of points $z^{(1)} = (x^{(1)}; p^{(1)})$ that solve

$$\begin{aligned} p^{(1=2)} &= p^{(0)} - \frac{h}{2} \nabla_x H_2(x^{(0)}; p^{(1=2)}) ; \\ x^{(1)} &= x^{(0)} + \frac{h}{2} [\nabla_x H_2(x^{(0)}; p^{(1=2)}) + \nabla_x H_2(x^{(1)}; p^{(1=2)})] ; \\ p^{(1)} &= p^{(1=2)} - \frac{h}{2} \nabla_x H_2(x^{(1)}; p^{(1=2)}) : \end{aligned}$$

=) G_h is symplectic and reversible ([Hairer et al., 2006](#))

=) above, there may be either **0, 1, 2; : solutions !**

$F_h = G_h^{-1} \circ s$ is a set-valued map.

F_h approximates the dynamics of H_2 on a step-size h .

If $j(F_h(z)) > 0$ then $z \in (F_h^{-1} \circ F_h)(z)$ (almost an involution).

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

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

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

$$H_2(x; p) = \frac{1}{2} p^T g(x)^{-1} p ; \quad (\text{non separable})$$

⌚ Explicit integrator of H_1

Map $S_{h=2} : T^*M \rightarrow T^*M$ defined by $S_{h=2}(x; p) = (x; p - \frac{h}{2} \nabla_x H_1(x; p))$.

⌚ Implicit integrator of H_2

Set-valued map $F_h = G_h^{-1} \circ S$, where $G_h : T^*M \rightarrow 2^{T^*M}$ is the **Leapfrog integrator** of H_2 with step-size h (**none, one or multiple solutions**).

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

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

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

$$H_2(x; p) = \frac{1}{2} p^T g(x)^{-1} p ; \quad (\text{non separable})$$

Explicit integrator of H_1

Map $S_{h=2} : T^2M \rightarrow T^2M$ defined by $S_{h=2}(x; p) = (x; p - \frac{h}{2} \nabla_x H_1(x; p))$.

Implicit integrator of H_2

Set-valued map $F_h = G_h^{-1} \circ s$, where $G_h : T^2M \rightarrow \mathbb{R}^{2^T M}$ is the Leapfrog integrator of H_2 with step-size h (none, one or multiple solutions).

Implicit integrator of H

Set-valued map $R_h = (s \circ S_{h=2}) \circ F_h \circ (s \circ S_{h=2})$.

By composition, R_h is symplectic and reversible

$s \circ R_h$ approximates the dynamics of H on a step-size h .

In practice, we do not have access to Φ_h but approximate it with a numerical map $\tilde{\Phi}_h$, defined on a domain $\text{dom } \tilde{\Phi}_h \subset T^*M$ with $\tilde{\Phi}_h(\text{dom } \tilde{\Phi}_h) \subset T^*M$.

We also define the numerical map $R_h : (s, S_{h=2})(\text{dom } \tilde{\Phi}_h) \rightarrow T^*M \rightarrow T^*M$

$$R_h = (s, S_{h=2}) \circ \tilde{\Phi}_h \circ (s, S_{h=2}) :$$

Similarly to R_h , $s \circ R_h$ approximates the dynamics of H on a step-size h .

How to implement $\tilde{\Phi}_h$?

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 \tilde{h} , defined on a domain $\text{dom } \tilde{h} \subset T^*M$ with $\tilde{h}(\text{dom } \tilde{h}) \subset T^*M$.

We also define the numerical map $R_h : (s, S_{h=2})(\text{dom } \tilde{h}) \rightarrow T^*M \rightarrow T^*M$

$$R_h = (s, S_{h=2}) \circ \tilde{h} \circ (s, S_{h=2}) :$$

Similarly to R_h , $s \circ R_h$ approximates the dynamics of H on a step-size h .

How to implement \tilde{h} ?

Fixed-point solver (ours, [Kook et al. \(2022\)](#)).

Newton's solver ([Brofos and Lederman, 2021a,b](#)).

For any $z \in T^*M$, we define on T^*M the norm $\|k\|_z$ by

$$\|k\|_z = \|kx\|_{g(x)} + \|kp\|_{g(x)^{-1}} ; z = (x^0; p^0) :$$

On x^0 : the Dikin norm.

On p^0 : the "natural" norm induced by $g(x)^{-1}$.

BHMC: CRHMC with “involution checking step”

Algorithm 2: Barrier HMC (BHMC)

HMC Input: $(x_0; p_0) \in T^2M$, $\tau \in (0; 1]$, $N \in \mathbb{N}$

ODE Input: $h > 0$, $\beta > 0$, numerical integrator h with domain $\text{dom } h$

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

```

1 for  $n = 1; \dots; N$  do
2   Step 1:  $\beta \sim N(0; g(x_{n-1}))$ ;  $p_{n-1} \leftarrow p_{n-1} + \beta$ 
3   Step 2: solving discretized ODE (1) with  $h$ 
4    $x^0; p^0 \leftarrow x_{n-1}; p_{n-1}$ ;  $x^{(0)}; p^{(0)} \leftarrow (S_{h=2})(x_{n-1}; p_{n-1})$ 
5   if  $z^{(0)} \in \text{dom } h$  then
6      $z^{(1)} = h(z^{(0)})$ ,  $\text{err} = k z^{(0)} \cdot h(z^{(1)})_{z^{(0)}} + k z^{(0)} \cdot h(z^{(1)})_{z^{(1)}}$ 
7     if  $z^{(1)} \in \text{dom } h$  &  $\text{err}$  then  $x^0; p^0 \leftarrow (S_{h=2})(x^{(1)}; p^{(1)})$ ;
8   Step 3:  $a = \min(1; \exp[-H(x^0; p^0) + H(x_{n-1}; p_{n-1})])$ ;  $u \sim U[0; 1]$ 
9   if  $u < a$  then  $x_n; p_n \leftarrow x^0; p^0$ ;
10  else  $x_n; p_n \leftarrow x_{n-1}; p_{n-1}$ ;
11  Step 4:  $x_n; p_n \leftarrow x_n; p_n$  (guarantees reversibility and exploration)
    
```

⊙ Checking that h is well defined on the iterates.

⊙ New: checking that $(h \circ h)(z^{(0)}) = z^{(0)}$! involution checking !

Outline

- 1 Motivations and background
- 2 Description of BHMC
- 3 Results**
 - Assumptions
 - Results on integrators
 - Results on reversibility
 - Numerical experiments
- 4 Conclusion

Assumptions

We aim at sampling from a target distribution supported on M

$$d(x) = dx / \exp[-V(x)] \quad ; \quad V \in C^2(M; \mathbb{R}) :$$

A1 (Assumption on M .)

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

A2 (Assumption on g .)

There exists , *-s.-c. barrier* on M such that $g = D^2$.

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 A1, A2. Let $z^{(0)} \in T^*M$, then there exists $h^? > 0$ (explicit) such that for any $h \in (0; h^?)$, there exist $z_h^{(1)} \in F_h(z^{(0)})$, a neighborhood $U \subset T^*M$ of $z^{(0)}$ and a C^1 -diffeomorphism $\eta_h : U \rightarrow \eta_h(U) \subset T^*M$ with

- (a) $\eta_h(z^{(0)}) = z_h^{(1)}$ and $j \det \text{Jac}(\eta_h) = 1$.
- (b) $\eta_h(z)$ is the only element of $F_h(z)$ in $\eta_h(U)$ for $z \in U$.

Following Proposition 1, we derive a technical assumption on the corresponding numerical integrator η_h , denoted by A3: basically, we assume that η_h is locally involutive.

$Q_0 : T^?M \rightarrow B(T^?M) \times [0; 1]$: the transition kernel for Step 1.

! Q_0 is reversible up to m.-r. w.r.t.

$Q_1 : T^?M \rightarrow B(T^?M) \times [0; 1]$, the transition kernel for Steps 2 to 4.

! Q_1 is reversible up to m.-r. w.r.t. under A1, A2 and A3.

$Q : T^?M \rightarrow B(T^?M) \times [0; 1]$, the transition kernel for Steps 1 to 4 such that

$$Q(z; dz^0) = \int_{T^?M} \int_{T^?M} Q_0(z; dz_1) Q_1(z_1; dz^0) :$$

Assume A1, A2, A3. Then, Q is reversible up to momentum reversal
In particular, μ is an invariant measure for Q .

Parameters

$$\hat{O} = M = f \times 2 \mathbb{R}^d : \prod_{i=1}^d x_i = 1; x_i \in [0, 1]; \text{dgg with } d \in \{5, 10\}$$

$$\hat{O} = 10 \text{ if } d = 5 \text{ and } = 200 \text{ if } d = 10$$

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.

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

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

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

BHMC approximates the dynamics with a numerical integrator.

Ô We proved that M is invariant for BHMC.

Ô 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

- ! Investigate the coupled behaviour of the hyperparameters h and ϵ .
- ! Study the irreducibility of n-BHMC (not easy task).

Unbiased constrained sampling with Self-Concordant Barrier Hamiltonian Monte Carlo

Maxence Noblé¹ Valentin de Bortoli² Alain Durmus¹

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

²Département 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 de ned manifolds. In Arti cial 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 a ne interior point method for linear programming. In Proceedings of the forty- rst 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.