Geometric Convergence of HMC on Complete Riemannian Manifolds

Xinhao Yu and A. D. Kennedy

Higgs Centre
School *of* Physics *and* Astronomy
University of Edinburgh

Liverpool, Lattice2024

Daily work: sampling from $\tilde{\mu}$ on $\mathcal{M}$ using algorithm $\mathcal{P}$.

Our aim: conditions on $\mathcal{P} \Longrightarrow$ converge exponentially towards $\tilde{\mu}$.

Assuming sampling space $\mathcal{M}$ is a complete Riemannian manifold, we prove

- If $\mathcal{M}$ is compact, HMC converges.
- If $\mathcal{M}$ is non-compact, after mixed with a radial Metropolis(provided there is radial direction), HMC converges.

Outline

Convergence of Markov chains

What do we mean by convergence?

Banach fixed-point theorem

If (X, d) is a complete metric space and the transition $\mathcal{P} : X \to X$ is a contraction mapping,

$$d(\mathcal{P}\mu, \mathcal{P}\nu) \leq a \cdot d(\mu, \nu)$$

for some $a \in [0, 1)$ and $\forall \mu, \nu \in X$, then there is a unique fixed-point $\tilde{\mu}$ such that $\lim_{n\to\infty} \mathcal{P}^n\mu = \tilde{\mu} \quad \forall \mu \in X$.

So one has to prepare

- a metric on the space of probability measures.
- show $\mathcal{P}$ is a contraction on it.

Compact spaces: Doeblin's condition

A sufficient condition for convergence on compact spaces is

Doeblin's condition

If $\exists \alpha \in (0,1)$ and a probability measure ν such that

$$\mathcal{P}(x,\cdot) \geq \alpha\nu(\cdot) \qquad \forall x \in \mathcal{M}$$

then the Markov chain converges geometrically.

- Total Variation (TV) metric:

$$d(\mu,\nu) \equiv \|\mu - \nu\|_{\mathsf{TV}} = \sup_{A \in \mathcal{B}(\mathcal{M})} |\mu(A) - \nu(A)| \, .$$

- It may be shown that $\forall \mu, \nu \in X$:

$$\|\mathcal{P}\mu - \mathcal{P}\nu\|_{\mathsf{TV}} \leq (1-\alpha) \cdot \|\mu - \nu\|_{\mathsf{TV}}$$

Harris' Theorem

Not able to construct Doeblin's condition on the whole non-compact $\mathcal{M}$ since

$$1 = \int_{\mathcal{M}} \mathcal{P}(x, \mathcal{M}) \cdot \mu_V(dx) \geq \alpha\nu(\mathcal{M}) \cdot \mu_V(\mathcal{M})$$

and the volume $\mu_V(\mathcal{M})$ can be infinite.

Harris' theorem

Hairer and Mattingly gave an elegant simplification of Harris theorem[1].

Geometric Drift Condition (GDC)

There is a Lyapunov function $L : \mathcal{M} \to [0, \infty)$, $\gamma \in (0, 1)$, and $K \geq 0$ such that $\forall x \in \mathcal{M}$ we have

$$(\mathcal{P}L)(x) \leq \gamma \cdot L(x) + K.$$

Doeblin's Condition (DC)

$\exists \alpha \in (0, 1)$, a probability measure ν, and a *small set* $\mathcal{C} = \{x \in \mathcal{M} : L(x) \leq R\}$ where $R > \frac{2K}{1 - \gamma}$, such that

$$\inf_{x \in \mathcal{C}} \mathcal{P}(x, \cdot) \geq \alpha \nu(\cdot).$$

- if $\gamma \in [0, 1]$, we say weak GDC.

- use a more simple condition: $\inf_{x,y \in \mathcal{C}} P(x \to y) \geq \alpha$.

[1] Martin Hairer and Jonathan C. Mattingly (2011). "Yet Another Look at Harris' Ergodic Theorem for Markov Chains". In: *Seminar on Stochastic Analysis, Random Fields and Applications VI*. ed. by Robert Dalang, Marco Dozzi, and Francesco Russo. Basel: Springer, pp. 109–117. ISBN: 978-3-0348-0021-1.

Outline

Introduction

HMC assigns each degree of freedom a fictitious momentum:

$$q \in \mathcal{M} \longrightarrow (q, p) \in T^*\mathcal{M}$$

The cotangent bundle $T^*\mathcal{M}$ admits a symplectic structure because

- $\exists$ Liouville one-form: $\exists \theta \in \Lambda^1(T^*\mathcal{M})$ with $\beta^*\theta = \beta, \ \forall \beta \in \Lambda^1(\mathcal{M})$.
- $\exists$ a closed, non-singular fundamental two-form ω:

$$\omega \equiv -d\theta, \quad d\omega = 0, \quad \det \omega \neq 0.$$

Then for every smooth function F, there is a unique Hamiltonian vector field $\hat{F}$ such that

$$\omega(\hat{F}, \cdot) = -dF.$$

Trajectories, denoted as $\sigma_{\hat{F}}$, are local flows tangential to $\hat{F}$.

Doeblin's condition for probability density

Probability Densities

A probability density is usually just the Radon–Nikodym derivative of the transition probability, but sometimes this needs to be extended to a distribution (generalized function).

Measure of $T^*\mathcal{M}$ is the volume form of the phase space:

$$\mathrm{Vol} \equiv \omega^n.$$

Define probability densities:

$$P(x \to y) \equiv \frac{d\mathcal{P}(x,\cdot)}{d\mathrm{Vol}}(y), \quad Q(y) \equiv \frac{d\nu}{d\mathrm{Vol}}(y).$$

Then $\forall x, y \in \mathcal{C}$:

$$P(x \to y) \geq c > 0 \underset{\not\Longleftarrow}{\Longrightarrow} P(x \to y) \geq \alpha Q(y) \Leftrightarrow \mathcal{P}(x,\cdot) \geq \alpha\nu.$$

Doeblin's condition for probability density

It is convenient to use on $T^*\mathcal{M}$ and $\mathcal{M}$ that are invariant under transition:

- symplectomorphisms (canonical transformations) preserve the volume form:

$$\mathcal{L}_{\hat{V},\hat{T}} \mathrm{Vol} = 0.$$

- $\hat{T}$ is an isometry: Riemannian measure μ_g is preserved.

As a result, the extended target distribution

$$\int_{\mathrm{Vol}} e^{-H} = \int_{\mathrm{Vol}} e^{-(V+T)}$$

has well-defined state density e^{-H}, and after integration over momentum:

$$\int_{\mathrm{Vol}} e^{-(T+V)} \Rightarrow \int_{\mu_g} e^{-V}$$

also has well-defined e^{-V}.

Algorithm

The algorithm we use:

1. (Partial) Momentum Refreshment

$$S_{\mathrm{MR}} : (q,p) \mapsto (q',p') = p\cdot\cos\theta + \eta\cdot\sin\theta, \quad \eta\sim\mu_G.$$

2. Molecular Dynamics Monte Carlo (S_{MDMC}) Which is made up of
   - A Hamiltonian trajectory

$$S_{\mathrm{MD}} : (q,p) \mapsto (q',p') = \sigma(t).$$

   - A Metropolis accept/reject test S_{MC}.
   - A momentum flip if rejected:

$$S_{\mathrm{Flip}} : (q,p) \mapsto (q,-p).$$

Algorithm

We use the Leapfrog (Verlet, Störmer) integrator S_{LF} to approximate Hamiltonian dynamics S_{MD}; given a step size τ, it is

$$S_{\mathrm{MD}} = S_{\mathrm{LF}} \equiv \sigma_{\hat{V}}\left(\frac{\tau}{2}\right) \circ \sigma_{\hat{T}}(\tau) \circ \sigma_{\hat{V}}\left(\frac{\tau}{2}\right).$$

Kinetic energy T is naturally defined by the inverse Riemannian metric

$$T(q,p) \equiv \frac{1}{2} g_q^{-1}(p,p).$$

Thus the Gibbs sampler of S_{MR} is the distribution

$$\mu_G(A) \propto \int_A e^{-T(q,\eta)}\, d\eta.$$

Leapfrog on the cotangent bundle

Levi-Civita connection ∇:

$$\nabla g = 0.$$
$$\nabla_X Y - \nabla_Y X - [X, Y] = 0, \quad \forall X, Y \in \mathfrak{X}(\mathcal{M}).$$

∇ is an Ehresmann connection:

$$T_x T^*\mathcal{M} = \mathcal{V}_x \oplus \mathcal{H}_x.$$

Denote $\sharp$ the musical isomorphism of g, we have:

- $\hat{V}$ is vertical:

$$\hat{V} = (0, -dV) \qquad \sigma_{\hat{V}}(t) : (q, p) \mapsto (q, p - t \cdot dV_q).$$

- $\hat{T}$ is horizontal:

$$\hat{T} = (p^\sharp, 0) \quad \sigma_{\hat{T}}(t) : (q, p) \mapsto \left(\exp_q(t \cdot p^\sharp), p'\right), \quad T(x) = T(x').$$

Doeblin's condition for HMC

So, An update is:

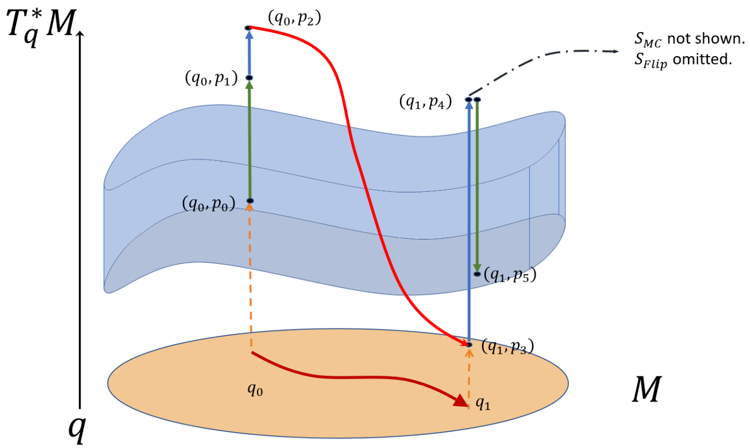
$$(q_0, p_0) \xrightarrow{S_{\mathrm{MR}}} (q_0, p_1) \xrightarrow{\sigma_{\hat{V}}} (q_0, p_2) \xrightarrow{\sigma_{\hat{T}}} (q_1, p_3) \xrightarrow{\sigma_{\hat{V}}} (q_1, p_4) \xrightarrow{S_{\mathrm{MR}}} (q_1, p_5)$$

The step $S_{\mathrm{MC}} \circ S_{\mathrm{Flip}}$ after (q_1, p_4) is not shown explicitly.

small set

The Lyapunov function is chosen to be Hamiltonian: $L \equiv H$.
The small set is:

$$\mathcal{C} = \{x \in T^*\mathcal{M} \big| H(x) \leq V_R\}$$


T_q^*M
(q_0, p_2)
(q_0, p_1)
(q_0, p_0)
(q_1, p_4)
(q_1, p_5)
(q_1, p_3)
S_{MC} not shown.
S_{Flip} omitted.
q_0
q_1
M
q

1. We have
   - $|V(q_1) - V(q_0)| \leq V_R \Rightarrow |q_1 - q_0| \leq R$.
   - $|dV_{q_0}| \leq M_1$
2. $\sigma_{\hat{T}}$ exists:
   By Hopf-Rinow theorem, there is a geodesic connecting any two points on $\mathcal{M}$, $\sigma_{\hat{T}}$ is the unique horizontal lift of it.
3. $T(q_0, p_2) = T(q_1, p_3)$ is bounded, since

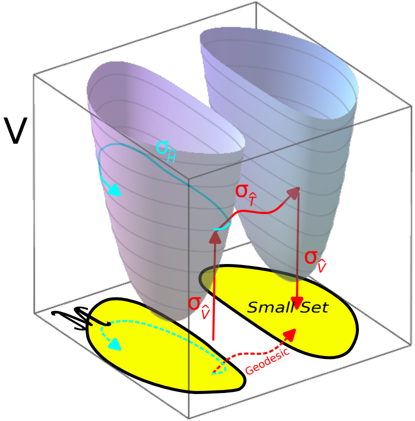
$$T = d(q_0, q_1)^2 / 2\tau^2$$

4. $p_2 = p_0 \cos\theta + \eta \sin\theta - \frac{\tau}{2} dV_{q_0} \Rightarrow T(\eta)$ and $e^{-T(\eta)}$ bounded.
5. As a result, the probability density is bounded by

$$P(x \to y) \geq c = \exp\left\{ \frac{-2}{\sin^2\theta} \left[\frac{R^2}{2\tau^2} + \cos^2\theta \, V_R + \frac{\tau^2}{4} M_1^2 \right] \right\} \cdot e^{-2V_R},$$

   the multiplication of probability densities from two S_{MR} and a Metropolis test.

Leapfrog integrator



Reasons for using a single Leapfrog step:

- When $\mathcal{C}$ disconnected, $\sigma_{\hat{T}}$ can join state between subsets.
  This will not work if use exact integrator $\sigma_{\hat{H}}$.
- A trajectory consists of a random number of Leapfrog steps, and there is a positive probability of taking a single step.
- This depends on ability of the algorithm to cross a potential barrier, HMC was not designed to deal with such barriers.

Geometric drift condition: Compact $\mathcal{M}$

- No need for Harris' theorem if total momentum refreshment.
- However in general we must consider the non-compact phase space $T^*\mathcal{M}$ as the state space; e.g., when partial momentum refreshment is used.

Lyapunov function L

Choose the Lyapunov function to be the Hamiltonian, $L = H$. In general it is always minus the logarithm of the target distribution.

The strategy of our proof is

1. Momentum Refreshment satisfies the strong GDC.
2. Molecular Dynamics satisfies weak GDC.
3. Thus combining them HMC satisfies strong GDC on compact $\mathcal{M}$.

Generalized Drift Condition for compact $\mathcal{M}$

$\mathcal{M}$ is compact and V is smooth, so must be bounded $V \leq V_{\text{max}}$.

1. Partial momentum refreshment ($\mathcal{P}_{\text{MR}}$) satisfies the strong GDC

$$\begin{aligned}
(\mathcal{P}_{\text{MR}} H)(q,p) &= \langle H(q,p) \rangle_\eta \\
&\propto \int_\Omega \left[T(q, S_{\text{MR}}(p)) + V(q) \right] e^{-T(\eta)} \, d\eta \\
&= V(q) + (\cos\theta)^2 T(q,p) + (\sin\theta)^2 \\
&= (\cos\theta)^2 H(q,p) + (\sin\theta)^2 (1 + V(q)) \\
&\leq (\cos\theta)^2 H(q,p) + (\sin\theta)^2 (1 + V_{\text{max}}).
\end{aligned}$$

This also works for pseudofermions, since they are generated using a Gibbs sampler (heatbath) from a distribution with exponentially small tails.

2. $\mathcal{P}_{\text{MD}}$ satisfies weak GDC.

Weak Generalized Drift Condition for the Metropolis Algorithm

In general any Metropolis algorithm satisfies the weak GDC with minus log probability as the Lyapunov function.

Let $\tilde{x} = (\tilde{q}, \tilde{p}) = S_{\text{MD}}(x)$, then the acceptance rate is

$$\mathcal{A}(x, \tilde{x}) = \min\left(1, e^{-H(\tilde{x})+H(x)}\right) = \min\left(1, e^{-\delta H}\right).$$

Thus we have

$$\begin{aligned}(\mathcal{P}_{\text{MD}} H)(x) &= \mathcal{A} \cdot H(\tilde{x}) + (1 - \mathcal{A}) \cdot H(x) \\ &= H(x) + \mathcal{A} \cdot \delta H.\end{aligned}$$

The term $\mathcal{A} \cdot \delta H$ is bounded from above since

- If $\delta H \leq 0$ then $\mathcal{A} \cdot \delta H = \delta H \leq 0$.
- If $\delta H > 0$ then $\mathcal{A} \cdot \delta H = e^{-\delta H} \delta H \leq 1/e$, the maximum value being attained at $\delta H = 1$.

Generalized Drift Condition for compact $\mathcal{M}$

3. The combination of steps satisfy the strong GDC.
   If a bounded number of transitions $\{\mathcal{P}_i\}$ with $i = 1, \ldots, n$ all satisfy the weak GDC, and furthermore one of them $\mathcal{P}_k$ satisfies the strong GDC, then the composite transition satisfies the strong GDC whose parameters are

$$\gamma = \gamma_k, \qquad K = \sum_i K_i.$$

   Thus HMC on a compact Riemannian manifold satisfies the strong GDC.

Drift Condition on non-compact $\mathcal{M}$

What goes wrong in the non-compact case?

- Doeblin's Condition is fine.
- previous results, such as weak GDC for Metropolis still hold.
- V is no longer bounded thus S_{MR} merely satisfies the weak GDC

$$
\begin{aligned}
(\mathcal{P}_{\mathrm{MR}} H)(q,p) &= V(q) + (\cos\theta)^2 T(p) + (\sin\theta)^2 \\
&= H(q,p) + \sin\theta^2 (1 - T(p)),
\end{aligned}
$$

We fix this by introducing a new Markov step that has the desired fixed-point distribution and satisfies the strong GDC by construction. It is a Metropolis algorithm in the radial direction.

Radial Metropolis Algorithm

The algorithm is on the base manifold $\mathcal{M}$. With a reference point q_0 as the origin, the radius is defined as the distance $r_q \equiv d(q, q_0)$, the complementary angular direction is denoted by $\theta \in \Omega_\theta$, so we have the parameterization $q = (r, \theta)$.

Define a forward step $f : r \rightarrow f(r) = R_f$ and the corresponding backward step $b : r \rightarrow g(r) = R_b$ such that $b = f^{-1}$. Then, the algorithm works as follows

1. $r \rightarrow R_f$ or $r \rightarrow R_b$ with equal probability $= \frac{1}{2}$.
2. Apply a Metropolis test with the acceptance rate $\mathcal{A} = \min(1, e^{-\delta V})$.

GDC for Radial Metropolis

At angle θ_0. Denote the acceptance rate $\mathcal{A}_x(r \to R_x) = \min\left(1, e^{-V(R_x)+V(r)}J(R_x, r)\right)$, transition on V is:

$$(\mathcal{P}_r V)(r) = \frac{1}{2}\sum_{x\in\{f,b\}} \{V(R_x)\mathcal{A}_x + V(r)\,(1-\mathcal{A}_x)\}$$

Three conditions sufficient for Radial Metropolis to satisfy strong GDC:

1. $\exists \tilde{R}$ such that $e^{-V(R_f)+V(r)}J(R_f, r) \leq 1 \leq e^{-V(R_f)+V(r)}J(R_b, r)$ for all $r > \tilde{R}$.
2. Backward step shrinks V: $\exists \rho \in (0,1)$ such that $\forall r \geq \tilde{R}$ one has $V(R_b) \leq \rho V(r) + N$.
3. $\exists M \geq 0$ such that $J(R_f, r)\delta V \cdot e^{-\delta V} \leq M$.

GDC for Radial Metropolis

Under these requirements

$$
\begin{aligned}
(\mathcal{P}_r V)(r) &= \frac{1}{2} \sum_{x \in \{f,b\}} \{V(R_x)\mathcal{A}_x + V(r)\,(1 - \mathcal{A}_x)\} \\
&= \frac{1}{2} \left\{ J(R_f, r)\delta V e^{-\delta V} + V(r) + V(R_b) \right\} \\
&\leq \frac{1+\rho}{2} V(r) + M + N \\
&= \gamma V(r) + K.
\end{aligned}
$$

We now provide a choice of forward/backward steps that meet these requirements.

Radial Metropolis: polynomial potential

Case 1: $V(r) = kr^\alpha + o(r^\alpha)$.

- Forward step: $r \to R_f = (1+\epsilon)r$,
- Backward step: $r \to R_b = r/(1+\epsilon)$.

All three conditions met, with parameters:

1. $\tilde{R} \geq \left\{ \dfrac{\log(1+\epsilon)}{k(1-(1+\epsilon)^{-\alpha})} \right\}^{1/\alpha}$.
2. $\gamma = \dfrac{1}{(1+\epsilon)^\alpha}$.
3. $K = (1+\epsilon)/e$.

Example: ϕ^4

Consider the lattice action with volume N and n dimension of ϕ (thus $\mathcal{M} = \mathbb{R}^{N\times n}$):

$$S = \sum_{x,i} \frac{1}{2}|\nabla_\mu \phi_{x,i}|^2 + \frac{1}{2}m^2|\phi_{x,i}|^2 + \frac{\lambda}{4!}(|\phi_{x,i}|^2)^2.$$

Set a basis $(\theta, r) \in \mathbb{R}^{N\times n}$ such that $\phi_{x,i} = f_{x,i}(\theta)r$ and $\sum_{x,i}|f_{x,i}(\theta)|^2 = 1$, so:

$$S = k_1(\theta)r^4 + k_2(\theta)r^2.$$

Radial Metropolis: logarithmic potential

Case 2: $V(r) = \beta \log r + o(\log r)$.

$\beta \geq D - 1 \geq 0$ for normalization

- Forward step: $R_f = r(1 + \epsilon \cdot r)^\delta$,
- Backward step: $r = R_b(1 + \epsilon \cdot R_b)^\delta$.

All three conditions are met, with parameters:

1. $\tilde{R}_b \geq \frac{1}{\epsilon}\left\{\left((1+\delta)^{\frac{1}{\beta-1}} - 1\right)^{1/\delta}\right\}$
2. $\gamma = \frac{1}{\delta}$.
3. $K = \frac{\beta(1+\delta)}{(\beta-1)e}$.

Radial Metropolis: uniform GDC

What we have done so far: strong GDC in any direction θ_0 by radial Metropolis.

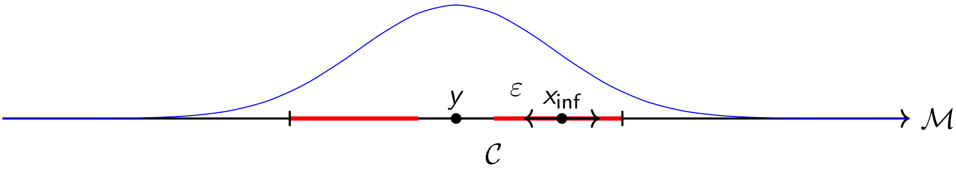
What we need: strong GDC of the whole space.

It requires combine the GDCs for all angluar directions and obtain one single GDC bound. Fortunately Ω_θ is compact, at least in finite dimensional spaces, so we have the following:

For a family of strong drift conditions along radial directions with $\gamma(\theta) \in (0,1)$ and $K(\theta) > 0$ are continuous functions of $\theta \in \Omega_\theta$ for a compact state subspace Ω, there exist a constant $\gamma \in [0,1)$ and a constant $K > 0$ such that they are maxima of the corresponding functions at some θ, hence the family of the strong GDCs yields uniformly a strong drift condition with γ and K:

$$(\mathcal{P}_r L)(r,\theta) \leq \gamma L(r,\theta) + K.$$

Thank you !



If the $\mathcal{C}$ chosen by GDC has $R > \varepsilon$, $\inf_{x,y} P(x \to y)$ vanishes, if $R \leq \varepsilon$ it is fine.