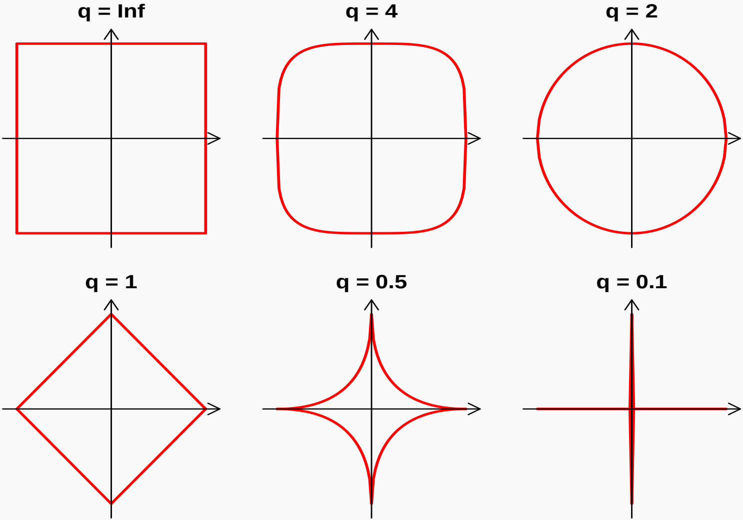
School of Mathematical and Statistical Sciences
Arizona State University

Robust & Efficient Hamiltonian Monte Carlo Algorithms on Manifolds

Shiwei Lan
slan@asu.edu

JUNE 05, 2020 Bayesian Computation @ SDSS, VIRTUAL

Motivation



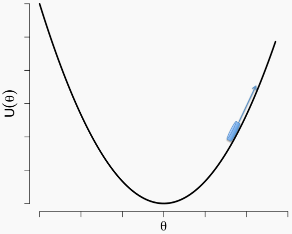
- Many probability distributions have natural constraints, e.g. $\|x\|_q \leq 1$.
  - Ridge regression: $q = 2$
  - Lasso, copula: $q = 1$
  - Constrained Gaussian Process: $q = \infty$
- Some probability distributions are defined directly on manifolds.
  - Latent Dirichlet Allocation: Simplex
  - Covariance Matrix: Space of Positive Definite Matrices
  - Eigen Vectors: Stiefel Manifolds
- Direct truncation may be doable but computationally wasteful.
- All these challenges call for efficient inference methods for Bayesian models in various settings.
- We consider a particular MCMC, Hamiltonian Monte Carlo, defined on several manifolds to resolve issues arising from different applications.

Table of contents

Hamiltonian Monte Carlo

HMC

Hamiltonian Dynamics
[8]



$$\dot{\boldsymbol{\theta}} = \frac{\partial H}{\partial \mathbf{p}}$$
$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \boldsymbol{\theta}}$$

- Position $\boldsymbol{\theta} \in \mathbb{R}^D \Longleftarrow$ variables of interest
- Momentum $\mathbf{p} \in \mathbb{R}^D \Longleftarrow$ fictitious, usually $\sim \mathcal{N}(\mathbf{0}, \mathbf{M})$
- Potential energy $U(\boldsymbol{\theta}) \Longleftarrow$ minus log of target density $\pi(\cdot)$
- Kinetic energy $K(\mathbf{p}) \Longleftarrow$ minus log of momentum density
- Hamiltonian $H(\boldsymbol{\theta}, \mathbf{p}) = U(\boldsymbol{\theta}) + K(\mathbf{p}) \Longleftarrow$ constant.

Bayesian Posterior Sampling

- We are interested in Posterior sampling $\pi(\boldsymbol{\theta}|\mathcal{D}) \propto \pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathcal{D})$.

$$U(\boldsymbol{\theta}) = -\log\pi(\boldsymbol{\theta}|\mathcal{D}) = -[\log\pi(\boldsymbol{\theta}) + \sum_{i=1}^{N}\log\pi(x_i|\boldsymbol{\theta})] \quad \boxed{+C}$$

- Sample $\mathbf{p} \sim \mathcal{N}(\mathbf{0}, \mathbf{M}^1)$, then set

$$K(\mathbf{p}) = -\log\pi(\mathbf{p}) = \frac{1}{2}p^{\mathsf{T}}\mathbf{M}^{-1}\mathbf{p} \quad \boxed{+C}$$

- Thus the joint density of $(\boldsymbol{\theta}, \mathbf{p})$ is

$$\pi(\boldsymbol{\theta}, \mathbf{p}) \propto \exp\{-H(\boldsymbol{\theta}, \mathbf{p})\} = \exp\{-U(\boldsymbol{\theta})\}\exp\{-K(\mathbf{p})\}$$

[1]Often set $\mathbf{M} = \mathbf{I}_d$ for simplicity, but more informative $\mathbf{M}$ works better.

Definition (Hamiltonian dynamics)

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \frac{\partial}{\partial \mathbf{p}} H(\boldsymbol{\theta}, \mathbf{p}) &= \mathbf{M}^{-1}\mathbf{p} \\ \dot{\mathbf{p}} &= -\frac{\partial}{\partial \boldsymbol{\theta}} H(\boldsymbol{\theta}, \mathbf{p}) &= -\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}) \end{aligned}$$

Leapfrog: numerical integrator

$$\begin{aligned} \mathbf{p}(t+\varepsilon/2) &= \mathbf{p}(t) - (\varepsilon/2)\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}(t)) \\ \boldsymbol{\theta}(t+\varepsilon) &= \boldsymbol{\theta}(t) + \varepsilon \mathbf{M}^{-1}\mathbf{p}(t+\varepsilon/2) \\ \mathbf{p}(t+\varepsilon) &= \mathbf{p}(t+\varepsilon/2) - (\varepsilon/2)\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}(t+\varepsilon)) \end{aligned}$$

- Run for **L** steps and accept the joint proposal of $\mathbf{z}^* := (\boldsymbol{\theta}^*, \mathbf{p}^*)$ with

$$\alpha = \min\{1, \exp(-H(\mathbf{z}^*) + H(\mathbf{z}))\}$$

On the manifold $\{f(\cdot;\boldsymbol{\theta})\}$ with metric $G(\boldsymbol{\theta}) = -\mathrm{E}_{\mathbf{x}|\boldsymbol{\theta}}[\nabla^2_{\boldsymbol{\theta}} \log f(\mathbf{x};\boldsymbol{\theta})]$:

$$
\begin{aligned}
H(\boldsymbol{\theta},\mathbf{p}) &= U(\boldsymbol{\theta}) + K(\mathbf{p},\boldsymbol{\theta}) \\
&= -\log \pi(\boldsymbol{\theta}) + \frac{1}{2}\log\det \mathbf{G}(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{p}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})^{-1}\mathbf{p} \\
&\equiv \phi(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{p}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})^{-1}\mathbf{p}
\end{aligned}
$$

where $\mathbf{p}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}))$. Girolami and Calderhead (2011) propose:

Definition (Riemannian Hamiltonian dynamics)

$$
\begin{aligned}
\dot{\boldsymbol{\theta}} &= \frac{\partial}{\partial p} H(\boldsymbol{\theta},\mathbf{p}) = \mathbf{G}(\boldsymbol{\theta})^{-1}\mathbf{p} \\
\dot{\mathbf{p}} &= -\frac{\partial}{\partial \boldsymbol{\theta}} H(\boldsymbol{\theta},\mathbf{p}) = -\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{p}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})^{-1}\partial\mathbf{G}(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta})^{-1}\mathbf{p}
\end{aligned}
$$

To resolve the implicitness of RHMC, [6] propose

Definition (Lagrangian Dynamics)

$$\dot{\boldsymbol{\theta}} = \mathbf{G}(\boldsymbol{\theta})^{-1}\mathbf{p}$$

$$\dot{\mathbf{p}} = -\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{p}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})^{-1}\partial\mathbf{G}(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta})^{-1}\mathbf{p}$$

$$\boxed{\mathbf{p} \to \mathbf{v}} \Big\Downarrow \textbf{Lagrangian Dynamics}$$

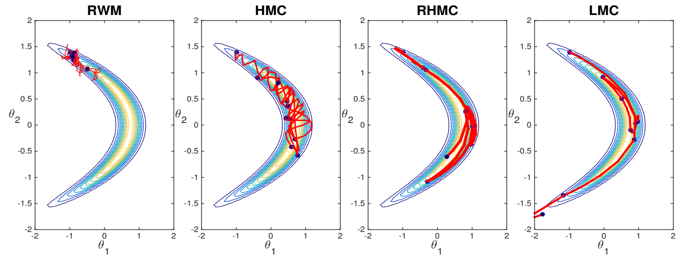
$$\dot{\boldsymbol{\theta}} = \mathbf{v}$$

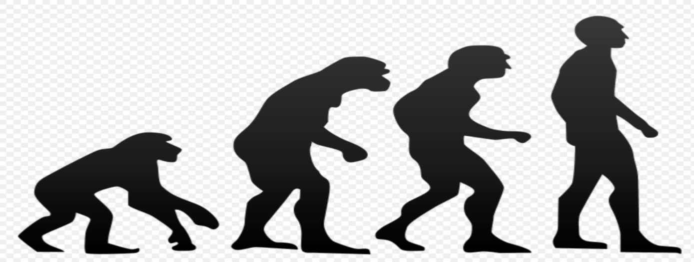
$$\dot{\mathbf{v}} = -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{v} - \mathbf{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta})$$

- Not Hamiltonian dynamics of $(\boldsymbol{\theta}, \mathbf{v})$!
- An *explicit* integrator can be found more stable and efficient.

Geometric Monte Carlo

[10, 8, 2, 6]





Geodesic HMC/LMC

Splitting Lagrangian Dynamics [1, 4]

$$\begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{v}\\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{v} - \mathbf{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta})\end{aligned} \tag{1}$$

$$\begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{0}\\ \dot{\mathbf{v}} &= -\frac{1}{2}\mathbf{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta})\end{aligned} \quad \longleftrightarrow \quad \begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{v}\\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{v}\end{aligned}$$

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{v} - \mathbf{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}) \end{aligned} \tag{1}$$

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{0} \\ \dot{\mathbf{v}} &= -\frac{1}{2}\mathbf{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}) \end{aligned} \quad \longleftrightarrow \quad \begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{v} \end{aligned}$$

$$\downarrow \qquad\qquad \downarrow$$

$$\begin{bmatrix} \boldsymbol{\theta}(t) \\ \mathbf{v}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}(0) \\ \mathbf{v}(0) - \frac{t}{2}\mathbf{G}(\boldsymbol{\theta}(0))^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}(0)) \end{bmatrix}$$

$$\begin{bmatrix} \boldsymbol{\theta}(t) \\ \mathbf{v}(t) \end{bmatrix} = \mathrm{geod}_t\left(\begin{bmatrix} \boldsymbol{\theta}(0) \\ \mathbf{v}(0) \end{bmatrix}\right)$$

Geodesic HMC/LMC

Splitting Lagrangian Dynamics [1, 4]

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{v} - \mathbf{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}) \end{aligned} \tag{1}$$

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{0} \\ \dot{\mathbf{v}} &= -\frac{1}{2}\mathbf{G}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}) \end{aligned} \longleftrightarrow \begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{v} \end{aligned}$$

$$\downarrow \qquad\qquad \downarrow$$

$$\begin{bmatrix} \boldsymbol{\theta}(t) \\ \mathbf{v}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}(0) \\ \mathbf{v}(0) - \frac{t}{2}\mathbf{G}(\boldsymbol{\theta}(0))^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}(0)) \end{bmatrix} \qquad \begin{bmatrix} \boldsymbol{\theta}(t) \\ \mathbf{v}(t) \end{bmatrix} = \text{geod}_t\left(\begin{bmatrix} \boldsymbol{\theta}(0) \\ \mathbf{v}(0) \end{bmatrix}\right)$$

$$\downarrow \qquad\qquad \downarrow$$

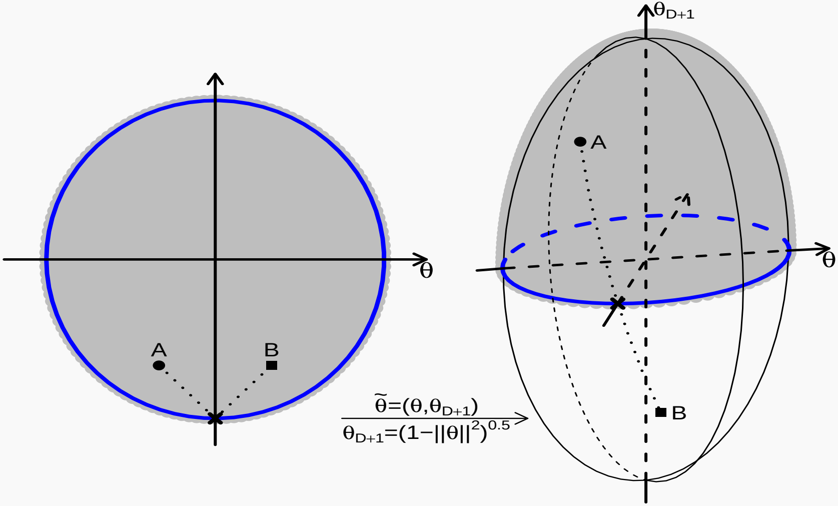
$$\begin{aligned} \mathbf{v}(t+\varepsilon/2) &= \mathbf{v}(t) - (\varepsilon/2)\mathbf{G}(\boldsymbol{\theta}(t))^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}(t)) \\ \begin{bmatrix} \boldsymbol{\theta}(t+\varepsilon) \\ \mathbf{v}^*(t+\varepsilon) \end{bmatrix} &= \text{geod}_\varepsilon\left(\begin{bmatrix} \boldsymbol{\theta}(t) \\ \mathbf{v}(t+\varepsilon/2) \end{bmatrix}\right) \\ \mathbf{v}(t+\varepsilon) &= \mathbf{v}^*(t+\varepsilon) - (\varepsilon/2)\mathbf{G}(\boldsymbol{\theta}(t+\varepsilon))^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}(t+\varepsilon)) \end{aligned}$$

Spherical HMC

in the Cartesian coordinate

Ball Type Constraints

Change of domain: from unit ball $\mathcal{B}_{\mathbf{0}}^D(1)$ to sphere $\mathcal{S}^D$



Spherical HMC in the Cartesian coordinate

for ball type constraints

$$\mathcal{B}_{\mathbf{0}}^D(1) := \{\boldsymbol{\theta} \in \mathbb{R}^D : \|\boldsymbol{\theta}\|_2 = \sqrt{\sum_{i=1}^D \theta_i^2} \leq 1\}$$

$$\xrightarrow[\theta_{D+1} = \pm\sqrt{1-\|\boldsymbol{\theta}\|_2^2}]{\boldsymbol{\theta} \mapsto \tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}, \theta_{D+1})}$$

$$\mathcal{S}^D := \{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{D+1} : \|\tilde{\boldsymbol{\theta}}\|_2 = 1\}$$

Spherical HMC in the Cartesian coordinate

for ball type constraints

$\mathcal{B}_{\mathbf{0}}^D(1) := \{\boldsymbol{\theta} \in \mathbb{R}^D : \|\boldsymbol{\theta}\|_2 = \sqrt{\sum_{i=1}^D \theta_i^2} \le 1\}$

$$\xrightarrow[\theta_{D+1}=\pm\sqrt{1-\|\boldsymbol{\theta}\|_2^2}]{\boldsymbol{\theta}\mapsto\tilde{\boldsymbol{\theta}}=(\boldsymbol{\theta},\theta_{D+1})}$$

$\mathcal{S}^D := \{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{D+1} : \|\tilde{\boldsymbol{\theta}}\|_2 = 1\}$

Change of variables

$$\int_{\mathcal{B}_0^D(1)} f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{\mathcal{B}} = \int_{\mathcal{S}_+^D} f(\tilde{\boldsymbol{\theta}}) \left| \frac{d\boldsymbol{\theta}_{\mathcal{B}}}{d\boldsymbol{\theta}_{\mathcal{S}_c}} \right| d\boldsymbol{\theta}_{\mathcal{S}_c} = \int_{\mathcal{S}_+^D} f(\tilde{\boldsymbol{\theta}}) |\theta_{D+1}| d\boldsymbol{\theta}_{\mathcal{S}_c}$$

where $f(\tilde{\boldsymbol{\theta}}) \equiv f(\boldsymbol{\theta})$.

Spherical HMC in the Cartesian coordinate

for ball type constraints

$\mathcal{B}_{\mathbf{0}}^D(1) := \{\boldsymbol{\theta} \in \mathbb{R}^D : \|\boldsymbol{\theta}\|_2 = \sqrt{\sum_{i=1}^D \theta_i^2} \le 1\}$

$$\xrightarrow[\theta_{D+1}=\pm\sqrt{1-\|\boldsymbol{\theta}\|_2^2}]{\boldsymbol{\theta}\mapsto\tilde{\boldsymbol{\theta}}=(\boldsymbol{\theta},\theta_{D+1})}$$

$\mathcal{S}^D := \{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{D+1} : \|\tilde{\boldsymbol{\theta}}\|_2 = 1\}$

Change of variables

$$\int_{\mathcal{B}_0^D(1)} f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{\mathcal{B}} = \int_{\mathcal{S}_+^D} f(\tilde{\boldsymbol{\theta}}) \left| \frac{d\boldsymbol{\theta}_{\mathcal{B}}}{d\boldsymbol{\theta}_{\mathcal{S}_c}} \right| d\boldsymbol{\theta}_{\mathcal{S}_c} = \int_{\mathcal{S}_+^D} f(\tilde{\boldsymbol{\theta}}) |\theta_{D+1}| d\boldsymbol{\theta}_{\mathcal{S}_c}$$

where $f(\tilde{\boldsymbol{\theta}}) \equiv f(\boldsymbol{\theta})$.

What We Want:
$\boldsymbol{\theta} \sim f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{\mathcal{B}}$

$$\xleftarrow[\text{weigh it by } |\theta_{D+1}|]{\text{drop } \theta_{D+1}}$$

What We Sample:
$\tilde{\boldsymbol{\theta}} \sim f(\tilde{\boldsymbol{\theta}}) d\boldsymbol{\theta}_{\mathcal{S}_c}$

Split Lagrangian dynamics on sphere
in the Cartesian coordinate [7, 5]

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}} \boldsymbol{\Gamma}_{\mathcal{S}_c}(\boldsymbol{\theta}) \mathbf{v} - \mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})^{-1} \nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}) \end{aligned} \tag{2}$$

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{0} \\ \dot{\mathbf{v}} &= -\frac{1}{2} \mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})^{-1} \nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}) \end{aligned} \longleftrightarrow \begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}} \boldsymbol{\Gamma}_{\mathcal{S}_c}(\boldsymbol{\theta}) \mathbf{v} \end{aligned}$$

Split Lagrangian dynamics on sphere

in the Cartesian coordinate [7, 5]

$$\begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}_{\mathcal{S}_c}(\boldsymbol{\theta})\mathbf{v} - \mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta})\end{aligned} \tag{2}$$

$$\begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{0} \\ \dot{\mathbf{v}} &= -\frac{1}{2}\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta})\end{aligned} \longleftrightarrow \begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}_{\mathcal{S}_c}(\boldsymbol{\theta})\mathbf{v}\end{aligned}$$

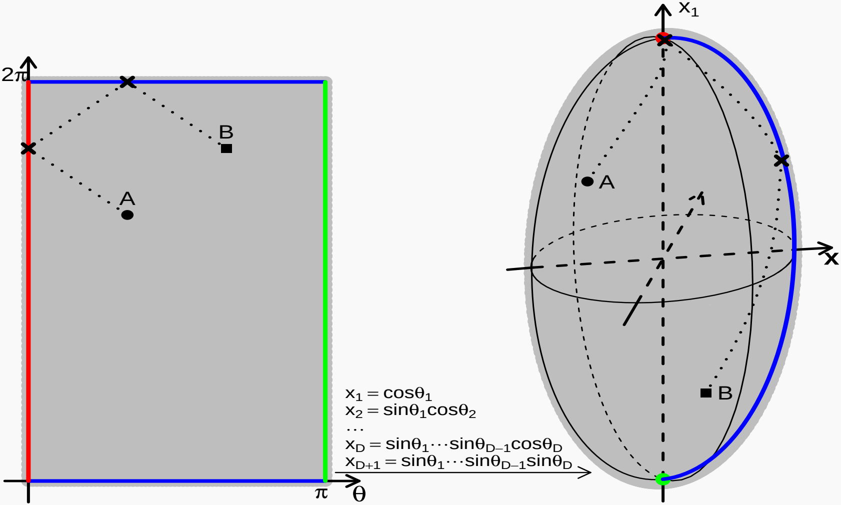
$$\downarrow$$

$$\begin{aligned}\tilde{\boldsymbol{\theta}}(t) &= \tilde{\boldsymbol{\theta}}(0) \\ \tilde{\mathbf{v}}(t) &= \tilde{\mathbf{v}}(0) \\ &- \frac{t}{2}\left[\begin{bmatrix}\mathbf{I}_D \\ \mathbf{0}^{\mathsf{T}}\end{bmatrix} - \tilde{\boldsymbol{\theta}}(0)\boldsymbol{\theta}(0)^{\mathsf{T}}\right]\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}(0))\end{aligned}$$

$$\begin{aligned}\tilde{\boldsymbol{\theta}}(t) =& \tilde{\boldsymbol{\theta}}(0)\cos(\|\tilde{\mathbf{v}}(0)\|_2 t) \\ &+ \frac{\tilde{\mathbf{v}}(0)}{\|\tilde{\mathbf{v}}(0)\|_2}\sin(\|\tilde{\mathbf{v}}(0)\|_2 t) \\ \tilde{\mathbf{v}}(t) =& -\tilde{\boldsymbol{\theta}}(0)\|\tilde{\mathbf{v}}(0)\|_2\sin(\|\tilde{\mathbf{v}}(0)\|_2 t) \\ &+ \tilde{\mathbf{v}}(0)\cos(\|\tilde{\mathbf{v}}(0)\|_2 t)\end{aligned}$$

Box Type Constraints

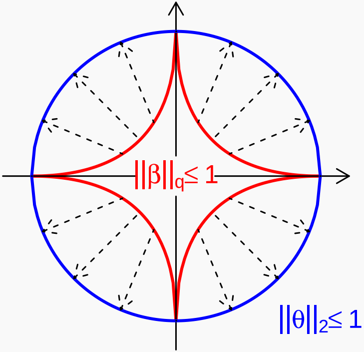
Change of domain: from rectangle $\mathcal{R}_0^D$ to sphere $\mathcal{S}^D$



General q-norm Constraints

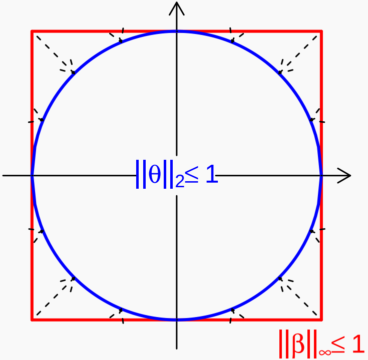
Mapping q-norm constrained domain to unit ball

$0 < q < \infty$



$$\theta = \mathrm{sgn}(\beta)|\beta|^{(q/2)}$$

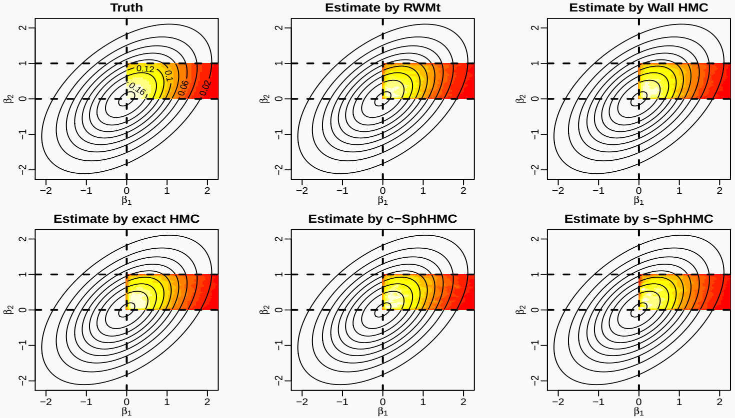
$q = \infty$



$$\theta = \beta \frac{\|\beta\|_\infty}{\|\beta\|_2}$$

Truncated Multivariate Gaussian

$$\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}1 & .5\\ .5 & 1\end{bmatrix}\right), \qquad 0 \le \beta_1 \le 5, \quad 0 \le \beta_2 \le 1$$



Truncated Multivariate Gaussian

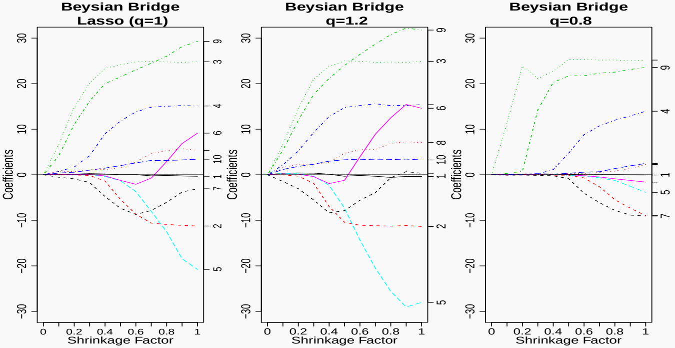
- To evaluate efficiency, we increase the dimensionality for $D = 10,\ 100$

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),\ \Sigma_{ij} = 1/(1+|i-j|); \quad 0 \le \beta_1 \le 5,\ 0 \le \beta_i \le 0.5,\ i \neq 1.$$

- RWMt: $> 95\%$ of times proposals rejected due to constraint violation.
- Wall HMC: average wall hits 3.81 (L=2, D=10), 6.19 (L=5, D=100).

Dim	Method	AP	s/iter	ESS(min,med,max)	Min(ESS)/s	spdup
D=10	RWMt	0.62	5.72E-05	(48,691,736)	7.58	1.00
	Wall HMC	0.83	1.19E-04	(31904,86275,87311)	2441.72	322.33
	exact HMC	1.00	7.60E-05	(1e+05,1e+05,1e+05)	11960.29	1578.87
	c-SphHMC	0.82	2.53E-04	(62658,85570,86295)	2253.32	297.46
	s-SphHMC	0.79	2.02E-04	(76088,1e+05,1e+05)	3429.56	452.73
D=100	RWMt	0.81	5.45E-04	(1,4,54)	0.01	1.00
	Wall HMC	0.74	2.23E-03	(17777,52909,55713)	72.45	5130.21
	exact HMC	1.00	4.65E-02	(97963,1e+05,1e+05)	19.16	1356.64
	c-SphHMC	0.73	3.45E-03	(55667,68585,72850)	146.75	10390.94
	s-SphHMC	0.87	2.30E-03	(74476,99670,1e+05)	294.31	20839.43

Bayesian Bridge: regularized regression



- Obtain the coefficients β by minimizing the residual sum of squares (RSS) subject to a constraint on the magnitude of β

$$\min_{\|\beta\|_q \le t} \mathrm{RSS}(\beta), \qquad \mathrm{RSS}(\beta) := \sum_i (y_i - \beta_0 - x_i^{\mathrm{T}} \beta)^2$$

- Polson et al (2013) have Bayesian Bridge with complicated priors

Spherical LMC

on the probability simplex

The Probability Simplex

- A class of models having probability distributions defined on *simplex*

$$\Delta^K := \{\boldsymbol{\pi} \in \mathbb{R}^D |\ \pi_k \geq 0, \sum_{k=1}^{K} \pi_d = 1\}$$

- *Latent Dirichlet Allocation (LDA)* (Blei et al., 2003) is a hierarchical Bayesian model frequently used to model document topics.
- 1-norm constraint: identify the first (all positive) orthant with others.
- $T_{\Delta \to \sqrt{\Delta}} : \boldsymbol{\pi} \mapsto \boldsymbol{\theta} = \sqrt{\boldsymbol{\pi}}$ maps the simplex to the sphere

$$\sqrt{\Delta}^K := \{\boldsymbol{\theta} \in \mathcal{S}^{K-1} | \theta_k \geq 0, \forall\, k = 1, \cdots, K\} \subset \mathcal{S}^{K-1}$$

Distribution on the Probability Simplex

- Prototype example: Dirichlet-Multinomial distribution

$$p(x_i = k|\boldsymbol{\pi}) = \pi_k, \quad k = 1, \cdots, K$$

$$p(\boldsymbol{\pi}) \propto \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$$

$$p(\boldsymbol{\pi}|\mathbf{x}) \propto \prod_{k=1}^{K} \pi_k^{n_k + \alpha_k - 1}, \quad n_k = \sum_{i=1}^{N} I(x_i = k), \; n = \sum_{k=1}^{K} n_k$$

- Fisher metric on $\sqrt{\Delta}$ coincides $\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})$ on $\mathcal{S}^{K-1}$ up to a constant.

$$\mathbf{G}_{\Delta}(\boldsymbol{\pi}_{-K}) = n[\mathrm{diag}(1/\boldsymbol{\pi}_{-K}) + \mathbf{1}\mathbf{1}^{\mathsf{T}}/\pi_K]$$

$$\mathbf{G}_{\sqrt{\Delta}}(\boldsymbol{\theta}) = \frac{d\boldsymbol{\pi}_{-K}^{\mathsf{T}}}{d\boldsymbol{\theta}_{-K}} \mathbf{G}_{\Delta}(\boldsymbol{\pi}_{-K}) \frac{d\boldsymbol{\pi}_{-K}}{d\boldsymbol{\theta}_{-K}^{\mathsf{T}}} = 4n\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})$$

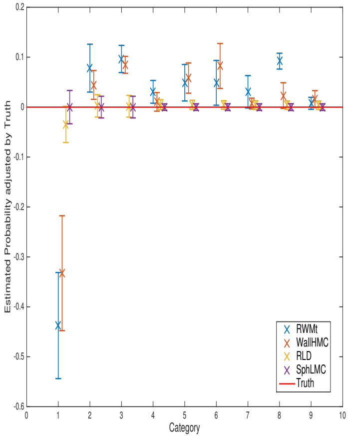
Spherical LMC
on the Probability Simplex [5]

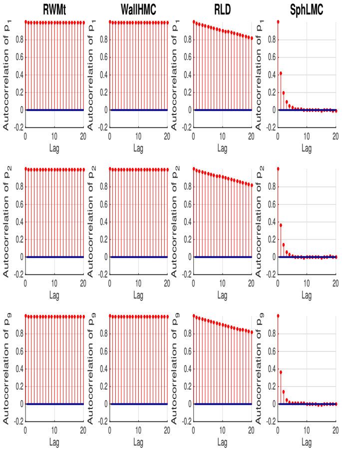
- Use $\mathbf{G}_{\sqrt{\Delta}}(\boldsymbol{\theta})$ instead of $\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})$ in c-SphHMC.
- Include the volume adjustment term, $\left|\frac{d\boldsymbol{\beta}_{\mathcal{D}}}{d\boldsymbol{\theta}_{\mathcal{S}}}\right|$ in the Hamiltonian

$$H(\boldsymbol{\theta}, \mathbf{v}) = \phi(\boldsymbol{\theta}) + \frac{1}{2}\langle \mathbf{v}, \mathbf{v}\rangle_{\mathbf{G}_{\sqrt{\Delta}}(\boldsymbol{\theta})}, \quad \phi(\boldsymbol{\theta}) = U(\boldsymbol{\theta}) - \log\left|\frac{d\boldsymbol{\beta}_{\mathcal{D}}}{d\boldsymbol{\theta}_{\mathcal{S}}}\right|$$

- No afterward re-weight: online learning
- c-SphHMC $\xrightarrow{\textit{above modifications}}$ Spherical Lagrangian Monte Carlo.
- SphLMC: stems from the Fisher metric on the simplex.

Spherical LMC
on the Probability Simplex





Spherical HMC

in the infinite dimension

Representation of Probability Densities

- Consider probability distributions over smooth manifolds $\mathcal{D}$. Having fixed a background measure μ, let

$$\mathcal{P} := \left\{ p : \mathcal{D} \to \mathbb{R} \mid p \geq 0, \int_{\mathcal{D}} p(x)\, \mu(dx) = 1 \right\} \tag{3}$$

- Define the following nonparametric Fisher metric on the tangent space $T_p\mathcal{P} := \left\{ \phi \in C^\infty(\mathcal{D}) \mid \int_{\mathcal{D}} \phi(x)\, \mu(dx) = 0 \right\}$:

$$g_F(\phi, \psi)_p := \int_{\mathcal{D}} \frac{\phi(x)\psi(x)}{p(x)} \mu(dx). \tag{4}$$

- The square-root mapping $S : (\mathcal{P}, g_F) \to (\mathcal{Q}, \langle \cdot, \cdot \rangle_2)$, $S(p) = q = 2\sqrt{p}$ is a Riemannian isometry, where $\mathcal{Q}$ is ∞-dimensional sphere in $L^2(\mathcal{D})$

$$\mathcal{Q} := \left\{ q : \mathcal{D} \to \mathbb{R} \mid \int_{\mathcal{D}} q(x)^2\, \mu(dx) = 1 \right\}, \quad \langle f, h \rangle_2 = \int_{\mathcal{D}} fh d\mu(x) \tag{5}$$

Nonparametric Density Modeling

[3]

- It is easier to work with root density $q \in \mathcal{Q}$ (e.g. clean geodesic flow).
- Restrict Gaussian process prior $q(\cdot) \sim \mathcal{GP}(0, K(\cdot))$ to $\mathcal{Q}$, where the covariance operator $K = \sigma^2(\alpha - \Delta)^{-s}$ has eigen-pairs $\{\lambda_i^2, \phi_i(x)\}_{i=1}^\infty$.
- Then $\|q(x)\|_2 = \|\sum_{i=1}^\infty q_i\phi_i(x)\|_2 = 1$ with $q_i \sim \mathcal{N}(0, \lambda_i^2)$ implies

$$\|q\|_2^2 := \sum_{i=1}^{\infty} q_i^2 = 1, \quad i.e.\ q := (q_i) \in \mathcal{S}^\infty \tag{6}$$

- Given data $x = \{x_n \in \mathcal{D}\}_n^N$, we have the posterior density

$$\pi(q|x) \propto \pi(q)\,\pi(x|q) = \prod_{i=1}^{\infty} \exp\left(-q_i^2/(2\lambda_i^2)\right)\delta_{\|q\|_2}(1) \prod_{n=1}^{N} q^2(x_n) \tag{7}$$

- Sampling $q = (q_i)$ can be done by spherical HMC [7].

Nonparametric Density Modeling

simulation result

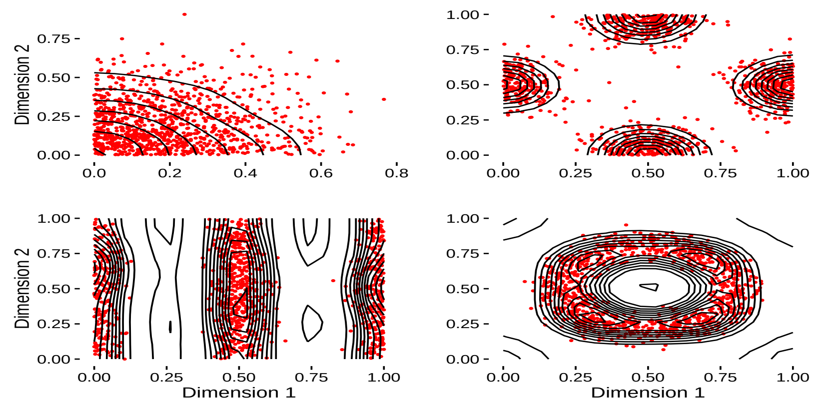


Figure: The contours (black) of the posterior median from 1,000 draws of the χ^2-process density sampler. Each posterior is conditioned on 1,000 data points (red).

PD-HMC

on the Space of PD Matrices

The Space of PD Matrices

- Denote $\mathcal{S}_d(\mathbb{C})$ as the space of $d \times d$ Hermitian matrices, and $\mathcal{S}_d^+(\mathbb{C})$ as its subspace of PD matrices. Note $\mathcal{S}_d^+(\mathbb{C}) = GL(d,\mathbb{C})/U(d)$.
- The group action is given by conjugation: for $G \in GL(d,\mathbb{C})$, $\Sigma \in \mathcal{S}_d^+$,

$$G^* \Sigma = G \Sigma G^H \ . \tag{8}$$

- $\mathcal{S}_d$ happens to be the tangent space to $\mathcal{S}_d^+$ at the identity, that is, $T_{Id}\mathcal{S}_d^+ = \mathcal{S}_d$. The action translates vectors between tangent spaces:

$$\Sigma^{1/2*} : T_{Id}\mathcal{S}_d^+ \to T_\Sigma \mathcal{S}_d^+, \ V \mapsto \Sigma^{1/2} V \Sigma^{1/2} \tag{9}$$

- Élie Cartan constructed a natural Riemannian metric $g_\Sigma(\cdot,\cdot)$

$$g_\Sigma(V_1, V_2) = \mathrm{tr}(\Sigma^{-1} V_1 \Sigma^{-1} V_2), \quad \forall V_1, V_2 \in T_\Sigma \mathcal{S}_d^+ \tag{10}$$

Split Lagrangian Dynamics on PD Matrices
[4]

$$\dot{\Sigma} = V$$
$$\mathrm{vech}(\dot{V}) = -\mathrm{vech}(V)^H \mathbf{\Gamma}(\Sigma)\mathrm{vech}(V) - G(\Sigma)^{-1}\mathrm{vech}(\nabla_\Sigma U(\Sigma))$$

$$\dot{\Sigma} = 0$$
$$\mathrm{vech}(\dot{V}) = -\frac{1}{2} G(\Sigma)^{-1}\mathrm{vech}(\nabla_\Sigma U(\Sigma))$$

$\longleftrightarrow$

$$\dot{\Sigma} = V$$
$$\mathrm{vech}(\dot{V}) = -\mathrm{vech}(V)^H \mathbf{\Gamma}(\Sigma)\mathrm{vech}(V)$$

Split Lagrangian Dynamics on PD Matrices
[4]

$$\dot{\Sigma} = V$$
$$\mathrm{vech}(\dot{V}) = -\mathrm{vech}(V)^H \mathbf{\Gamma}(\Sigma)\mathrm{vech}(V) - G(\Sigma)^{-1}\mathrm{vech}(\nabla_\Sigma U(\Sigma))$$

$$\dot{\Sigma} = 0$$
$$\mathrm{vech}(\dot{V}) = -\frac{1}{2} G(\Sigma)^{-1}\mathrm{vech}(\nabla_\Sigma U(\Sigma))$$

$\longleftrightarrow$

$$\dot{\Sigma} = V$$
$$\mathrm{vech}(\dot{V}) = -\mathrm{vech}(V)^H \mathbf{\Gamma}(\Sigma)\mathrm{vech}(V)$$

$\downarrow$

$$\Sigma(t) = \Sigma(0)$$
$$\mathrm{vech}(V(t)) = \mathrm{vech}(V(0)) - \frac{t}{2} G(\Sigma(0))^{-1}\mathrm{vech}(\nabla_\Sigma U(\Sigma(0)))$$

$\downarrow$

$$\Sigma(t) = \Sigma(0)^{1/2} \exp\left(t\Sigma(0)^{-1/2} V(0) \Sigma(0)^{-1/2}\right) \Sigma(0)^{1/2}$$
$$V(t) = V(0)\Sigma(0)^{-1/2} \exp\left(t\Sigma(0)^{-1/2} V(0) \Sigma(0)^{-1/2}\right) \Sigma(0)^{1/2}$$

- Given multivariate time series $y(t) = (y_1(t), \ldots, y_d(t))^{\mathsf{T}} \in \mathbb{R}^d$, $t = 1, \ldots, T$, The power spectral density matrix is the Fourier transform of the lagged variance-covariance matrix Γ_ℓ:

$$\begin{aligned}\Sigma(\omega) &= \sum_{\ell=-\infty}^{\infty} \Gamma_\ell \exp(-i2\pi\omega\ell) \\ \Gamma_\ell &= \mathrm{Cov}\big(y(t),\, y(t-\ell)\big) = \mathrm{E}\Big(\big(y(t)-\mu\big)\big(y(t-\ell)-\mu\big)^T\Big)\end{aligned} \tag{11}$$

- For certain frequency band, e.g. alpha-band 7.5-12.5Hz, we assume the discrete Fourier transformed time series $Y(\omega_k) \in \mathbb{C}^d$ follow a complex multivariate Gaussian distribution:

$$Y(\omega_k) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y(t) \exp(-i2\pi\omega_k t) \overset{iid}{\sim} \mathrm{CN}_d(0, \Sigma_\alpha)\ , \tag{12}$$

  where the spectral density matrix Σ_α shared by the entire band.
- Σ can be given inverse-Wishart or reference priors and its posterior can be sampled by PD-HMC.

Validation of PD-HMC

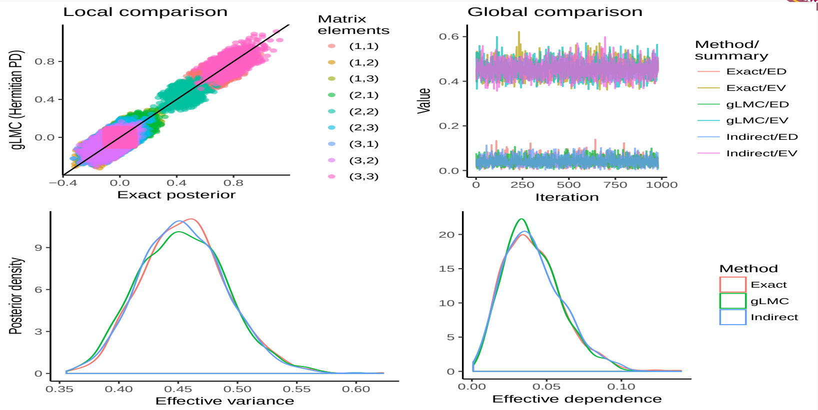


Figure: These figures provide empirical validation for the well-posedness of PDHMC. On the left is a quantile-quantile plot comparing the Hermitian PDHMC posterior sample with that of the closed-form posterior for the complex Gaussian inverse-Wishart model. Both real and imaginary elements are included, and points are jittered for visibility. On the right are posterior samples of 'global' matrix statistics pertaining both to (symmetric) PDHMC and the closed-form solution. These statistics are the effective variance and the effective dependence, built off the covariance matrix and the correlation matrix, respectively.

Estimation of Coherence of LFP signals

recorded in the CA1 region of a rat hippocampus

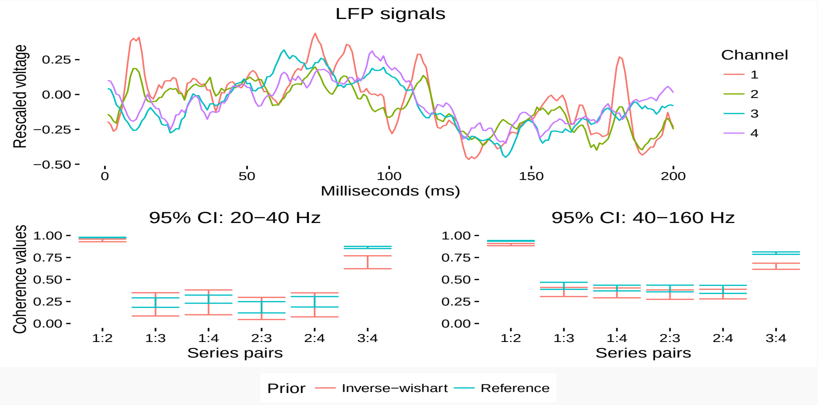


Figure: A 4-dimensional LFP signal with credible intervals for 6 coherences measured at 20-40 Hz (left) and 40-160 Hz (right). First 200 samples are shown for ease of visualization; the multi-dimensional time series totals 4,000 samples in length. Coherence profiles are remarkably similar between the two frequency bands considered.

Conclusion

- *Geomtry* can help to derive a natural and efficient framework to handle manifold constraints in Bayesian inference.
- Spherical HMC and Spherical LMC move on sphere freely while implicitly handling constraints, demonstrating substantial advantage over existing methods.
- PD-HMC provides a direct and efficient way to sample PD matrices in Bayesian statistics. It has potential impact on Bayesian covariance modeling.
- They open a door for more research in efficient Bayesian computation, e.g. other manifold MCMC, generalization to infinite-dimensions, etc.

References I

[1] Byrne, S. and Girolami, M. (2013). Geodesic monte carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845.

[2] Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, (with discussion) 73(2):123–214.

[3] Holbrook, A., Lan, S., Streets, J., and Shahbaba, B. (2017). The nonparametric Fisher information geometry and the chi-square process density prior. arXiv:1707.03117.

[4] Holbrook, A., Lan, S., Vandenberg-Rodes, A., and Shahbaba, B. (2018). Geodesic lagrangian monte carlo over the space of positive definite matrices: with application to bayesian spectral density estimation. *Journal of Statistical Computation and Simulation*, 88(5):982–1002.

[5] Lan, S. and Shahbaba, B. (2016). *Algorithmic Advances in Riemannian Geometry and Applications*, chapter 2-Sampling Constrained Probability Distributions Using Spherical Augmentation, pages 25–71. Advances in Computer Vision and Pattern Recognition. Springer International Publishing, 1 edition.

[6] Lan, S., Stathopoulos, V., Shahbaba, B., and Girolami, M. (2015). Markov Chain Monte Carlo from Lagrangian Dynamics. *Journal of Computational and Graphical Statistics*, 24(2):357–378.

[7] Lan, S., Zhou, B., and Shahbaba, B. (2014). Spherical Hamiltonian Monte Carlo for Constrained Target Distributions. In Xing, E. P. and Jebara, T., editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 629–637, Bejing, China.

[8] Neal, R. M. (2010). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G., and Meng, X. L., editors, *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC.

[9] Pennec, X., Fillard, P., and Ayache, N. (2006). A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66.

[10] Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363.

Thank you !

https://math.asu.edu/∼slan

Canonical spherical metric

- Here, the proper metric on $\mathcal{S}^D$ is called *canonical spherical metric*:

Definition (canonical spherical metric)

$$\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta}) = \mathbf{I}_D + \frac{\boldsymbol{\theta}\boldsymbol{\theta}^{\mathsf{T}}}{\theta_{D+1}^2} = \mathbf{I}_D + \frac{\boldsymbol{\theta}\boldsymbol{\theta}^{\mathsf{T}}}{1 - \|\boldsymbol{\theta}\|_2^2} \tag{13}$$

- For any vector $\tilde{\mathbf{v}} = (\mathbf{v}, v_{D+1}) \in T_{\tilde{\boldsymbol{\theta}}}\mathcal{S}^D := \{\tilde{\mathbf{v}} \in \mathbb{R}^{D+1} : \tilde{\boldsymbol{\theta}}^{\mathsf{T}}\tilde{\mathbf{v}} = 0\}$, $\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})$ can be viewed as a way to express the length of $\tilde{\mathbf{v}}$ in $\mathbf{v}$:

$$\begin{aligned}\mathbf{v}^{\mathsf{T}}\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})\mathbf{v} &= \|\mathbf{v}\|_2^2 + \frac{\mathbf{v}^{\mathsf{T}}\boldsymbol{\theta}\boldsymbol{\theta}^{\mathsf{T}}\mathbf{v}}{\theta_{D+1}^2} = \|\mathbf{v}\|_2^2 + \frac{(-\theta_{D+1}v_{D+1})^2}{\theta_{D+1}^2} \\ &= \|\mathbf{v}\|_2^2 + v_{D+1}^2 = \|\tilde{\mathbf{v}}\|_2^2\end{aligned}$$

Hamiltonian (Lagrangian) dynamics on sphere

in the Cartesian coordinate

On $\mathcal{B}_{\mathbf{0}}^{D}(1)$

$$H(\boldsymbol{\theta}, \mathbf{v}) = U(\boldsymbol{\theta}) + K(\mathbf{v}) = -\log f(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{v}^{\mathsf{T}}\mathbf{I}\mathbf{v}$$

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$$

$$\xrightarrow{\boldsymbol{\theta}\mapsto\tilde{\boldsymbol{\theta}}}$$

$$\xrightarrow{\mathbf{v}\mapsto\tilde{\mathbf{v}}}$$

On $\mathcal{S}^{D}$

$$H^{*}(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{v}}) = U(\tilde{\boldsymbol{\theta}}) + K(\tilde{\mathbf{v}}) = -\log f(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{v}^{\mathsf{T}}\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})\mathbf{v}$$

$$\tilde{\mathbf{v}} \sim (\mathbf{I}_{D+1} - \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^{\mathsf{T}})\mathcal{N}(\mathbf{0}, \mathbf{I}_{D+1})$$

Hamiltonian (Lagrangian) dynamics on sphere

in the Cartesian coordinate

On $\mathcal{B}_{\mathbf{0}}^{D}(1)$ | On $\mathcal{S}^{D}$

$$H(\boldsymbol{\theta}, \mathbf{v}) = U(\boldsymbol{\theta}) + K(\mathbf{v}) = -\log f(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{v}^{\mathsf{T}}\mathbf{I}\mathbf{v}$$

$\xrightarrow{\boldsymbol{\theta}\mapsto\tilde{\boldsymbol{\theta}}}$

$$H^{*}(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{v}}) = U(\tilde{\boldsymbol{\theta}}) + K(\tilde{\mathbf{v}}) = -\log f(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{v}^{\mathsf{T}}\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})\mathbf{v}$$

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$$

$\xrightarrow{\mathbf{v}\mapsto\tilde{\mathbf{v}}}$

$$\tilde{\mathbf{v}} \sim (\mathbf{I}_{D+1} - \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^{\mathsf{T}})\mathcal{N}(\mathbf{0}, \mathbf{I}_{D+1})$$

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}) \\ \|\boldsymbol{\theta}\|_2 &\le 1 \end{aligned}$$

$\longrightarrow$

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}_{\mathcal{S}_c}(\boldsymbol{\theta})\mathbf{v} - \mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}) \\ \theta_{D+1} &= \sqrt{1 - \|\boldsymbol{\theta}\|_2^2}, \quad v_{D+1} = -\boldsymbol{\theta}^{\mathsf{T}}\mathbf{v}/\theta_{D+1} \end{aligned}$$

Algorithm 1 Spherical HMC in the Cartesian coordinate $\boxed{(c - SphHMC)}$

Initialize $\tilde{\boldsymbol{\theta}}^{(1)}$ at current $\tilde{\boldsymbol{\theta}}$ after transformation $T_{\mathcal{D}\to\mathcal{S}}$
Sample a new velocity value $\tilde{\mathbf{v}}^{(1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{D+1})$
Set $\tilde{\mathbf{v}}^{(1)} \leftarrow \tilde{\mathbf{v}}^{(1)} - \tilde{\boldsymbol{\theta}}^{(1)}(\tilde{\boldsymbol{\theta}}^{(1)})^{\mathsf{T}}\tilde{\mathbf{v}}^{(1)}$
Calculate $H(\tilde{\boldsymbol{\theta}}^{(1)}, \tilde{\mathbf{v}}^{(1)}) = U(\boldsymbol{\theta}^{(1)}) + K(\tilde{\mathbf{v}}^{(1)})$
for $\ell = 1$ to L **do**

$$\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})} = \tilde{\mathbf{v}}^{(\ell)} - \frac{\varepsilon}{2}\left(\begin{bmatrix}\mathbf{I}_D \\ \mathbf{0}^{\mathsf{T}}\end{bmatrix} - \tilde{\boldsymbol{\theta}}^{(\ell)}(\boldsymbol{\theta}^{(\ell)})^{\mathsf{T}}\right)\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}^{(\ell)})$$

$$\tilde{\boldsymbol{\theta}}^{(\ell+1)} = \tilde{\boldsymbol{\theta}}^{(\ell)}\cos(\|\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})}\|\varepsilon) + \frac{\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})}}{\|\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})}\|}\sin(\|\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})}\|\varepsilon)$$

$$\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})} \leftarrow -\tilde{\boldsymbol{\theta}}^{(\ell)}\|\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})}\|\sin(\|\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})}\|\varepsilon) + \tilde{\mathbf{v}}^{(\ell+\frac{1}{2})}\cos(\|\tilde{\mathbf{v}}^{(\ell+\frac{1}{2})}\|\varepsilon)$$

$$\tilde{\mathbf{v}}^{(\ell+1)} = \tilde{\mathbf{v}}^{(\ell+\frac{1}{2})} - \frac{\varepsilon}{2}\left(\begin{bmatrix}\mathbf{I}_D \\ \mathbf{0}^{\mathsf{T}}\end{bmatrix} - \tilde{\boldsymbol{\theta}}^{(\ell+1)}(\boldsymbol{\theta}^{(\ell+1)})^{\mathsf{T}}\right)\nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}^{(\ell+1)})$$

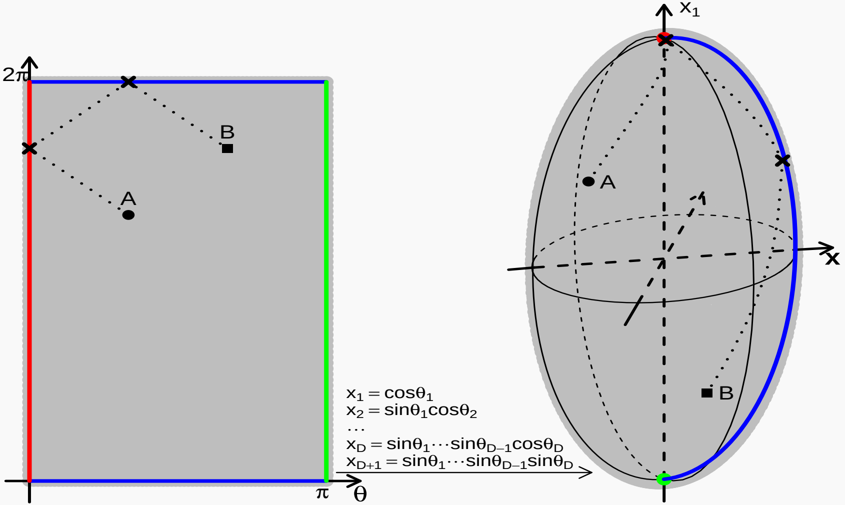
end for
Calculate $H(\tilde{\boldsymbol{\theta}}^{(L+1)}, \tilde{\mathbf{v}}^{(L+1)}) = U(\boldsymbol{\theta}^{(L+1)}) + K(\tilde{\mathbf{v}}^{(L+1)})$
Calculate the acceptance probability $\alpha = \min\{1, \exp[-H(\tilde{\boldsymbol{\theta}}^{(L+1)}, \tilde{\mathbf{v}}^{(L+1)}) + H(\tilde{\boldsymbol{\theta}}^{(1)}, \tilde{\mathbf{v}}^{(1)})]\}$
Accept or reject the proposal according to α for the next state $\tilde{\boldsymbol{\theta}}'$
Calculate $T_{\mathcal{S}\to\mathcal{D}}(\tilde{\boldsymbol{\theta}}')$ and the corresponding weight $|dT_{\mathcal{S}\to\mathcal{D}}|$

Spherical HMC

in the spherical coordinate

Box Type Constraints

Change of domain: from rectangle $\mathcal{R}_0^D$ to sphere $\mathcal{S}^D$



$\mathcal{R}_{\mathbf{0}}^D := [0,\pi]^{D-1} \times [0,2\pi)$

$\xrightarrow[x_d=\cos\theta_d\prod_{i=1}^{d-1}\sin\theta_i]{\boldsymbol{\theta}\mapsto\mathbf{x}}$

$\mathcal{S}^D := \{\mathbf{x} \in \mathbb{R}^{D+1} : \|\mathbf{x}\|_2 = 1\}$

Change of measure

$$\int_{\mathcal{R}_{\mathbf{0}}^D} f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{\mathcal{R}_{\mathbf{0}}} = \int_{\mathcal{S}^D} f(\boldsymbol{\theta}) \left| \frac{d\boldsymbol{\theta}_{\mathcal{R}_{\mathbf{0}}}}{d\boldsymbol{\theta}_{\mathcal{S}_r}} \right| d\boldsymbol{\theta}_{\mathcal{S}_r} = \int_{\mathcal{S}^D} f(\boldsymbol{\theta}) \prod_{d=1}^{D-1} \sin^{d-D}\theta_d d\boldsymbol{\theta}_{\mathcal{S}_r}$$

where $f(\boldsymbol{\theta}) = f(\boldsymbol{\theta}(\mathbf{x}))$ on $\mathcal{S}^D$.

What We Want:
$\boldsymbol{\theta} \sim f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{\mathcal{R}_{\mathbf{0}}}$

weigh sample $\boldsymbol{\theta}$ $\longleftarrow$ by $\prod_{d=1}^{D-1}\sin^{d-D}\theta_d$

What We Sample:
$\boldsymbol{\theta} \sim f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{\mathcal{S}_r}$

Round spherical metric

- Here, the natural metric on $\mathcal{S}^D$ is called *round spherical metric*:

Definition (round spherical metric)

$$\mathbf{G}_{\mathcal{S}_r}(\boldsymbol{\theta}) = \mathrm{diag}\left[1, \sin^2\theta_1, \cdots, \prod_{d=1}^{D-1}\sin^2\theta_d\right] \tag{14}$$

- For any vector $\mathbf{v} \in T_{\boldsymbol{\theta}}\mathcal{R}_{\mathbf{0}}^D$, we have

$$\mathbf{v}^{\mathsf{T}}\mathbf{G}_{\mathcal{S}_r}(\boldsymbol{\theta})\mathbf{v} \leq \|\mathbf{v}\|_2^2 \leq \|\tilde{\mathbf{v}}\|_2^2 = \mathbf{v}^{\mathsf{T}}\mathbf{G}_{\mathcal{S}_c}(\boldsymbol{\theta})\mathbf{v}$$

On $\mathcal{R}_{\mathbf{0}}^{D}$ | On $\mathcal{S}^D$

$$H(\boldsymbol{\theta},\mathbf{v}) = U(\boldsymbol{\theta}) + K(\mathbf{v}) = -\log f(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{v}^{\mathsf{T}}\mathbf{I}\mathbf{v} \quad \xrightarrow{\boldsymbol{\theta}\mapsto\mathbf{x}} \quad H^*(\mathbf{x},\dot{\mathbf{x}}) = U(\mathbf{x}) + K(\dot{\mathbf{x}}) = -\log f(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{v}^{\mathsf{T}}\mathbf{G}_{\mathcal{S}_r}(\boldsymbol{\theta})\mathbf{v}$$

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0},\mathbf{I}_D) \quad \xrightarrow{\mathbf{v}\mapsto\dot{\mathbf{x}}} \quad \mathbf{v}(\mathbf{x},\dot{\mathbf{x}}) \sim \mathbf{G}_{\mathcal{S}_r}(\boldsymbol{\theta})^{-\frac{1}{2}}\mathcal{N}(\mathbf{0},\mathbf{I}_D)$$

$$\begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{v}\\ \dot{\mathbf{v}} &= -\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta})\\ \mathbf{l} \le \boldsymbol{\theta} &\le \mathbf{u}\end{aligned} \quad \longrightarrow \quad \begin{aligned}\dot{\boldsymbol{\theta}} &= \mathbf{v}\\ \dot{\mathbf{v}} &= -\mathbf{v}^{\mathsf{T}}\boldsymbol{\Gamma}_{\mathcal{S}_r}(\boldsymbol{\theta})\mathbf{v} - \mathbf{G}_{\mathcal{S}_r}(\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta})\\ \boldsymbol{\theta} &= \boldsymbol{\theta}(\mathbf{x}), \quad \mathbf{v} = \mathbf{v}(\mathbf{x},\dot{\mathbf{x}})\end{aligned}$$

Split Lagrangian dynamics on sphere

in the spherical coordinate [5]

$$\dot{\boldsymbol\theta} = \mathbf{v}$$
$$\dot{\mathbf{v}} = -\mathbf{v}^{\mathsf T}\boldsymbol\Gamma_{\mathcal{S}_r}(\boldsymbol\theta)\mathbf{v} - \mathbf{G}_{\mathcal{S}_r}(\boldsymbol\theta)^{-1}\nabla_{\boldsymbol\theta}U(\boldsymbol\theta)$$

$$\dot{\boldsymbol\theta} = \mathbf{0}$$
$$\dot{\mathbf{v}} = -\frac{1}{2}\mathbf{G}_{\mathcal{S}_r}(\boldsymbol\theta)^{-1}\nabla_{\boldsymbol\theta}U(\boldsymbol\theta)$$

$\longleftrightarrow$

$$\dot{\boldsymbol\theta} = \mathbf{v}$$
$$\dot{\mathbf{v}} = -\mathbf{v}^{\mathsf T}\boldsymbol\Gamma_{\mathcal{S}_r}(\boldsymbol\theta)\mathbf{v}$$

$\downarrow$

$$\boldsymbol\theta(t) = \boldsymbol\theta(0)$$
$$\mathbf{v}(t) = \mathbf{v}(0) - \frac{t}{2}\cdot \operatorname{diag}\left[1,\cdots,\prod_{d=1}^{D-1}\sin^{-2}\theta_d\right]\nabla_{\boldsymbol\theta}U(\boldsymbol\theta(0))$$

$\downarrow$

$$(\boldsymbol\theta(0),\mathbf{v}(0)) \longrightarrow (\mathbf{x}(0),\dot{\mathbf{x}}(0))$$
$$\downarrow$$
$$(\mathbf{x}(t),\dot{\mathbf{x}}(t)) = g_r(\mathbf{x}(0),\dot{\mathbf{x}}(0))$$
$$\downarrow$$
$$(\boldsymbol\theta(t),\mathbf{v}(t)) \longleftarrow (\mathbf{x}(t),\dot{\mathbf{x}}(t))$$

Algorithm 2 Spherical HMC in the spherical coordinate (s-SphHMC)

Initialize $\boldsymbol{\theta}^{(1)}$ at current $\boldsymbol{\theta}$ after transformation $T_{\mathcal{D}\to\mathcal{S}}$
Sample a new velocity value $\mathbf{v}^{(1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$
Set $v_d^{(1)} \leftarrow v_d^{(1)} \prod_{i=1}^{d-1} \sin^{-1}(\theta_i^{(1)}),\ d = 1, \cdots, D$
Calculate $H(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)}) = U(\boldsymbol{\theta}^{(1)}) + K(\mathbf{v}^{(1)})$
for $\ell = 1$ to L **do**

$v_d^{(\ell+\frac{1}{2})} = v_d^{(\ell)} - \frac{\varepsilon^d}{2} \frac{\partial}{\partial \theta_d} U(\boldsymbol{\theta}^{(\ell)}) \prod_{i=1}^{d-1} \sin^{-2}(\theta_i^{(\ell)}),\ d = 1, \cdots, D$

$(\boldsymbol{\theta}^{(\ell+1)}, \mathbf{v}^{(\ell+\frac{1}{2})}) \leftarrow \tilde{T}_{\mathcal{S}\to\mathcal{R}_\mathbf{0}} \circ g_\varepsilon \circ \tilde{T}_{\mathcal{R}_\mathbf{0}\to\mathcal{S}}(\boldsymbol{\theta}^{(\ell)}, \mathbf{v}^{(\ell+\frac{1}{2})})$

$v_d^{(\ell+1)} = v_d^{(\ell+\frac{1}{2})} - \frac{\varepsilon^d}{2} \frac{\partial}{\partial \theta_d} U(\boldsymbol{\theta}^{(\ell+1)}) \prod_{i=1}^{d-1} \sin^{-2}(\theta_i^{(\ell+1)}),\ d = 1, \cdots, D$

end for
Calculate $H(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)}) = U(\boldsymbol{\theta}^{(L+1)}) + K(\mathbf{v}^{(L+1)})$
Calculate the acceptance probability $\alpha = \min\{1, \exp[-H(\boldsymbol{\theta}^{(L+1)}, \mathbf{v}^{(L+1)}) + H(\boldsymbol{\theta}^{(1)}, \mathbf{v}^{(1)})]\}$
Accept or reject the proposal according to α for the next state $\boldsymbol{\theta}'$
Calculate $T_{\mathcal{S}\to\mathcal{D}}(\boldsymbol{\theta}')$ and the corresponding weight $|dT_{\mathcal{S}\to\mathcal{D}}|$

The Geodesic Flow
in the Space of PD Matrices

- The metric tensor and inverse metric tensor are given by the $\frac{d}{2}(d+1) \times \frac{d}{2}(d+1)$ dimensional matrices

$$G(\Sigma) = D_d^{\mathsf{T}}(\Sigma^{-1} \otimes \Sigma^{-1})D_d \quad \text{and} \quad G^{-1}(\Sigma) = D_d^{+}(\Sigma \otimes \Sigma)D_d^{+T} \,. \tag{15}$$

- The geodesic flow under the foregoing metric in $\mathcal{S}_d^+$ is given [9]

$$\begin{aligned} \Sigma(t) &= \exp_{\Sigma} tV(0) = \Sigma(0)^{1/2*} \exp_{Id}(t\Sigma(0)^{-1/2*}V(0)) \\ &= \Sigma(0)^{1/2} \exp\left(t\Sigma(0)^{-1/2}V(0)\Sigma(0)^{-1/2}\right)\Sigma(0)^{1/2} \,. \end{aligned} \tag{16}$$

- The corresponding flow on the tangent bundle:

$$\begin{aligned} V(t) = \dot{\Sigma}(t) &= \frac{d}{dt}\exp_{\Sigma} tV(0) \\ &= V(0)\Sigma(0)^{-1/2} \exp\left(t\Sigma(0)^{-1/2}V(0)\Sigma(0)^{-1/2}\right)\Sigma(0)^{1/2} \,. \end{aligned} \tag{17}$$

PD-HMC
[4]

Algorithm 3 Hermitian PDHMC

$\mathrm{vech}(V) \sim \mathrm{CN}(0, G^{-1}(\Sigma_t))$
$e \leftarrow -\log\pi(\Sigma_t) - (d+1)\log|\Sigma_t| + \mathrm{vech}(V)^H G(\Sigma_t)\mathrm{vech}(V)$
$\Sigma^* \leftarrow \Sigma_t$
for $\tau = 1, \ldots, T$ **do**
  $\mathrm{vech}(V) \leftarrow \mathrm{vech}(V) + \frac{\varepsilon}{2} G^{-1}(\Sigma^*)\mathrm{vech}\Big(\nabla_\Sigma\big(\log\pi(\Sigma^*) + (d+1)\log|\Sigma^*|\big)\Big)$
  Progress (Σ^*, V) along the geodesic flow for time ε.
  $\mathrm{vech}(V) \leftarrow \mathrm{vech}(V) + \frac{\varepsilon}{2} G^{-1}(\Sigma^*)\mathrm{vech}\Big(\nabla_\Sigma\big(\log\pi(\Sigma^*) + (d+1)\log|\Sigma^*|\big)\Big)$
end for
$e^* \leftarrow -\log\pi(\Sigma^*) - (d+1)\log|\Sigma^*| + \mathrm{vech}(V)^H G(\Sigma^*)\mathrm{vech}(V)$
$u \sim U(0,1)$
if $u < \exp(e - e^*)$ **then**
  $\Sigma_{t+1} \leftarrow \Sigma^*$
end if