Understanding MCMC Dynamics as Flows on the Wasserstein Space

Chang Liu, Jingwei Zhuo, Jun Zhu
chang-li14@mails.tsinghua.edu.cn

Introduction

Known before: Langevin Dynamics (LD) = gradient flow of KL on the Wasserstein space (Jordan *et al.*, 1998).

This work: general MCMC = fGH flow of KL on the Wasserstein space of an fRP manifold,

- so the behavior of a general MCMC is clear;
- so more MCMCs could inspire more ParVIs.

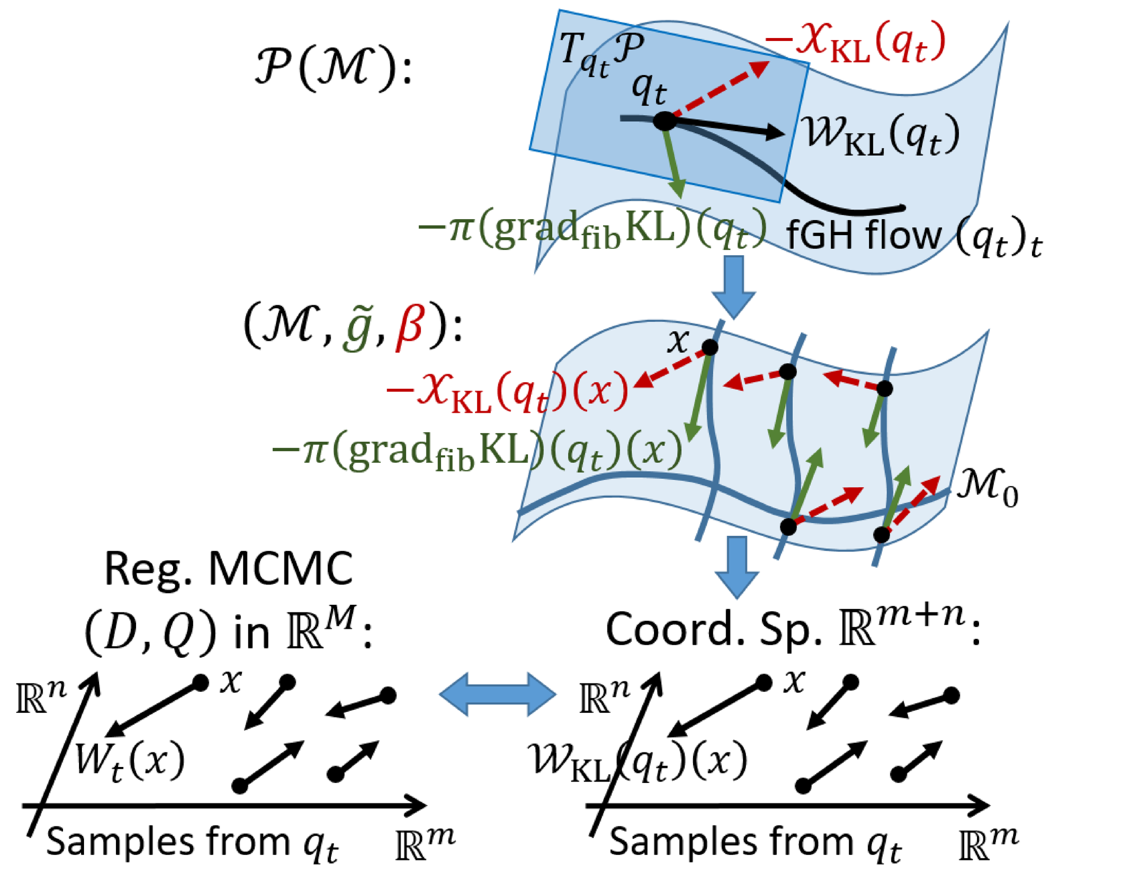
MCMC as fGH Flow

Theorem 5. (The unified framework) Fiber-Riemannian Poisson (fRP) manifold: $(\mathcal{M}, \tilde{g}, \beta)$. Fiber-Grad. Hamiltonian (fGH) flow on $\mathcal{P}(\mathcal{M})$:

$$\mathcal{W}_{\mathrm{KL}_p} := -\pi(\mathrm{grad}_{\mathrm{fib}}\,\mathrm{KL}_p) - \mathcal{X}_{\mathrm{KL}_p},$$

$$\big(\mathcal{W}_{\mathrm{KL}_p}(q)\big)^i = \pi_q\big((\tilde{g}^{ij} + \beta^{ij})\partial_j \log(p/q)\big).$$

Then: Regular MCMC dynamics $\iff$ fGH flow with fRP $\mathcal{M}$, $(D, Q) \iff (\tilde{g}, \beta)$.



Simulation as ParVIs

Deterministic dynamics of SGHMC:

By Lemma 1 (pSGHMC-det):

$$\begin{cases} \frac{\mathrm{d}\theta}{\mathrm{d}t} = \Sigma^{-1} r, \\ \frac{\mathrm{d}r}{\mathrm{d}t} = \nabla_\theta \log p(\theta) - C\Sigma^{-1} r - C\nabla_r \log q(r). \end{cases}$$

By Theorem 5 (pSGHMC-fGH):

$$\begin{cases} \frac{\mathrm{d}\theta}{\mathrm{d}t} = \Sigma^{-1} r + \nabla_r \log q(r), \\ \frac{\mathrm{d}r}{\mathrm{d}t} = \nabla_\theta \log p(\theta) - C\Sigma^{-1} r - C\nabla_r \log q(r) - \nabla_\theta \log q(\theta). \end{cases}$$

Estimate $\nabla \log q$ using ParVI techniques, *e.g.*, Blob (Chen *et al.*, 2018).

Preliminary

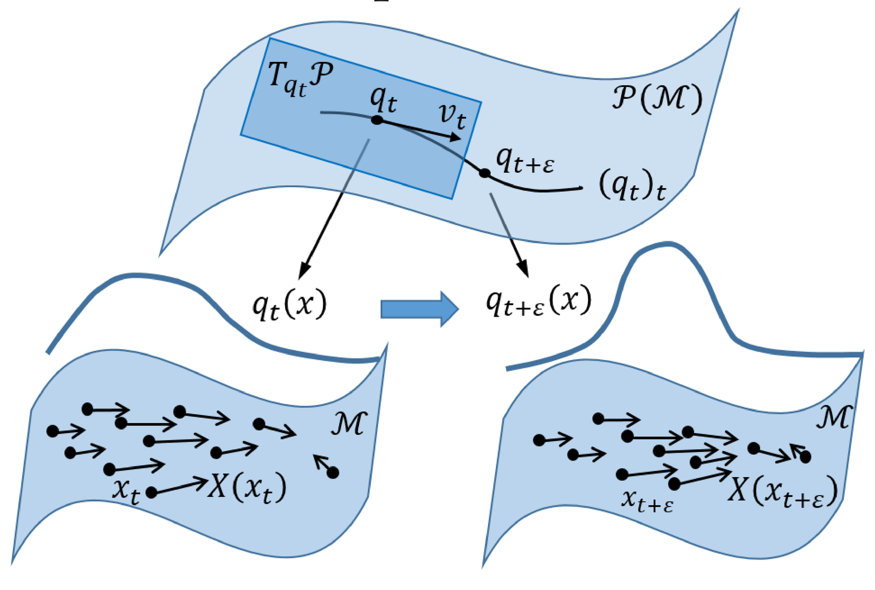
- The complete recipe of general MCMC dynamics on $\mathbb{R}^M$ (Ma *et al.*, 2015)

$$\mathrm{d}x = V(x)\,\mathrm{d}t + \sqrt{2D(x)}\,\mathrm{d}B_t(x),$$

$$V^i(x) = \frac{1}{p(x)}\partial_j\Big(p(x)\big(D^{ij}(x) + Q^{ij}(x)\big)\Big), \quad (1)$$

$D_{M\times M}$: pos. semi-def.; $Q_{M\times M}$: skew-symm.

- Wasserstein space $\mathcal{P}(\mathcal{M})$



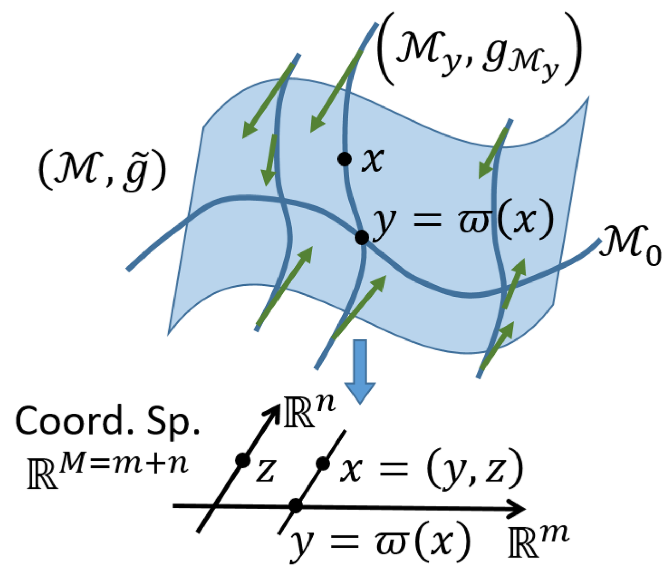
Tangent vector $v \Leftrightarrow$ vector field X on $\mathcal{M}$.

- Gradient flow on $\mathcal{P}(\mathcal{M})$
  - Riemannian structure: $\langle X, Y\rangle_{T_q\mathcal{P}} = \mathbb{E}_{q(x)}[\langle X(x), Y(x)\rangle_{T_x\mathcal{M}}]$, $\langle u, v\rangle_{T_x\mathcal{M}} = g_{ij}(x)u^i v^j$, (g_{ij}): pos. def.
  - Ortho. proj. $\pi_q : \mathcal{L}_q^2 \to T_q\mathcal{P}$ preserves distribution evolution.
  - $\mathrm{grad}_{\mathcal{P}(\mathcal{M})}\mathrm{KL}_p(q) = \mathrm{grad}_{\mathcal{M}} \log(q/p) = g^{ij}\partial_j \log(q/p)\partial_i$.
- Hamiltonian flow on $\mathcal{P}(\mathcal{M})$
  - Poisson structure: $\{F_f, F_h\}_{\mathcal{P}(\mathcal{M})} := F_{\{f,h\}_{\mathcal{M}}}$, $\{f, h\}_{\mathcal{M}} = \beta(\mathrm{d}f, \mathrm{d}h) = \beta^{ij}\partial_i f \partial_j h$, $F_f[q] := \mathbb{E}_q[f]$. (β^{ij}): skew-symm., Jacobi identity.
  - Hamiltonian flow: $\mathcal{X}_F(q) = \pi_q(X_f) = \pi_q(\beta^{ij}\partial_j f \partial_i)$.

Technical Developments

Lemma 1. MCMC dynamics Eq. (1) with symm. D is equiv. to $\mathrm{d}x = W_t(x)\mathrm{d}t$, $(W_t)^i(x) = D^{ij}(x)\,\partial_j \log(p(x)/q_t(x)) + Q^{ij}(x)\,\partial_j \log p(x) + \partial_j Q^{ij}(x)$.

Lemma 2. $\mathcal{X}_{\mathrm{KL}_p}(q) = \pi_q(X_{\log(q/p)})$, where $\big(X_{\log(q/p)}(x)\big)^i = \beta^{ij}(x)\,\partial_j \log(q(x)/p(x))$.



Def. 3. *Fiber-Riemannian manifold*: a fiber bundle with a Riemannian structure $g_{\mathcal{M}_y}$ on each *fiber* $\mathcal{M}_y$.

- Fiber-gradient $\big(\mathrm{grad}_{\mathrm{fib}}\, f(x)\big)^i = \tilde{g}^{ij}(x)\,\partial_j f(x), \quad 1 \le i, j \le M,$

$$(\tilde{g}^{ij}(x)) := \begin{pmatrix} 0_{m\times m} & 0_{m\times n} \\ 0_{n\times m} & \big((g_{\mathcal{M}_{\varpi(x)}}(z))^{ab}\big)_{n\times n} \end{pmatrix}.$$

- Fiber-gradient on $\mathcal{P}(\mathcal{M})$: $\big(\mathrm{grad}_{\mathrm{fib}}\,\mathrm{KL}_p(q)(x)\big)_{\mathcal{M}} = \big(\tilde{g}^{ij}(x)\,\partial_j \log(q(x)/p(x))\big)_{\mathcal{M}}$.

MCMCs under the Framework

Type 1: D is non-singular ($m = 0$).

- $\mathcal{M}_0$ degenerates, $\mathcal{M}$ is the unique fiber.
- fGH flow = grad. flow + Ham. flow,
  grad. flow: min. KL_p on $\mathcal{P}(\mathcal{M})$.
  Ham. flow: conserves KL_p on $\mathcal{P}(\mathcal{M})$.
- Robust to stochastic gradient (SG).

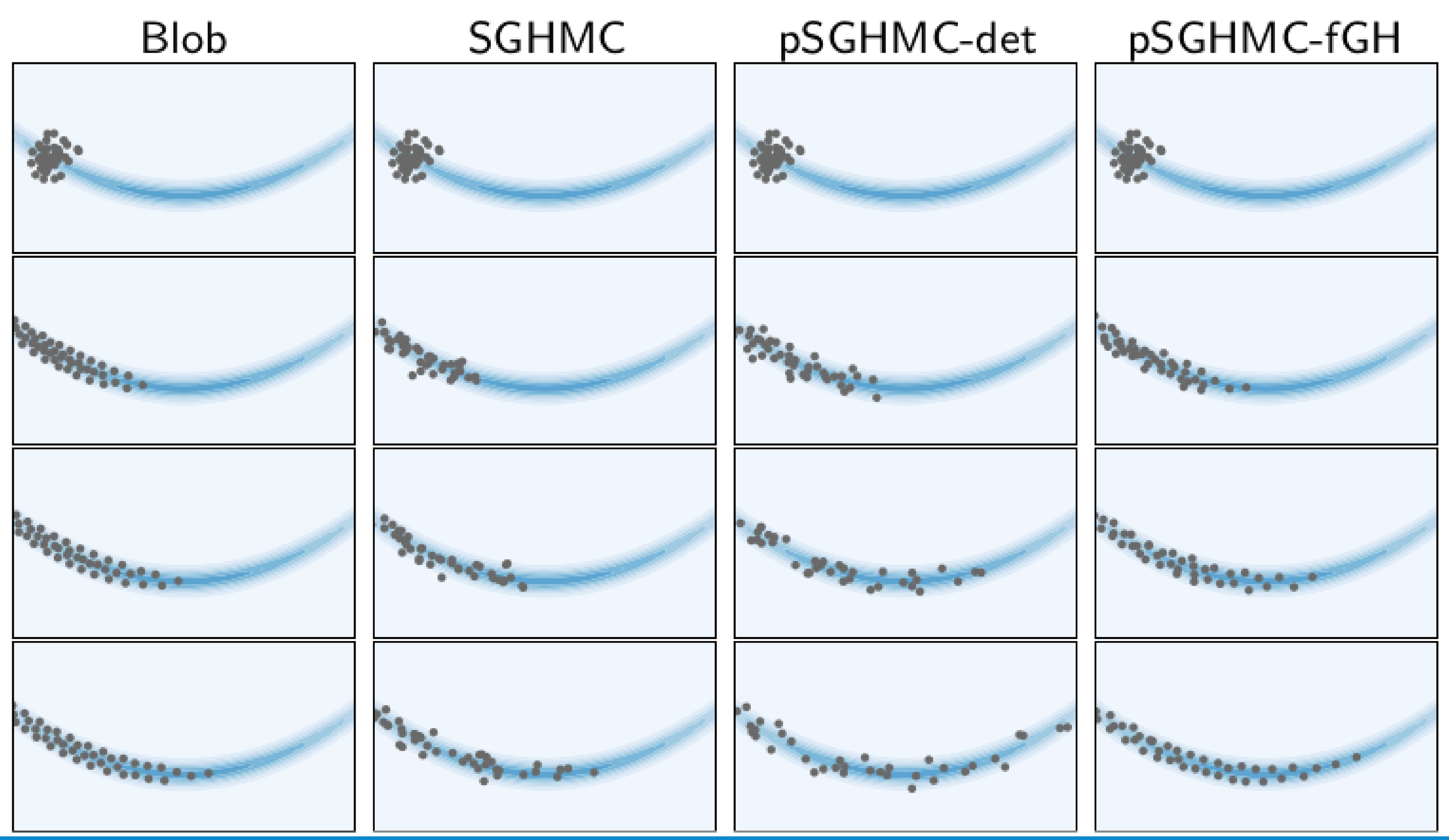
Type 2: $D = 0$ ($n = 0$).

- $\mathcal{M}_0 = \mathcal{M}$, fibers degenerate.
- fGH flow = Ham. flow.
- Fragile against SG: no stablizing forces (fiber-gradient flows)

Type 3: $D \neq 0$ and D is singular ($m, n \ge 1$).

- Non-degenerate $\mathcal{M}_0$ and $\mathcal{M}_y$.
- fGH = fib. grad. + Ham.,
  fib. grad.: min. $\mathrm{KL}_{p(\cdot|y)}(q(\cdot|y))$ on each fiber $\mathcal{P}(\mathcal{M}_y)$.
  Ham.: conserves KL_p on $\mathcal{P}(\mathcal{M})$ and helps mixing/exploration.
- Robust to SG (SG appears on each fiber).

Experiments



- Latent Dirichlet Allocation