MCMC methods derived from Diffusions, Geodesics and Foliations on Riemannian manifolds

Mark Girolami

Department of Statistical Science
University College London

CSML Lunchtime Meetings
April, 2012
MCMC from Diffusions, Geodesics and Foliations

- Riemann manifold Langevin and Hamiltonian Monte Carlo Methods
  Girolami, M. & Calderhead, B.

[Link to website](www.ucl.ac.uk/statistics/research/rmhmc)
Motivation to improve MCMC capability for challenging problems
Talk Outline

- Motivation to improve MCMC capability for challenging problems
- Exploring geometric concepts in MCMC methodology
Talk Outline

- Motivation to improve MCMC capability for challenging problems
- Exploring geometric concepts in MCMC methodology
- Diffusions across Riemann manifold as proposal mechanism
Talk Outline

- Motivation to improve MCMC capability for challenging problems
- Exploring geometric concepts in MCMC methodology
- Diffusions across Riemann manifold as proposal mechanism
- Geodesic flows on manifold form basis of MCMC methods
Talk Outline

• Motivation to improve MCMC capability for challenging problems
• Exploring geometric concepts in MCMC methodology
• Diffusions across Riemann manifold as proposal mechanism
• Geodesic flows on manifold form basis of MCMC methods
• Hierarchical Bayesian models - Introducing Foliations
Talk Outline

- Motivation to improve MCMC capability for challenging problems
- Exploring geometric concepts in MCMC methodology
- Diffusions across Riemann manifold as proposal mechanism
- Geodesic flows on manifold form basis of MCMC methods
- Hierarchical Bayesian models - Introducing Foliations
- Nonlinear dynamic systems - deterministic and stochastic
Talk Outline

- Motivation to improve MCMC capability for challenging problems
- Exploring geometric concepts in MCMC methodology
- Diffusions across Riemann manifold as proposal mechanism
- Geodesic flows on manifold form basis of MCMC methods
- Hierarchical Bayesian models - Introducing Foliations
- Nonlinear dynamic systems - deterministic and stochastic
- Further Work and Conclusions
Motivation Simulation Based Inference

- Monte Carlo method employs samples from $\pi(\theta)$ to obtain estimate

$$\int \phi(\theta)\pi(\theta)d\theta = \frac{1}{N} \sum_n \phi(\theta^n) + O(N^{-\frac{1}{2}})$$
Motivation Simulation Based Inference

- Monte Carlo method employs samples from $\pi(\theta)$ to obtain estimate

$$\int \phi(\theta)\pi(\theta)d\theta = \frac{1}{N} \sum_n \phi(\theta^n) + \mathcal{O}(N^{-\frac{1}{2}})$$

- Draw $\theta^n$ from ergodic Markov process with stationary distribution $\pi(\theta)$
Motivation Simulation Based Inference

- Monte Carlo method employs samples from $\pi(\theta)$ to obtain estimate

$$\int \phi(\theta) \pi(\theta) d\theta = \frac{1}{N} \sum_{n} \phi(\theta^n) + O(N^{-\frac{1}{2}})$$

- Draw $\theta^n$ from ergodic Markov process with stationary distribution $\pi(\theta)$
- Construct transition kernel as product of two components

Success of MCMC reliant upon appropriate proposal design
Motivation Simulation Based Inference

- Monte Carlo method employs samples from $\pi(\theta)$ to obtain estimate

$$\int \phi(\theta)\pi(\theta)d\theta = \frac{1}{N} \sum_n \phi(\theta^n) + O(N^{-\frac{1}{2}})$$

- Draw $\theta^n$ from ergodic Markov process with stationary distribution $\pi(\theta)$

- Construct transition kernel as product of two components
  - Propose a move $\theta \rightarrow \theta'$ with probability $p_p(\theta'|\theta)$

- Convergence rate and asymptotic variance dependent on $p_p(\theta'|\theta)$

- Success of MCMC reliant upon appropriate proposal design
Motivation Simulation Based Inference

- Monte Carlo method employs samples from $\pi(\theta)$ to obtain estimate

$$\int \phi(\theta)\pi(\theta)d\theta = \frac{1}{N} \sum_n \phi(\theta^n) + O(N^{-\frac{1}{2}})$$

- Draw $\theta^n$ from ergodic Markov process with stationary distribution $\pi(\theta)$

- Construct transition kernel as product of two components
  - Propose a move $\theta \rightarrow \theta'$ with probability $p_p(\theta'|\theta)$
  - accept or reject proposal with probability

$$p_a(\theta'|\theta) = \min\left[1, \frac{\pi(\theta')p_p(\theta'|\theta')}{\pi(\theta)p_p(\theta'|\theta)}\right]$$
Motivation Simulation Based Inference

- Monte Carlo method employs samples from $\pi(\theta)$ to obtain estimate

$$\int \phi(\theta) \pi(\theta) d\theta = \frac{1}{N} \sum_n \phi(\theta^n) + \mathcal{O}(N^{-\frac{1}{2}})$$

- Draw $\theta^n$ from ergodic Markov process with stationary distribution $\pi(\theta)$

- Construct transition kernel as product of two components
  - Propose a move $\theta \rightarrow \theta'$ with probability $p_p(\theta'|\theta)$
  - accept or reject proposal with probability

$$p_a(\theta'|\theta) = \min \left[ 1, \frac{\pi(\theta') p_p(\theta'|\theta)}{\pi(\theta) p_p(\theta'|\theta)} \right]$$

- Convergence rate and asymptotic variance dependent on $p_p(\theta'|\theta)$
Motivation Simulation Based Inference

- Monte Carlo method employs samples from $\pi(\theta)$ to obtain estimate
  \[ \int \phi(\theta)\pi(\theta)d\theta = \frac{1}{N} \sum_n \phi(\theta^n) + O(N^{-\frac{1}{2}}) \]

- Draw $\theta^n$ from ergodic Markov process with stationary distribution $\pi(\theta)$

- Construct transition kernel as product of two components
  - Propose a move $\theta \rightarrow \theta'$ with probability $p_p(\theta'|\theta)$
  - accept or reject proposal with probability
    \[ p_a(\theta'|\theta) = \min \left[ 1, \frac{\pi(\theta')p_p(\theta'|\theta)}{\pi(\theta)p_p(\theta'|\theta)} \right] \]

- Convergence rate and asymptotic variance dependent on $p_p(\theta'|\theta)$

- Success of MCMC reliant upon appropriate proposal design
Geometric Concepts in MCMC

• Denote expected Fisher Information as $G(\theta) = \text{cov}(\nabla_\theta L(\theta))$

• Rao, 1945 to first order

$$\chi^2(\delta \theta) = \int |p(y; \theta + \delta \theta) - p(y; \theta)|^2 p(y; \theta) \, dy \approx \delta \theta^T G(\theta) \delta \theta$$

• Jeffreys, 1948 to first order

$$D(\theta||\delta \theta) = \int p(y; \theta + \delta \theta) \log p(y; \theta + \delta \theta) p(y; \theta) \, dy \approx \delta \theta^T G(\theta) \delta \theta$$

• Expected Fisher Information $G(\theta)$ is metric tensor of a Riemann manifold

• Non-Euclidean geometry - invariants, connections, curvature, geodesics

• Asymptotic statistical analysis. e.g. Amari, 1981; Murray & Rice, 1993; Critchley et al., 1993; Kass, 1989; Dawid, 1975; Lauritsen, 1989

• Statistical shape analysis Kent et al., 1996; Dryden & Mardia, 1998

• Can geometric structure be employed in Monte Carlo methodology?
Geometric Concepts in MCMC

- Denote expected Fisher Information as $G(\theta) = \text{cov}(\nabla_\theta \mathcal{L}(\theta))$

- Asymptotic statistical analysis. e.g. Amari, 1981; Murray & Rice, 1993; Critchley et al., 1993; Kass, 1989; Dawid, 1975; Lauritsen, 1989

- Statistical shape analysis Kent et al., 1996; Dryden & Mardia, 1998

- Can geometric structure be employed in Monte Carlo methodology?
Geometric Concepts in MCMC

- Denote expected Fisher Information as $G(\theta) = \text{cov}(\nabla_\theta L(\theta))$
- Rao, 1945 to first order

$$\chi^2(\delta\theta) = \int \frac{|p(y; \theta + \delta\theta) - p(y; \theta)|^2}{p(y; \theta)} \, dy \approx \delta\theta^T G(\theta) \delta\theta$$
Geometric Concepts in MCMC

• Denote expected Fisher Information as $G(\theta) = \text{cov}(\nabla_\theta \mathcal{L}(\theta))$

• Rao, 1945 to first order

$$\chi^2(\delta \theta) = \int \frac{|p(y; \theta + \delta \theta) - p(y; \theta)|^2}{p(y; \theta)} dy \approx \delta \theta^T G(\theta) \delta \theta$$

• Jeffreys, 1948 to first order

$$D(\theta || \delta \theta) = \int p(y; \theta + \delta \theta) \log \frac{p(y; \theta + \delta \theta)}{p(y; \theta)} dy \approx \delta \theta^T G(\theta) \delta \theta$$
Geometric Concepts in MCMC

• Denote expected Fisher Information as $G(\theta) = \text{cov}(\nabla \theta \mathcal{L}(\theta))$

• Rao, 1945 to first order

$$\chi^2(\delta \theta) = \int \frac{|p(y; \theta + \delta \theta) - p(y; \theta)|^2}{p(y; \theta)} \, dy \approx \delta \theta^T G(\theta) \delta \theta$$

• Jeffreys, 1948 to first order

$$D(\theta || \delta \theta) = \int p(y; \theta + \delta \theta) \log \frac{p(y; \theta + \delta \theta)}{p(y; \theta)} \, dy \approx \delta \theta^T G(\theta) \delta \theta$$

• Expected Fisher Information $G(\theta)$ is metric tensor of a Riemann manifold
Geometric Concepts in MCMC

- Denote expected Fisher Information as \( G(\theta) = \text{cov}(\nabla_{\theta} \mathcal{L}(\theta)) \)
- Rao, 1945 to first order
  \[
  \chi^2(\delta \theta) = \int \frac{|p(y; \theta + \delta \theta) - p(y; \theta)|^2}{p(y; \theta)} \, dy \approx \delta \theta^T G(\theta) \delta \theta
  \]
- Jeffreys, 1948 to first order
  \[
  D(\theta \| \delta \theta) = \int p(y; \theta + \delta \theta) \log \frac{p(y; \theta + \delta \theta)}{p(y; \theta)} \, dy \approx \delta \theta^T G(\theta) \delta \theta
  \]
- Expected Fisher Information \( G(\theta) \) is metric tensor of a Riemann manifold
- Non-Euclidean geometry - invariants, connections, curvature, geodesics
Geometric Concepts in MCMC

• Denote expected Fisher Information as \( \mathbf{G}(\theta) = \text{cov}(\nabla_\theta \mathcal{L}(\theta)) \)

• Rao, 1945 to first order

\[
\chi^2(\delta \theta) = \int \frac{|p(y; \theta + \delta \theta) - p(y; \theta)|^2}{p(y; \theta)} dy \approx \delta \theta^T \mathbf{G}(\theta) \delta \theta
\]

• Jeffreys, 1948 to first order

\[
D(\theta || \delta \theta) = \int p(y; \theta + \delta \theta) \log \frac{p(y; \theta + \delta \theta)}{p(y; \theta)} dy \approx \delta \theta^T \mathbf{G}(\theta) \delta \theta
\]

• Expected Fisher Information \( \mathbf{G}(\theta) \) is metric tensor of a Riemann manifold

• Non-Euclidean geometry - invariants, connections, curvature, geodesics

• Asymptotic statistical analysis. e.g. Amari, 1981; Murray & Rice, 1993; Critchley et al, 1993; Kass, 1989; Dawid, 1975; Lauritsen, 1989
Geometric Concepts in MCMC

- Denote expected Fisher Information as $G(\theta) = \text{cov}(\nabla_\theta \mathcal{L}(\theta))$
- Rao, 1945 to first order
  \[
  \chi^2(\delta \theta) = \int \frac{|p(y; \theta + \delta \theta) - p(y; \theta)|^2}{p(y; \theta)} \, dy \approx \delta \theta^T G(\theta) \delta \theta
  \]
- Jeffreys, 1948 to first order
  \[
  D(\theta || \delta \theta) = \int p(y; \theta + \delta \theta) \log \frac{p(y; \theta + \delta \theta)}{p(y; \theta)} \, dy \approx \delta \theta^T G(\theta) \delta \theta
  \]
- Expected Fisher Information $G(\theta)$ is metric tensor of a Riemann manifold
- Non-Euclidean geometry - invariants, connections, curvature, geodesics
- Asymptotic statistical analysis. e.g. Amari, 1981; Murray & Rice, 1993; Critchley et al, 1993; Kass, 1989; Dawid, 1975; Lauritsen, 1989
- Statistical shape analysis Kent et al, 1996; Dryden & Mardia, 1998
Geometric Concepts in MCMC

- Denote expected Fisher Information as $G(\theta) = \text{cov} (\nabla_{\theta} L(\theta))$
- Rao, 1945 to first order
  $$\chi^2(\delta \theta) = \int \frac{|p(y; \theta + \delta \theta) - p(y; \theta)|^2}{p(y; \theta)} dy \approx \delta \theta^T G(\theta) \delta \theta$$
- Jeffreys, 1948 to first order
  $$D(\theta || \delta \theta) = \int p(y; \theta + \delta \theta) \log \left( \frac{p(y; \theta + \delta \theta)}{p(y; \theta)} \right) dy \approx \delta \theta^T G(\theta) \delta \theta$$
- Expected Fisher Information $G(\theta)$ is metric tensor of a Riemann manifold
- Non-Euclidean geometry - invariants, connections, curvature, geodesics
- Asymptotic statistical analysis. e.g. Amari, 1981; Murray & Rice, 1993; Critchley et al, 1993; Kass, 1989; Dawid, 1975; Lauritsen, 1989
- Statistical shape analysis Kent et al, 1996; Dryden & Mardia, 1998
- Can geometric structure be employed in Monte Carlo methodology?
Geometric Concepts in MCMC

- **Tangent space** - local metric defined by $\delta_{G}^{T}(\theta) \delta_{\theta} = g_{kl}\delta_{\theta}^{k}\delta_{\theta}^{l}$

- **Christoffel symbols** - characterise Levi-Civita connection on manifold

\[ \Gamma^{i}_{kl} = \frac{1}{2} g^{im} \left( \partial g_{mk} \partial \theta^{l} + \partial g_{ml} \partial \theta^{k} - \partial g_{kl} \partial \theta^{m} \right) \]

- **Geodesics** - shortest path between two points on manifold

\[ d_{\theta}^{2} dt^{2} + \sum_{k, l} \Gamma^{i}_{kl} d \theta^{k} d \theta^{l} = 0 \]
Geometric Concepts in MCMC

• Tangent space - local metric defined by \( \delta \theta^T G(\theta) \delta \theta = g_{kl} \delta \theta_k \delta \theta_l \)
Geometric Concepts in MCMC

- Tangent space - local metric defined by $\delta \theta^T G(\theta) \delta \theta = g_{kl} \delta \theta_k \delta \theta_l$
- Christoffel symbols - characterise Levi-Civita connection on manifold
Geometric Concepts in MCMC

- Tangent space - local metric defined by \( \delta \theta^T G(\theta) \delta \theta = g_{kl} \delta \theta_k \delta \theta_l \)
- Christoffel symbols - characterise Levi-Civita connection on manifold

\[
\Gamma^i_{kl} = \frac{1}{2} \sum_m g^{im} \left( \frac{\partial g_{mk}}{\partial \theta^l} + \frac{\partial g_{ml}}{\partial \theta^k} - \frac{\partial g_{kl}}{\partial \theta^m} \right)
\]
Geometric Concepts in MCMC

- Tangent space - local metric defined by $\delta \theta^T G(\theta) \delta \theta = g_{kl} \delta \theta_k \delta \theta_l$
- Christoffel symbols - characterise Levi-Civita connection on manifold
  
  $$\Gamma^i_{kl} = \frac{1}{2} \sum_m g^{im} \left( \frac{\partial g_{mk}}{\partial \theta^l} + \frac{\partial g_{ml}}{\partial \theta^k} - \frac{\partial g_{kl}}{\partial \theta^m} \right)$$

- Geodesics - shortest path between two points on manifold
Geometric Concepts in MCMC

- Tangent space - local metric defined by $\delta \theta^T G(\theta) \delta \theta = g_{kl} \delta \theta_k \delta \theta_l$
- Christoffel symbols - characterise Levi-Civita connection on manifold

\[
\Gamma^i_{kl} = \frac{1}{2} \sum_m g^{im} \left( \frac{\partial g_{mk}}{\partial \theta^l} + \frac{\partial g_{ml}}{\partial \theta^k} - \frac{\partial g_{kl}}{\partial \theta^m} \right)
\]

- Geodesics - shortest path between two points on manifold

\[
\frac{d^2 \theta^i}{dt^2} + \sum_{k,l} \Gamma^i_{kl} \frac{d \theta^k}{dt} \frac{d \theta^l}{dt} = 0
\]
Illustration of Geometric Concepts

- Consider Normal density $p(x|\mu, \sigma) = \mathcal{N}_x(\mu, \sigma)$
Illustration of Geometric Concepts

- Consider Normal density \( p(x|\mu, \sigma) = \mathcal{N}_x(\mu, \sigma) \)

- Local inner product on tangent space defined by metric tensor, i.e. \( \delta \theta^T G(\theta) \delta \theta \), where \( \theta = (\mu, \sigma)^T \)
Illustration of Geometric Concepts

- Consider Normal density \( p(x|\mu, \sigma) = \mathcal{N}_x(\mu, \sigma) \)

- Local inner product on tangent space defined by metric tensor, i.e. \( \delta \theta^T G(\theta) \delta \theta \), where \( \theta = (\mu, \sigma)^T \)

- Metric is Expected Fisher Information

\[
G(\mu, \sigma) = \begin{bmatrix}
\sigma^{-2} & 0 \\
0 & 2\sigma^{-2}
\end{bmatrix}
\]
Illustration of Geometric Concepts

• Consider Normal density \( p(x|\mu, \sigma) = N_x(\mu, \sigma) \)

• Local inner product on tangent space defined by metric tensor, i.e.
  \( \delta \theta^T G(\theta) \delta \theta \), where \( \theta = (\mu, \sigma)^T \)

• Metric is Expected Fisher Information
  \[
  G(\mu, \sigma) = \begin{bmatrix}
  \sigma^{-2} & 0 \\
  0 & 2\sigma^{-2}
  \end{bmatrix}
  \]

• Components of connection \( \partial_\mu G = 0 \) and \( \partial_\sigma G = -\text{diag}(2\sigma^{-3}, 4\sigma^{-3}) \)
Illustration of Geometric Concepts

- Consider Normal density $p(x|\mu, \sigma) = \mathcal{N}_x(\mu, \sigma)$

- Local inner product on tangent space defined by metric tensor, i.e. $\delta \theta^\top G(\theta) \delta \theta$, where $\theta = (\mu, \sigma)^\top$

- Metric is Expected Fisher Information

$$G(\mu, \sigma) = \begin{bmatrix} \sigma^{-2} & 0 \\ 0 & 2\sigma^{-2} \end{bmatrix}$$

- Components of connection $\partial_\mu G = 0$ and $\partial_\sigma G = -\text{diag}(2\sigma^{-3}, 4\sigma^{-3})$

- Metric on tangent space

$$\delta \theta^\top G(\theta) \delta \theta = \frac{(\delta \mu^2 + 2\delta \sigma^2)}{\sigma^2}$$
Illustration of Geometric Concepts

- Consider Normal density \( p(x|\mu, \sigma) = \mathcal{N}_x(\mu, \sigma) \)

- Local inner product on tangent space defined by metric tensor, i.e. \( \delta \theta^T G(\theta) \delta \theta \), where \( \theta = (\mu, \sigma)^T \)

- Metric is Expected Fisher Information

\[
G(\mu, \sigma) = \begin{bmatrix}
\sigma^{-2} & 0 \\
0 & 2\sigma^{-2}
\end{bmatrix}
\]

- Components of connection \( \partial_\mu G = 0 \) and \( \partial_\sigma G = - \text{diag}(2\sigma^{-3}, 4\sigma^{-3}) \)

- Metric on tangent space

\[
\delta \theta^T G(\theta) \delta \theta = \frac{(\delta \mu^2 + 2\delta \sigma^2)}{\sigma^2}
\]

- Metric tensor for univariate Normal defines a Hyperbolic Space
Illustration of Geometric Concepts

- Consider Normal density $p(x|\mu, \sigma) = \mathcal{N}_x(\mu, \sigma)$

- Local inner product on tangent space defined by metric tensor, i.e.
  $\delta \theta^T G(\theta) \delta \theta$, where $\theta = (\mu, \sigma)^T$

- Metric is Expected Fisher Information
  \[
  G(\mu, \sigma) = \begin{bmatrix}
  \sigma^{-2} & 0 \\
  0 & 2\sigma^{-2}
  \end{bmatrix}
  \]

- Components of connection $\partial_\mu G = 0$ and $\partial_\sigma G = -\text{diag}(2\sigma^{-3}, 4\sigma^{-3})$

- Metric on tangent space
  \[
  \delta \theta^T G(\theta) \delta \theta = \frac{(\delta \mu^2 + 2\delta \sigma^2)}{\sigma^2}
  \]

- Metric tensor for univariate Normal defines a Hyperbolic Space

- Consider densities $\mathcal{N}(0, 1) \& \mathcal{N}(1, 1)$ and $\mathcal{N}(0, 2) \& \mathcal{N}(1, 2)$
Normal Density - Euclidean Parameter space

\[ \mathcal{N}(0, 2) \quad \mathcal{N}(1, 2) \]

\[ \mathcal{N}(0, 1) \quad \mathcal{N}(1, 1) \]
Normal Density - Riemannian Functional space
M.C. Escher, Heaven and Hell, 1960
Langevin Diffusion on Riemannian manifold

- Discretised Langevin diffusion on manifold defines proposal mechanism

\[
\theta'_d = \theta_d + \frac{\epsilon^2}{2} \left( G^{-1}(\theta) \nabla_{\theta} L(\theta) \right)_d - \epsilon^2 \sum_{i,j}^D G(\theta)_{ij}^{-1} \Gamma_{ij}^d + \epsilon \left( \sqrt{G^{-1}(\theta)} z \right)_d
\]
Langevin Diffusion on Riemannian manifold

- Discretised Langevin diffusion on manifold defines proposal mechanism

\[ \theta'_d = \theta_d + \frac{\epsilon^2}{2} \left( G^{-1}(\theta) \nabla_\theta \mathcal{L}(\theta) \right)_d - \epsilon^2 \sum_{i,j}^D G(\theta)_{ij}^{-1} \Gamma^d_{ij} + \epsilon \left( \sqrt{G^{-1}(\theta)} z \right)_d \]

- Manifold with constant curvature then proposal mechanism reduces to

\[ \theta' = \theta + \frac{\epsilon^2}{2} G^{-1}(\theta) \nabla_\theta \mathcal{L}(\theta) + \epsilon \sqrt{G^{-1}(\theta)} z \]
Langevin Diffusion on Riemannian manifold

• Discretised Langevin diffusion on manifold defines proposal mechanism

\[ \theta'_d = \theta_d + \frac{\epsilon^2}{2} \left( G^{-1}(\theta) \nabla_\theta \mathcal{L}(\theta) \right)_d - \epsilon^2 \sum_{i,j}^D G(\theta)^{-1} \Gamma_{ij}^d + \epsilon \left( \sqrt{G^{-1}(\theta)} z \right)_d \]

• Manifold with constant curvature then proposal mechanism reduces to

\[ \theta' = \theta + \frac{\epsilon^2}{2} G^{-1}(\theta) \nabla_\theta \mathcal{L}(\theta) + \epsilon \sqrt{G^{-1}(\theta)} z \]

• MALA proposal with preconditioning

\[ \theta' = \theta + \frac{\epsilon^2}{2} M \nabla_\theta \mathcal{L}(\theta) + \epsilon \sqrt{M} z \]
Langevin Diffusion on Riemannian manifold

- Discretised Langevin diffusion on manifold defines proposal mechanism

\[
\theta'_{d} = \theta_{d} + \frac{\epsilon^2}{2} \left( G^{-1}(\theta) \nabla_{\theta} \mathcal{L}(\theta) \right)_{d} - \epsilon^2 \sum_{i,j}^{D} G(\theta)_{ij}^{-1} \Gamma_{ij}^{d} + \epsilon \left( \sqrt{G^{-1}(\theta)} \mathbf{z} \right)_{d}
\]

- Manifold with constant curvature then proposal mechanism reduces to

\[
\theta' = \theta + \frac{\epsilon^2}{2} G^{-1}(\theta) \nabla_{\theta} \mathcal{L}(\theta) + \epsilon \sqrt{G^{-1}(\theta)} \mathbf{z}
\]

- MALA proposal with preconditioning

\[
\theta' = \theta + \frac{\epsilon^2}{2} M \nabla_{\theta} \mathcal{L}(\theta) + \epsilon \sqrt{M} \mathbf{z}
\]

- Proposal and acceptance probability

\[
p_{p}(\theta' | \theta) = \mathcal{N}(\theta' | \mu(\theta, \epsilon), \epsilon^2 G^{-1}(\theta))
\]

\[
p_{a}(\theta' | \theta) = \min \left[ 1, \frac{\pi(\theta') p_{p}(\theta | \theta')}{{\pi(\theta)} p_{p}(\theta' | \theta)} \right]
\]
Langevin Diffusion on Riemannian manifold

- Discretised Langevin diffusion on manifold defines proposal mechanism
  \[ \theta' = \theta + \frac{\epsilon^2}{2} \left( G^{-1}(\theta) \nabla_\theta \mathcal{L}(\theta) \right)_d - \epsilon^2 \sum_{i,j}^D G(\theta)_{ij}^{-1} \Gamma_{ij}^d + \epsilon \left( \sqrt{G^{-1}(\theta)} \mathbf{z} \right)_d \]

- Manifold with constant curvature then proposal mechanism reduces to
  \[ \theta' = \theta + \frac{\epsilon^2}{2} G^{-1}(\theta) \nabla_\theta \mathcal{L}(\theta) + \epsilon \sqrt{G^{-1}(\theta)} \mathbf{z} \]

- MALA proposal with preconditioning
  \[ \theta' = \theta + \frac{\epsilon^2}{2} M \nabla_\theta \mathcal{L}(\theta) + \epsilon \sqrt{M} \mathbf{z} \]

- Proposal and acceptance probability
  \[ p_p(\theta' | \theta) = \mathcal{N}(\theta' | \mu(\theta, \epsilon), \epsilon^2 G^{-1}(\theta)) \]
  \[ p_a(\theta' | \theta) = \min \left[ 1, \frac{\pi(\theta') p_p(\theta | \theta')}{\pi(\theta) p_p(\theta' | \theta)} \right] \]

- Proposal mechanism diffuses approximately along the manifold
Langevin Diffusion on Riemannian manifold
Langevin Diffusion on Riemannian manifold
Geodesic flow as proposal mechanism

Intuitive that proposal mechanism follow direct paths - geodesics

Consider Thermodynamic Integral - basis for Path & Bridge sampling

For $\theta \in \mathbb{R}^D$ with $p(y|\theta) = q(y|\theta)$

$z(\theta) - 1$ denote $U(y, \theta) = \partial_{\theta} \log q(y|\theta)$

and let $\theta(t)$ be a function of index $t$

$\lambda = \log \left( \frac{z(\theta(t_2))}{z(\theta(t_1))} \right) = \int_{t_2}^{t_1} \mathbb{E}_{y|\theta(t)} \left\{ \sum d\dot{\theta} \cdot U d(y, \theta) \right\} dt$

Quality of $N$-sample path estimate $\hat{\lambda}$ is $\text{var}(\hat{\lambda})$ i.e. independent sampling
Geodesic flow as proposal mechanism

• Intuitive that proposal mechanism follow direct paths - geodesics
Geodesic flow as proposal mechanism

- Intuitive that proposal mechanism follow direct paths - geodesics
- Consider Thermodynamic Integral - basis for Path & Bridge sampling

\[ \theta \in \mathbb{R}^D \text{ with } p(y|\theta) = q(y|\theta) \]

\[ z(\theta) - 1 \]

\[ U(y, \theta) = \partial_\theta \log q(y|\theta) \]

\[ \theta(t) \text{ be a function of index } \lambda = \log \frac{z(\theta(t_2))}{z(\theta(t_1))} = \int_{t_2}^{t_1} E_{y|\theta(t)} \left\{ \sum d\dot{\theta} \cdot d(\theta) \right\} dt \]

\[ \text{Quality of } N \text{-sample path estimate } \hat{\lambda} \text{ is } \text{var}(\hat{\lambda}) \text{ i.e. independent sampling} \]
Geodesic flow as proposal mechanism

- Intuitive that proposal mechanism follow direct paths - geodesics
- Consider Thermodynamic Integral - basis for Path & Bridge sampling
- For $\theta \in \mathbb{R}^D$ with $p(y|\theta) = q(y|\theta)z(\theta)^{-1}$ denote $U(y, \theta) = \partial_\theta \log q(y|\theta)$ and let $\theta(t)$ be a function of index $t$
Geodesic flow as proposal mechanism

- Intuitive that proposal mechanism follow direct paths - geodesics
- Consider Thermodynamic Integral - basis for Path & Bridge sampling
- For $\theta \in \mathbb{R}^D$ with $p(y|\theta) = q(y|\theta)z(\theta)^{-1}$ denote $U(y, \theta) = \partial_\theta \log q(y|\theta)$ and let $\theta(t)$ be a function of index $t$

$$\lambda = \log \left( \frac{z(\theta(t_2))}{z(\theta(t_1))} \right) = \int_{t_1}^{t_2} E_{y|\theta(t)} \left\{ \sum_d \dot{\theta}_d(t) U_d(y, \theta) \right\} dt$$
Geodesic flow as proposal mechanism

• Intuitive that proposal mechanism follow direct paths - geodesics

• Consider Thermodynamic Integral - basis for Path & Bridge sampling

• For $\theta \in \mathbb{R}^D$ with $p(y|\theta) = q(y|\theta)z(\theta)^{-1}$ denote $U(y, \theta) = \partial_\theta \log q(y|\theta)$ and let $\theta(t)$ be a function of index $t$

$$\lambda = \log \left( \frac{z(\theta(t_2))}{z(\theta(t_1))} \right) = \int_{t_1}^{t_2} E_{y|\theta(t)} \left\{ \sum_d \dot{\theta}_d(t) U_d(y, \theta) \right\} dt$$

• Quality of $N$-sample path estimate $\hat{\lambda}$ is $\text{var}(\hat{\lambda})$ i.e. independent sampling
Geodesic flow as proposal mechanism

- Intuitive that proposal mechanism follow direct paths - geodesics

- Consider Thermodynamic Integral - basis for Path & Bridge sampling

- For $\theta \in \mathbb{R}^D$ with $p(y|\theta) = q(y|\theta)z(\theta)^{-1}$ denote $U(y, \theta) = \partial_{\theta} \log q(y|\theta)$ and let $\theta(t)$ be a function of index $t$

$$
\lambda = \log \left( \frac{z(\theta(t_2))}{z(\theta(t_1))} \right) = \int_{t_1}^{t_2} E_{y|\theta(t)} \left\{ \sum_d \dot{\theta}_d(t) U_d(y, \theta) \right\} dt
$$

- Quality of $N$-sample path estimate $\hat{\lambda}$ is $\text{var}(\hat{\lambda})$ i.e. independent sampling

$$
\frac{1}{N} \left[ \int_{t_1}^{t_2} \sum_{i,j} \left( g_{ij} \dot{\theta}_i(t) \dot{\theta}_j(t) \right) dt - \lambda^2 \right] = \frac{1}{N} \left[ \int_{t_1}^{t_2} \sum_{i,j} \left( g_{ij} p_i(t) p_j(t) \right) dt - \lambda^2 \right]
$$

where $g_{ij} = E_{y|\theta(t)} \{ U_i(y, \theta) U_j(y, \theta) \}$ and $p_i(t) = \sum_j g_{ij} \dot{\theta}_j(t)$
Geodesic flow as proposal mechanism

- Variance of estimate can be minimised by employing path that minimises

\[
\int_{t_1}^{t_2} \dot{\theta}(t)^T G(\theta(t)) \dot{\theta}(t) dt = \int_{t_1}^{t_2} p(t)^T G^{-1}(\theta(t)) p(t) dt = \int_{t_1}^{t_2} H(\theta(t), p(t)) dt
\]
Geodesic flow as proposal mechanism

- Variance of estimate can be minimised by employing path that minimises

\[
\int_{t_1}^{t_2} \dot{\theta}(t)^T G(\theta(t)) \dot{\theta}(t) dt = \int_{t_1}^{t_2} p(t)^T G^{-1}(\theta(t)) p(t) dt = \int_{t_1}^{t_2} H(\theta(t), p(t)) dt
\]

where \(H(\theta(t), p(t))\) is Hamiltonian defined on the manifold with position specific metric tensor \(g_{ij}(\theta)\)
Geodesic flow as proposal mechanism

- Variance of estimate can be minimised by employing path that minimises

\[
\int_{t_1}^{t_2} \theta(t)^T G(\theta(t)) \dot{\theta}(t) dt = \int_{t_1}^{t_2} p(t)^T G^{-1}(\theta(t)) p(t) dt = \int_{t_1}^{t_2} H(\theta(t), p(t)) dt
\]

where \( H(\theta(t), p(t)) \) is Hamiltonian defined on the manifold with position specific metric tensor \( g_{ij}(\theta) \)

- This is satisfied by Geodesic equations

\[
\frac{d^2 \theta^i}{dt^2} + \sum_{k,l} \Gamma^i_{kl} \frac{d \theta^k}{dt} \frac{d \theta^l}{dt} = 0
\]
Geodesic flow as proposal mechanism

- Variance of estimate can be minimised by employing path that minimises

\[
\int_{t_1}^{t_2} \dot{\theta}(t)^T G(\theta(t)) \dot{\theta}(t) dt = \int_{t_1}^{t_2} p(t)^T G^{-1}(\theta(t)) p(t) dt = \int_{t_1}^{t_2} H(\theta(t), p(t)) dt
\]

where \( H(\theta(t), p(t)) \) is Hamiltonian defined on the manifold with position specific metric tensor \( g_{ij}(\theta) \)

- This is satisfied by Geodesic equations

\[
\frac{d^2 \theta^i}{dt^2} + \sum_{k,l} \Gamma^i_{kl} \frac{d\theta^k}{dt} \frac{d\theta^l}{dt} = 0
\]

- Replace \( D \) 2-order ODEs with \( 2D \) 1-order ODEs in Hamilton-Jacobi form

\[
\frac{d\theta}{dt} = \frac{\partial}{\partial p} H(\theta, p) \quad \frac{dp}{dt} = -\frac{\partial}{\partial \theta} H(\theta, p)
\]
Geodesic flow as proposal mechanism

- Variance of estimate can be minimised by employing path that minimises
  \[ \int_{t_1}^{t_2} \dot{\theta}(t)^T G(\theta(t)) \dot{\theta}(t) dt = \int_{t_1}^{t_2} p(t)^T G^{-1}(\theta(t)) p(t) dt = \int_{t_1}^{t_2} H(\theta(t), p(t)) dt \]

  where \( H(\theta(t), p(t)) \) is Hamiltonian defined on the manifold with position specific metric tensor \( g_{ij}(\theta) \)

- This is satisfied by Geodesic equations
  \[ \frac{d^2 \theta^i}{dt^2} + \sum_{k,l} \Gamma^i_{kl} \frac{d\theta^k}{dt} \frac{d\theta^l}{dt} = 0 \]

- Replace \( D \) 2-order ODEs with \( 2D \) 1-order ODEs in Hamilton-Jacobi form
  \[ \frac{d\theta}{dt} = \frac{\partial}{\partial p} H(\theta, p) \quad \frac{dp}{dt} = -\frac{\partial}{\partial \theta} H(\theta, p) \]

- Solving Hamilton equations defines lowest variance / shortest path between two densities on manifold
Geodesic flow as proposal mechanism

- Desirable that proposals follow direct path on manifold - geodesics
Geodesic flow as proposal mechanism

- Desirable that proposals follow direct path on manifold - geodesics
- How can this be exploited in the design of a transition operator?
Geodesic flow as proposal mechanism

- Desirable that proposals follow direct path on manifold - geodesics
- How can this be exploited in the design of a transition operator?
- First define log-density under model as $\mathcal{L}(\theta)$
Geodesic flow as proposal mechanism

- Desirable that proposals follow direct path on manifold - geodesics
- How can this be exploited in the design of a transition operator?
- First define log-density under model as $\mathcal{L}(\theta)$
- Introduce auxiliary variable $p \sim \mathcal{N}(0, G(\theta))$
Geodesic flow as proposal mechanism

- Desirable that proposals follow direct path on manifold - geodesics
- How can this be exploited in the design of a transition operator?
- First define log-density under model as $L(\theta)$
- Introduce auxiliary variable $p \sim \mathcal{N}(0, G(\theta))$
- Negative joint log density is

$$H(\theta, p) = -L(\theta) + \frac{1}{2} \log 2\pi^D |G(\theta)| + \frac{1}{2} p^T G(\theta)^{-1} p$$
Geodesic flow as proposal mechanism

- Desirable that proposals follow direct path on manifold - geodesics
- How can this be exploited in the design of a transition operator?
- First define log-density under model as $\mathcal{L}(\theta)$
- Introduce auxiliary variable $p \sim \mathcal{N}(0, G(\theta))$
- Negative joint log density is

$$H(\theta, p) = -\mathcal{L}(\theta) + \frac{1}{2} \log 2\pi^D |G(\theta)| + \frac{1}{2} p^T G(\theta)^{-1} p$$

- Negative joint log-density $\equiv$ Hamiltonian defined on Riemann manifold

$$H(\theta, p) = -\mathcal{L}(\theta) + \frac{1}{2} \log 2\pi^D |G(\theta)| + \frac{1}{2} p^T G(\theta)^{-1} p$$

\[\text{Potential Energy} + \text{Kinetic Energy}\]
Riemannian Hamiltonian Monte Carlo

- Marginal density follows as required

\[ \pi(\theta) \propto \frac{\exp \{ \mathcal{L}(\theta) \}}{\sqrt{2\pi^D |G(\theta)|}} \int \exp \left\{ -\frac{1}{2} p^T G(\theta)^{-1} p \right\} dp = \exp \{ \mathcal{L}(\theta) \} \]
Riemannian Hamiltonian Monte Carlo

• Marginal density follows as required

\[ \pi(\theta) \propto \frac{\exp \{ \mathcal{L}(\theta) \}}{\sqrt{2\pi^D |\mathbf{G}(\theta)|}} \int \exp \left\{ -\frac{1}{2} \mathbf{p}^T \mathbf{G}(\theta)^{-1} \mathbf{p} \right\} d\mathbf{p} = \exp \{ \mathcal{L}(\theta) \} \]

• Obtain samples from marginal \( \pi(\theta) \) using Metropolis-within-Gibbs sampler for \( p(\theta, \mathbf{p}) \)

\[
\mathbf{p}^{n+1} | \theta^n \sim \mathcal{N}(0, \mathbf{G}(\theta^n)) \\
\theta^{n+1} | \mathbf{p}^{n+1} \sim p(\theta^{n+1} | \mathbf{p}^{n+1})
\]
Riemannian Hamiltonian Monte Carlo

- Marginal density follows as required

\[
\pi(\theta) \propto \frac{\exp\{\mathcal{L}(\theta)\}}{\sqrt{2\pi^D|\mathbf{G}(\theta)|}} \int \exp\left\{-\frac{1}{2} \mathbf{p}^\top \mathbf{G}(\theta)^{-1} \mathbf{p}\right\} d\mathbf{p} = \exp\{\mathcal{L}(\theta)\}
\]

- Obtain samples from marginal \(\pi(\theta)\) using Metropolis-within-Gibbs sampler for \(p(\theta, \mathbf{p})\)

\[
\begin{align*}
\mathbf{p}^{n+1}|\theta^n & \sim \mathcal{N}(\mathbf{0}, \mathbf{G}(\theta^n)) \\
\theta^{n+1}|\mathbf{p}^{n+1} & \sim p(\theta^{n+1}|\mathbf{p}^{n+1})
\end{align*}
\]

- Integrate across geodesics to propose samples for \(p(\theta^{n+1}|\mathbf{p}^{n+1})\). Numerical symplectic integration of Hamilton-Jacobi
Riemannian Hamiltonian Monte Carlo

• Marginal density follows as required

\[
\pi(\theta) \propto \exp \left\{ \mathcal{L}(\theta) \right\} \frac{1}{\sqrt{2\pi^D|G(\theta)|}} \int \exp \left\{ -\frac{1}{2} p^T G(\theta)^{-1} p \right\} dp = \exp \{ \mathcal{L}(\theta) \}
\]

• Obtain samples from marginal \(\pi(\theta)\) using Metropolis-within-Gibbs sampler for \(p(\theta, p)\)

\[
p^{n+1} | \theta^n \sim \mathcal{N}(0, G(\theta^n))
\]

\[
\theta^{n+1} | p^{n+1} \sim p(\theta^{n+1} | p^{n+1})
\]

• Integrate across geodesics to propose samples for \(p(\theta^{n+1} | p^{n+1})\).

Numerical symplectic integration of Hamilton-Jacobi

\[
\frac{d\theta}{dt} = \frac{\partial}{\partial p} H(\theta, p) \quad \frac{dp}{dt} = -\frac{\partial}{\partial \theta} H(\theta, p)
\]

Riemannian Manifold Hamiltonian Monte Carlo

- Consider the Hamiltonian \( \tilde{H}(\theta, p) = \frac{1}{2} p^T \tilde{G}(\theta)^{-1} p \)
Riemannian Manifold Hamiltonian Monte Carlo

- Consider the Hamiltonian \( \tilde{H}(\theta, p) = \frac{1}{2} p^T \tilde{G}(\theta)^{-1} p \)

- Hamiltonians with only a quadratic kinetic energy term exactly describe geodesic flow on the coordinate space \( \theta \) with metric \( \tilde{G} \)
Riemannian Manifold Hamiltonian Monte Carlo

- Consider the Hamiltonian $\tilde{H}(\theta, p) = \frac{1}{2} p^T \tilde{G}(\theta)^{-1} p$

- Hamiltonians with only a quadratic kinetic energy term exactly describe geodesic flow on the coordinate space $\theta$ with metric $\tilde{G}$

- However our Hamiltonian is $H(\theta, p) = V(\theta) + \frac{1}{2} p^T G(\theta)^{-1} p$
Riemannian Manifold Hamiltonian Monte Carlo

- Consider the Hamiltonian $\tilde{H}(\theta, p) = \frac{1}{2} p^T \tilde{G}(\theta)^{-1} p$

- Hamiltonians with only a quadratic kinetic energy term exactly describe geodesic flow on the coordinate space $\theta$ with metric $\tilde{G}$

- However our Hamiltonian is $H(\theta, p) = V(\theta) + \frac{1}{2} p^T G(\theta)^{-1} p$

- If we define $\tilde{G}(\theta) = G(\theta) \times (h - V(\theta))$, where $h$ is a constant $H(\theta, p)$
Riemannian Manifold Hamiltonian Monte Carlo

- Consider the Hamiltonian \( \tilde{H}(\theta, p) = \frac{1}{2} p^T \tilde{G}(\theta)^{-1} p \)

- Hamiltonians with only a quadratic kinetic energy term exactly describe geodesic flow on the coordinate space \( \theta \) with metric \( \tilde{G} \)

- However our Hamiltonian is \( H(\theta, p) = V(\theta) + \frac{1}{2} p^T G(\theta)^{-1} p \)

- If we define \( \tilde{G}(\theta) = G(\theta) \times (h - V(\theta)) \), where \( h \) is a constant \( H(\theta, p) \)

- Flow for \( H(\theta, p) \) and \( \tilde{H}(\theta, p) \) equivalent along energy level \( h \)
Riemannian Manifold Hamiltonian Monte Carlo

- Consider the Hamiltonian $\tilde{H}(\theta, p) = \frac{1}{2} p^T \tilde{G}(\theta)^{-1} p$

- Hamiltonians with only a quadratic kinetic energy term exactly describe geodesic flow on the coordinate space $\theta$ with metric $\tilde{G}$

- However our Hamiltonian is $H(\theta, p) = V(\theta) + \frac{1}{2} p^T G(\theta)^{-1} p$

- If we define $\tilde{G}(\theta) = G(\theta) \times (h - V(\theta))$, where $h$ is a constant $H(\theta, p)$

- Flow for $H(\theta, p)$ and $\tilde{H}(\theta, p)$ equivalent along energy level $h$

- The solution of

$$ \frac{d\theta}{dt} = \frac{\partial}{\partial p} H(\theta, p) \quad \frac{dp}{dt} = -\frac{\partial}{\partial \theta} H(\theta, p) $$
Riemannian Manifold Hamiltonian Monte Carlo

- Consider the Hamiltonian \( \tilde{H}(\theta, p) = \frac{1}{2} p^T \tilde{G}(\theta)^{-1} p \)

- Hamiltonians with only a quadratic kinetic energy term exactly describe geodesic flow on the coordinate space \( \theta \) with metric \( \tilde{G} \)

- However our Hamiltonian is \( H(\theta, p) = V(\theta) + \frac{1}{2} p^T G(\theta)^{-1} p \)

- If we define \( \tilde{G}(\theta) = G(\theta) \times (h - V(\theta)) \), where \( h \) is a constant \( H(\theta, p) \)

- Flow for \( H(\theta, p) \) and \( \tilde{H}(\theta, p) \) equivalent along energy level \( h \)

- The solution of

\[
\frac{d\theta}{dt} = \frac{\partial}{\partial p} H(\theta, p) \quad \frac{dp}{dt} = -\frac{\partial}{\partial \theta} H(\theta, p)
\]

is therefore equivalent to the solution of

\[
\frac{d^2 \theta^i}{dt^2} + \sum_{k,l} \tilde{\Gamma}_{kl}^i \frac{d\theta^k}{dt} \frac{d\theta^l}{dt} = 0
\]
Riemannian Manifold Hamiltonian Monte Carlo

- Consider the Hamiltonian \( \tilde{H}(\theta, p) = \frac{1}{2} p^T \tilde{G}(\theta)^{-1} p \)

- Hamiltonians with only a quadratic kinetic energy term exactly describe geodesic flow on the coordinate space \( \theta \) with metric \( \tilde{G} \)

- However our Hamiltonian is \( H(\theta, p) = V(\theta) + \frac{1}{2} p^T G(\theta)^{-1} p \)

- If we define \( \tilde{G}(\theta) = G(\theta) \times (h - V(\theta)) \), where \( h \) is a constant \( H(\theta, p) \)

- Flow for \( H(\theta, p) \) and \( \tilde{H}(\theta, p) \) equivalent along energy level \( h \)

- The solution of

\[
\frac{d\theta}{dt} = \frac{\partial}{\partial p} H(\theta, p) \quad \frac{dp}{dt} = -\frac{\partial}{\partial \theta} H(\theta, p)
\]

is therefore equivalent to the solution of

\[
\frac{d^2 \theta^i}{dt^2} + \sum_{k, l} \tilde{\Gamma}^i_{kl} \frac{d\theta^k}{dt} \frac{d\theta^l}{dt} = 0
\]

- RMHMC proposals are along the manifold geodesics
Gaussian Mixture Model

- Univariate finite mixture model

\[ p(x|\mu, \sigma^2) = 0.7 \times \mathcal{N}(x|0, \sigma^2) + 0.3 \times \mathcal{N}(x|\mu, \sigma^2) \]
Gaussian Mixture Model

- Univariate finite mixture model

\[ p(x|\mu, \sigma^2) = 0.7 \times \mathcal{N}(x|0, \sigma^2) + 0.3 \times \mathcal{N}(x|\mu, \sigma^2) \]

Figure: Arrows correspond to the gradients and ellipses to the inverse metric tensor. Dashed lines are isocontours of the joint log density.
Gaussian Mixture Model
Gaussian Mixture Model
Stochastic Volatility Model

A stochastic volatility model (SVM) is defined with the latent volatilities taking the form of an AR(1) process such that

\[ y_t = \epsilon_t \beta \exp \left( \frac{x_t}{2} \right) \]

with

\[ x_{t+1} = \phi x_t + \eta_{t+1} \]

where

\[ \epsilon_t \sim \mathcal{N}(0, 1) \]
\[ \eta_t \sim \mathcal{N}(0, \sigma^2) \]

and

\[ x_1 \sim \mathcal{N}(0, \sigma^2/(1 - \phi^2)) \]
Stochastic Volatility Model

A stochastic volatility model with an AR(1) latent volatility process has joint density

\[ p(y, x, \beta, \phi, \sigma) = \prod_{t=1}^{T} p(y_t|x_t, \beta)p(x_1) \prod_{t=2}^{T} p(x_t|x_{t-1}, \phi, \sigma) \pi(\beta) \pi(\phi) \pi(\sigma). \]
Stochastic Volatility Model

- A stochastic volatility model with an AR(1) latent volatility process has joint density

\[ p(y, x, \beta, \phi, \sigma) = \prod_{t=1}^{T} p(y_t|x_t, \beta)p(x_1) \prod_{t=2}^{T} p(x_t|x_{t-1}, \phi, \sigma)\pi(\beta)\pi(\phi)\pi(\sigma). \]

- Split up the sampling procedure into two steps, simulate from

\[
\begin{align*}
\beta, \phi, \sigma | y, x & \sim p(\beta, \phi, \sigma | y, x) \\
x | y, \beta, \phi, \sigma & \sim p(x | y, \beta, \phi, \sigma)
\end{align*}
\]
Stochastic Volatility Model

- A stochastic volatility model with an AR(1) latent volatility process has joint density

\[ p(y, x, \beta, \phi, \sigma) = \prod_{t=1}^{T} p(y_t | x_t, \beta) p(x_1) \prod_{t=2}^{T} p(x_t | x_{t-1}, \phi, \sigma) \pi(\beta) \pi(\phi) \pi(\sigma). \]

- Split up the sampling procedure into two steps, simulate from

\[ \beta, \phi, \sigma | y, x \sim p(\beta, \phi, \sigma | y, x) \]
\[ x | y, \beta, \phi, \sigma \sim p(x | y, \beta, \phi, \sigma) \]

- Metric tensor for parameters

\[ G(\beta, \phi, \sigma) = \begin{bmatrix} \frac{2T}{\beta^2} & 0 & 0 \\ 0 & 2T & 2\phi \\ 0 & 2\phi & 2\phi^2 + (T - 1)(1 - \phi^2) \end{bmatrix} \]
A stochastic volatility model with an AR(1) latent volatility process has joint density

\[ p(y, x, \beta, \phi, \sigma) = \prod_{t=1}^{T} p(y_t|x_t, \beta)p(x_1) \prod_{t=2}^{T} p(x_t|x_{t-1}, \phi, \sigma)\pi(\beta)\pi(\phi)\pi(\sigma). \]

Split up the sampling procedure into two steps, simulate from

\[ \beta, \phi, \sigma|y, x \sim p(\beta, \phi, \sigma|y, x) \]
\[ x|y, \beta, \phi, \sigma \sim p(x|y, \beta, \phi, \sigma) \]

Metric tensor for parameters

\[ G(\beta, \phi, \sigma) = \begin{bmatrix} \frac{2T}{\beta^2} & 0 & 0 \\ 0 & 2T & 2\phi \\ 0 & 2\phi & 2\phi^2 + (T-1)(1-\phi^2) \end{bmatrix} \]

Metric tensor for latent volatilities

\[ G(x) = \frac{1}{2} \times I + C^{-1} \]

with \( C(t+n, t) = E\{x_{t+n}x_t\} = \phi^{\lfloor n \rfloor} \sigma^2/(1 - \phi^2) \) defines a flat manifold
Metropolis, Parameters of Stoch Vol Model, Acc Rate 25%
Metropolis, Parameters of Stoch Vol Model, Acc Rate 25%
RMHMC Parameters of Stoch Vol Model, Acc Rate 95%
RMHMC Parameters of Stoch Vol Model, Acc Rate 95%
Stochastic Volatility Model - Performance

Table: 2000 simulated observations with $\beta = 0.65$, $\sigma = 0.15$ and $\phi = 0.98$ - Comparison of sampling the parameters $\beta$, $\sigma$ and $\phi$ after 20,000 posterior samples averaged over 10 runs

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Time</th>
<th>ESS ($\beta, \sigma, \phi$)</th>
<th>S.E. ($\beta, \sigma, \phi$)</th>
<th>s/(Min ESS)</th>
<th>Rel. Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALA</td>
<td>44.0</td>
<td>(19.1, 11.3, 30.1)</td>
<td>(1.9,0.8,2.1)</td>
<td>3.89</td>
<td>$\times$ 36.7</td>
</tr>
<tr>
<td>HMC</td>
<td>424.8</td>
<td>(117, 81, 198)</td>
<td>(9.3, 3.1, 10.3)</td>
<td>5.19</td>
<td>$\times$ 27.5</td>
</tr>
<tr>
<td>mMALA</td>
<td>2455.9</td>
<td>(17.2, 17.4, 44.5)</td>
<td>(2.8, 2.4, 9.2)</td>
<td>142.8</td>
<td>$\times$ 1</td>
</tr>
<tr>
<td>RM-HMC</td>
<td>329.4</td>
<td>(325, 139, 344)</td>
<td>(19.0, 7.3, 25.2)</td>
<td>2.37</td>
<td>$\times$ 60.3</td>
</tr>
</tbody>
</table>

Table: 2000 simulated observations with $\beta = 0.65$, $\sigma = 0.15$ and $\phi = 0.98$ - Comparison of sampling the latent volatilities after 20,000 posterior samples averaged over 10 runs

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Time</th>
<th>ESS (min, median, max)</th>
<th>s/(Min ESS)</th>
<th>Rel. Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALA</td>
<td>44.0</td>
<td>(9.7, 16.7, 28.4)</td>
<td>4.53</td>
<td>$\times$ 7.5</td>
</tr>
<tr>
<td>HMC</td>
<td>424.8</td>
<td>(409,624,1239)</td>
<td>1.04</td>
<td>$\times$ 32.9</td>
</tr>
<tr>
<td>mMALA</td>
<td>2455.9</td>
<td>(71.8, 131.0, 329.8)</td>
<td>34.2</td>
<td>$\times$ 1</td>
</tr>
<tr>
<td>RM-HMC</td>
<td>329.4</td>
<td>(977, 1689, 3376)</td>
<td>0.34</td>
<td>$\times$ 100.6</td>
</tr>
</tbody>
</table>
Stochastic Volatility Model - Performance

Figure: Posterior marginal densities for $\beta$, $\sigma$ and $\phi$ respectively, employing RM-HMC to draw 20,000 samples of the parameters and latent volatilities using a simulated dataset consisting of 2000 observations. The true values are $\beta = 0.65$, $\sigma = 0.15$ and $\phi = 0.98$. 
Log-Gaussian Cox Point Process with Latent Field

- The joint density for Poisson counts and latent Gaussian field

\[ p(y, x|\mu, \sigma, \beta) \propto \prod_{i,j}^{64} \exp\{y_{i,j}x_{i,j} - m \exp(x_{i,j})\} \exp\left(- (x - \mu 1)^T \Sigma^{-1}_\theta (x - \mu 1)/2\right) \]
Log-Gaussian Cox Point Process with Latent Field

- The joint density for Poisson counts and latent Gaussian field

\[ p(y, x|\mu, \sigma, \beta) \propto \prod_{i,j}^{64} \exp\{y_{i,j}x_{i,j} - m \exp(x_{i,j})\} \exp(-\mathbf{x} - \mu \mathbf{1})^T \Sigma \theta^{-1} (\mathbf{x} - \mu \mathbf{1}) / 2 \]

- Metric tensors

\[
\begin{align*}
G(\theta)_{i,j} &= \frac{1}{2} \text{trace} \left( \Sigma \theta^{-1} \frac{\partial \Sigma \theta}{\partial \theta_i} \Sigma \theta^{-1} \frac{\partial \Sigma \theta}{\partial \theta_j} \right) \\
G(x) &= \Lambda + \Sigma \theta^{-1}
\end{align*}
\]

where \( \Lambda \) is diagonal with elements \( m \exp(\mu + (\Sigma \theta)_{i,i}) \)
Log-Gaussian Cox Point Process with Latent Field

• The joint density for Poisson counts and latent Gaussian field

\[ p(y, x|\mu, \sigma, \beta) \propto \prod_{i,j}^{64} \exp\{y_{i,j}x_{i,j} - m \exp(x_{i,j})\} \exp(- (x - \mu 1)^T \Sigma^{-1}_\theta (x - \mu 1)/2) \]

• Metric tensors

\[
\begin{align*}
G(\theta)_{i,j} &= \frac{1}{2} \text{trace} \left( \Sigma^{-1}_\theta \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma^{-1}_\theta \frac{\partial \Sigma_\theta}{\partial \theta_j} \right) \\
G(x) &= \Lambda + \Sigma^{-1}_\theta
\end{align*}
\]

where \( \Lambda \) is diagonal with elements \( m \exp(\mu + (\Sigma_\theta)_{i,i}) \)

• Latent field metric tensor defining flat manifold is \( 4096 \times 4096, \mathcal{O}(N^3) \) obtained from parameter conditional
Log-Gaussian Cox Point Process with Latent Field

- The joint density for Poisson counts and latent Gaussian field

\[
p(y, x|\mu, \sigma, \beta) \propto \prod_{i,j}^{64} \exp\{y_{i,j}x_{i,j} - m \exp(x_{i,j})\} \exp(- (x - \mu 1)^T \Sigma^{-1} (x - \mu 1)/2)
\]

- Metric tensors

\[
G(\theta)_{i,j} = \frac{1}{2} \text{trace} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta} \right)
\]

\[
G(x) = \Lambda + \Sigma^{-1}
\]

where \(\Lambda\) is diagonal with elements \(m \exp(\mu + (\Sigma_{\theta})_{i,i})\)

- Latent field metric tensor defining flat manifold is \(4096 \times 4096, O(N^3)\) obtained from parameter conditional

- MALA requires transformation of latent field to sample - with separate tuning in transient and stationary phases of Markov chain
Table: Sampling the latent variables of a Log-Gaussian Cox Process - Comparison of sampling methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Time</th>
<th>ESS (Min, Med, Max)</th>
<th>s/Min ESS</th>
<th>Rel. Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALA (Transient)</td>
<td>31,577</td>
<td>(3, 8, 50)</td>
<td>10,605</td>
<td>×1</td>
</tr>
<tr>
<td>MALA (Stationary)</td>
<td>31,118</td>
<td>(4, 16, 80)</td>
<td>7836</td>
<td>×1.35</td>
</tr>
<tr>
<td>mMALA</td>
<td>634</td>
<td>(26, 84, 174)</td>
<td>24.1</td>
<td>×440</td>
</tr>
<tr>
<td>RMHMC</td>
<td>2936</td>
<td>(1951, 4545, 5000)</td>
<td>1.5</td>
<td>×7070</td>
</tr>
</tbody>
</table>
RMHMC for Log-Gaussian Cox Point Processes

Figure: Kernel density estimates of the hyperparameter samples obtained from Gibbs style sampling from the Log-Gaussian Cox model. The true values are $\sigma = 0.19$ (left hand plot) and $\beta = 0.03$ (right hand plot).
Figure: Kernel density estimates of the hyperparameter samples obtained from Gibbs style sampling from the Log-Gaussian Cox model. The true values are $\sigma = 0.19$ (left hand plot) and $\beta = 0.03$ (right hand plot).

The dark underbelly of hierarchical models.....
Hierarchical Models

- Consider a simple hierarchical model:

Hyper parameters: \( \phi_1, \ldots, \phi_m \)

Latent variables: \( \omega_1, \omega_2, \ldots, \omega_n \)

Observed data: \( y_1, y_2, \ldots, y_n \)

- Often convenient to sample \( \omega \mid \phi, y \), e.g. S.V. model.

- Alternately updating \( \phi \mid \omega, y \) (i.e. Gibbs sampling) results in poor mixing: \( \phi \) highly correlated with \( \omega \), and correlation increases with \( N \).

- Difficult to construct alternative proposal schemes, as shape of density can change dramatically, particularly if \( \phi \) includes a scale term for \( \omega \).

- Methods such as Ancillary-Sufficiency Interweaving seek to resolve this problem by alternating between complementary parameterisations.
Hierarchical Models

- Consider a simple hierarchical model:

  Hyper parameters
  \( \phi_1, \ldots, \phi_m \)

  Latent variables
  \( \omega_1, \omega_2, \ldots, \omega_n \)

  Observed data
  \( y_1, y_2, \ldots, y_n \)

- Often convenient to sample \( \omega \mid \phi, y \) e.g. S.V. model.
Hierarchical Models

- Consider a simple hierarchical model:

  Hyper parameters
  \[ \phi_1 \cdots \phi_m \]

  Latent variables
  \[ \omega_1 \omega_2 \cdots \omega_n \]

  Observed data
  \[ y_1 y_2 \cdots y_n \]

- Often convenient to sample \( \omega \mid \phi, y \) e.g. S.V. model.

- Alternately updating \( \phi \mid \omega, y \) (i.e. Gibbs sampling) results in poor mixing: \( \phi \) highly correlated with \( \omega \), and correlation increases with \( N \).
Hierarchical Models

- Consider a simple hierarchical model:

  Hyper parameters \( \phi_1, \ldots, \phi_m \)

  Latent variables \( \omega_1, \omega_2, \ldots, \omega_n \)

  Observed data \( y_1, y_2, \ldots, y_n \)

- Often convenient to sample \( \omega \mid \phi, y \) e.g. S.V. model.

- Alternately updating \( \phi \mid \omega, y \) (i.e. Gibbs sampling) results in poor mixing: \( \phi \) highly correlated with \( \omega \), and correlation increases with \( N \).

- Difficult to construct alternative proposal schemes, as shape of density can change dramatically, particularly if \( \phi \) includes a scale term for \( \omega \).
Hierarchical Models

- Consider a simple hierarchical model:

  Hyper parameters \( \phi_1, \phi_2, \ldots, \phi_m \)

  Latent variables \( \omega_1, \omega_2, \ldots, \omega_n \)

  Observed data \( y_1, y_2, \ldots, y_n \)

- Often convenient to sample \( \omega | \phi, y \) e.g. S.V. model.

- Alternately updating \( \phi | \omega, y \) (i.e. Gibbs sampling) results in poor mixing: \( \phi \) highly correlated with \( \omega \), and correlation increases with \( N \).

- Difficult to construct alternative proposal schemes, as shape of density can change dramatically, particularly if \( \phi \) includes a scale term for \( \omega \).

- Methods such as Ancillary-Sufficiency Interweaving seek to resolve this problem by alternating between complementary parameterisations.
Hierarchical Models

- Consider a simple hierarchical model:

Hyper parameters: $\phi_1, \ldots, \phi_m$

Latent variables: $\omega_1, \omega_2, \ldots, \omega_n$

Observed data: $y_1, y_2, \ldots, y_n$

- Often convenient to sample $\omega \mid \phi, y$ e.g. S.V. model.

- Alternately updating $\phi \mid \omega, y$ (i.e. Gibbs sampling) results in poor mixing: $\phi$ highly correlated with $\omega$, and correlation increases with $N$.

- Difficult to construct alternative proposal schemes, as shape of density can change dramatically, particularly if $\phi$ includes a scale term for $\omega$.

- Methods such as Ancillary-Sufficiency Interweaving seek to resolve this problem by alternating between complementary parameterisations.
Foliations

- Seek to solve the problem directly:
  (a) Update high-dimensional $\omega \mid \phi, y$ as before.

- Ideally, do (b) by:
  1. Let $U_i = F(\omega_i \mid \phi, y)$ (conditional CDF)
  2. Update $\phi \mid y$ (since $\phi$ is independent of $U_i$).
  3. Let $\omega_i = F^{-1}(U_i \mid \phi, y)$.

- Geometrically, each value of $U_i$ characterises a submanifold of the parameter space; these form a foliation: a system of locally parallel submanifolds.

- Requires "global" information: if we could do this no need for MCMC.
Foliations

- Seek to solve the problem directly:
  (a) Update high-dimensional $\omega | \phi, y$ as before.
  (b) Jointly update $(\phi, \omega)$ to be “as independent as possible” to (a).

- Geometrically, each value of $U$ characterises a submanifold of the parameter space; these form a foliation: a system of locally parallel submanifolds.

- Requires “global” information: if we could do this no need for MCMC.
Foliations

- Seek to solve the problem directly:
  (a) Update high-dimensional $\omega \mid \phi, y$ as before.
  (b) Jointly update $(\phi, \omega)$ to be “as independent as possible” to (a).

- Ideally, do (b) by:
  1. Let $U_i = F(\omega_i \mid \phi, y)$ (conditional CDF)
Foliations

• Seek to solve the problem directly:
  (a) Update high-dimensional \( \omega \mid \phi, y \) as before.
  (b) Jointly update \((\phi, \omega)\) to be “as independent as possible” to (a).

• Ideally, do (b) by:
  1. Let \( U_i = F(\omega_i \mid \phi, y) \) (conditional CDF)
  2. Update \( \phi \mid y \) (since \( \phi \) is independent of \( U_i \)).
Foliations

- Seek to solve the problem directly:
  (a) Update high-dimensional $\omega \mid \phi, y$ as before.
  (b) Jointly update $(\phi, \omega)$ to be “as independent as possible” to (a).

- Ideally, do (b) by:
  1. Let $U_i = F(\omega_i \mid \phi, y)$ (conditional CDF)
  2. Update $\phi \mid y$ (since $\phi$ is independent of $U_i$).
  3. Let $\omega_i = F^{-1}(U_i \mid \phi, y)$. 
Foliations

- Seek to solve the problem directly:
  - (a) Update high-dimensional $\omega \mid \phi, y$ as before.
  - (b) Jointly update $(\phi, \omega)$ to be “as independent as possible” to (a).

- Ideally, do (b) by:
  1. Let $U_i = F(\omega_i \mid \phi, y)$ (conditional CDF)
  2. Update $\phi \mid y$ (since $\phi$ is independent of $U_i$).
  3. Let $\omega_i = F^{-1}(U_i \mid \phi, y)$.

- Geometrically, each value of $U$ characterises a submanifold of the parameter space; these form a foliation: a system of locally parallel submanifolds.
Foliations

- Seek to solve the problem directly:
  (a) Update high-dimensional $\omega | \phi, y$ as before.
  (b) Jointly update $(\phi, \omega)$ to be “as independent as possible” to (a).

- Ideally, do (b) by:
  1. Let $U_i = F(\omega_i | \phi, y)$ (conditional CDF)
  2. Update $\phi|y$ (since $\phi$ is independent of $U_i$).
  3. Let $\omega_i = F^{-1}(U_i | \phi, y)$.

- Geometrically, each value of $U$ characterises a submanifold of the parameter space; these form a foliation: a system of locally parallel submanifolds.

- Requires “global” information: if we could do this no need for MCMC.
Locally orthogonal foliations

- Define a notion of “local independence”: \( \omega, \phi \) locally independent at \((\omega_0, \phi_0)\) if:

\[
\frac{\partial^2 \log p(\omega, \phi)}{\partial \omega \partial \phi}(\omega_0, \phi_0) = 0
\]

- Local independence everywhere \( \Rightarrow \) independence
Locally orthogonal foliations

- Define a notion of “local independence”: $\omega, \phi$ locally independent at $(\omega_0, \phi_0)$ if:
  $$\frac{\partial^2 \log p(\omega, \phi)}{\partial \omega \partial \phi}(\omega_0, \phi_0) = 0$$

- Local independence everywhere $\Rightarrow$ independence
- Equivalent to orthogonality under the observed information metric $G_{\text{obs}}$

- General idea: in step (b) locally propose new values $(\phi, \omega) + (\Delta \phi, \Delta \omega)$ to be orthogonal to step (a). That is:
  $$(\Delta \phi, \Delta \omega)^\top G_{\text{obs}}(0, 1) = 0$$

- Define constrained Hamiltonian dynamics: in essence, RMHMC constrained to the orthogonal submanifold.

- Early results on Student-t hopeful.... more to do.... stay tuned....
Locally orthogonal foliations

- Define a notion of “local independence”: \( \omega, \phi \) locally independent at \((\omega_0, \phi_0)\) if:

\[
\frac{\partial^2 \log p(\omega, \phi)}{\partial \omega \partial \phi}(\omega_0, \phi_0) = 0
\]

- Local independence everywhere \( \Rightarrow \) independence
- Equivalent to orthogonality under the observed information metric \( G_{\text{obs}} \)
- General idea: in step (b) locally propose new values \((\phi, \omega) + (\Delta \phi, \Delta \omega)\) to be orthogonal to step (a). That is:

\[
(\Delta \phi, \Delta \omega)^\top G_{\text{obs}} (0, 1) = 0
\]
Locally orthogonal foliations

- Define a notion of “local independence”: \( \omega, \phi \) locally independent at \((\omega_0, \phi_0)\) if:

\[
\frac{\partial^2 \log p(\omega, \phi)}{\partial \omega \partial \phi}(\omega_0, \phi_0) = 0
\]

- Local independence everywhere \( \Rightarrow \) independence
- Equivalent to orthogonality under the observed information metric \( G_{obs} \)
- General idea: in step (b) locally propose new values

\[
(\phi, \omega) + (\Delta \phi, \Delta \omega)
\]

to be orthogonal to step (a). That is:

\[
(\Delta \phi, \Delta \omega)^\top G_{obs} (0, 1) = 0
\]

- Define constrained Hamiltonian dynamics: in essence, RMHMC constrained to the orthogonal submanifold.
Locally orthogonal foliations

- Define a notion of “local independence”: \( \omega, \phi \) locally independent at \((\omega_0, \phi_0)\) if:

\[
\frac{\partial^2 \log p(\omega, \phi)}{\partial \omega \partial \phi}(\omega_0, \phi_0) = 0
\]

- Local independence everywhere \( \Rightarrow \) independence
- Equivalent to orthogonality under the observed information metric \( G_{\text{obs}} \)
- General idea: in step (b) locally propose new values \((\phi, \omega) + (\Delta \phi, \Delta \omega)\) to be orthogonal to step (a). That is:

\[
(\Delta \phi, \Delta \omega) \top G_{\text{obs}} (0, 1) = 0
\]

- Define constrained Hamiltonian dynamics: in essence, RMHMC constrained to the orthogonal submanifold.
- Early results on Student-t hopeful.... more to do....
Locally orthogonal foliations

- Define a notion of “local independence”: $\omega, \phi$ locally independent at $(\omega_0, \phi_0)$ if:

$$\frac{\partial^2 \log p(\omega, \phi)}{\partial \omega \partial \phi}(\omega_0, \phi_0) = 0$$

- Local independence everywhere $\Rightarrow$ independence
- Equivalent to orthogonality under the observed information metric $G_{obs}$
- General idea: in step (b) locally propose new values $(\phi, \omega) + (\Delta \phi, \Delta \omega)$ to be orthogonal to step (a). That is:

$$(\Delta \phi, \Delta \omega)^\top G_{obs} (0, 1) = 0$$

- Define constrained Hamiltonian dynamics: in essence, RMHMC constrained to the orthogonal submanifold.
- Early results on Student-t hopeful.... more to do.... stay tuned.........
Nonlinear Dynamic System - Circadian Clock Gene Control
Nonlinear Dynamic System - Circadian Clock Gene Control

\[
\begin{align*}
\frac{d[LHY]_m}{dt} &= \frac{n_1[TOC1]_n}{g_1^a + [TOC1]_n^a} - \frac{m_1[LHY]_m}{k_1 + [LHY]_m} \\
\frac{d[LHY]_c}{dt} &= p_1[LHY]_m - r_1[LHY]_c + r_2[LHY]_n - \frac{m_2[LHY]_c}{k_2 + [LHY]_c} \\
\frac{d[LHY]_n}{dt} &= r_1[LHY]_c - r_2[LHY]_n - \frac{m_3[LHY]_n}{k_3 + [LHY]_n} \\
\frac{d[TOC1]_m}{dt} &= \frac{n_2g_2^b}{g_2^b + [LHY]_n^b} - \frac{m_4[TOC1]_m}{k_4 + [TOC1]_m} \\
\frac{d[TOC1]_c}{dt} &= p_2[TOC1]_m - r_3[TOC1]_c + r_4[TOC1]_n - \frac{m_5[TOC1]_c}{k_5 + [TOC1]_c} \\
\frac{d[TOC1]_n}{dt} &= r_3[TOC1]_c - r_4[TOC1]_n - \frac{m_6[TOC1]_n}{k_6 + [TOC1]_n}
\end{align*}
\]
Nonlinear Dynamic System - Circadian Clock Gene Control
Nonlinear Dynamic System - Circadian Clock Gene Control
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
- Transitions happening at random times.
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
- Transitions happening at random times.
- Transition rates depend on current state and unknown rate parameters.
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
- Transitions happening at random times.
- Transition rates depend on current state and unknown rate parameters.
- Markov property, \( p(x_t | x_{t-1}, \ldots, x_0) = p(x_t | x_{t-1}) \)
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
- Transitions happening at random times.
- Transition rates depend on current state and unknown rate parameters.
- Markov property, \( p(x_t|x_{t-1},...,x_0) = p(x_t|x_{t-1}) \)

- \( x_t = (x_{1,t},...,x_{N,t})^T \) is the state of the system at time t for N random variables.
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
- Transitions happening at random times.
- Transition rates depend on current state and unknown rate parameters.
- Markov property, \( p(x_t|x_{t-1}, \ldots, x_0) = p(x_t|x_{t-1}) \)
- \( x_t = (x_{1,t}, \ldots, x_{N,t})^T \) is the state of the system at time \( t \) for \( N \) random variables.
- State change vectors \( s_j = (s_{1,j}, \ldots, s_{N,j})^T, \ j \in \{1, \ldots, M\} \)
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
- Transitions happening at random times.
- Transition rates depend on current state and unknown rate parameters.
- Markov property, \( p(\mathbf{x}_t|\mathbf{x}_{t-1}, \ldots, \mathbf{x}_0) = p(\mathbf{x}_t|\mathbf{x}_{t-1}) \)

\( \mathbf{x}_t = (x_{1,t}, \ldots, x_{N,t})^T \) is the state of the system at time \( t \) for \( N \) random variables.

- State change vectors \( \mathbf{s}_j = (s_{1,j}, \ldots, s_{N,j})^T, \ j \in \{1, \ldots, M\} \)
- Transition rates, \( f_j(\mathbf{x}, \theta)dt \) the transition probability for state \( j \) in the interval \( [t, t + dt) \).

Master Equation (forward equation)

\[
\frac{dp_t(x)}{dt} = \sum_{j=1}^{M} \left[ f_j(x + s_j, \theta) p_t(x + s_j) - f_j(x, \theta) p_t(x) \right]
\]
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
- Transitions happening at random times.
- Transition rates depend on current state and unknown rate parameters.
- Markov property, \( p(x_t|x_{t-1}, \ldots, x_0) = p(x_t|x_{t-1}) \)

- \( x_t = (x_{1,t}, \ldots, x_{N,t})^T \) is the state of the system at time \( t \) for \( N \) random variables.
- State change vectors \( s_j = (s_{1,j}, \ldots, s_{N,j})^T, j \in \{1, \ldots, M\} \)
- Transition rates, \( f_j(x, \theta)dt \) the transition probability for state \( j \) in the interval \([t, t + dt)\).

Master Equation (forward equation)

\[
\frac{dp_t(x)}{dt} = \sum_{j=1}^{M} [f_j(x + s_j, \theta)p_t(x + s_j) - f_j(x, \theta)p_t(x)]
\]
Markov Jump Processes

- Continuous stochastic process on a discrete state space.
- Transitions happening at random times.
- Transition rates depend on current state and unknown rate parameters.
- Markov property, \( p(x_t|x_{t-1}, \ldots, x_0) = p(x_t|x_{t-1}) \)
- \( x_t = (x_{1,t}, \ldots, x_{N,t})^T \) is the state of the system at time \( t \) for \( N \) random variables.
- State change vectors \( s_j = (s_{1,j}, \ldots, s_{N,j})^T, \ j \in \{1, \ldots, M\} \)
- Transition rates, \( f_j(x, \theta)dt \) the transition probability for state \( j \) in the interval \([t, t + dt)\).

Master Equation (forward equation)

\[
\frac{dp_t(x)}{dt} = \sum_{j=1}^{M} [f_j(x + s_j, \theta)p_t(x + s_j) - f_j(x, \theta)p_t(x)]
\]
Diffusion Approximation

• Informal derivation, \( \tau \)-leaping:

\[
\begin{align*}
\text{Choose } \tau > 0 \text{ such that:} \\
& f_j(x_{t'}, \theta) \approx f_j(x_t, \theta), \quad \forall t' \in [t, t+\tau], \forall j \in [1, M] \\
& f_j(x_t, \theta) \gg 1, \quad \forall j \in [1, M] \\
\end{align*}
\]

Conditions (1) and (2) can be satisfied if \( x_i \gg 1 \).

• (1) implies that the number of transitions to states \( j \) are independently Poisson distributed with mean \( f_j(x_t, \theta) \tau \).

• (2) implies that the number of transitions can be reasonably approximated by a Normal distribution.

Langevin Equation

\[
dx_t = S f(x_t, \theta) \, dt + \frac{1}{\sqrt{\Omega S}} \sqrt{\text{diag}(f(x_t, \theta))} \, dB_t
\]
Diffusion Approximation

- Informal derivation, $\tau$-leaping:
- Choose $\tau > 0$ such that:

$$f_j(x_{t'}, \theta) \approx f_j(x_t, \theta), \quad \forall t' \in [t, t + \tau], \forall j \in [1, M] \quad (1)$$

$$f_j(x_t, \theta) \tau \gg 1, \quad \forall j \in [1, M] \quad (2)$$
Diffusion Approximation

- Informal derivation, \( \tau \)-leaping:
- Choose \( \tau > 0 \) such that:

\[
f_j(x_{t'}, \theta) \approx f_j(x_t, \theta), \quad \forall t' \in [t, t + \tau], \forall j \in [1, M] \tag{1}
\]

\[
f_j(x_t, \theta) \tau \gg 1, \quad \forall j \in [1, M] \tag{2}
\]

- Conditions (1) and (2) can be satisfied if \( x_i \gg 1 \).
Diffusion Approximation

• Informal derivation, $\tau$-leaping:
• Choose $\tau > 0$ such that:

$$
f_j(x_{t'}, \theta) \approx f_j(x_t, \theta), \quad \forall t' \in [t, t + \tau], \forall j \in [1, M] \quad (1)
$$

$$
f_j(x_t, \theta) \tau \gg 1, \quad \forall j \in [1, M] \quad (2)
$$

• Conditions (1) and (2) can be satisfied if $x_i \gg 1$.
• (1) implies that the number of transitions to states $j$ are independently Poisson distributed with mean $f_j(x_t, \theta)\tau$. 

Langevin Equation

$$
dx_t = Sf(x_t, \theta) \, dt + 1/\sqrt{\Omega} S\sqrt{\text{diag}}(f(x_t, \theta)) \, dB_t \quad (3)
$$
Diffusion Approximation

• Informal derivation, $\tau$-leaping:
• Choose $\tau > 0$ such that:

$$f_j(x_{t'}, \theta) \approx f_j(x_t, \theta), \quad \forall t' \in [t, t + \tau], \forall j \in [1, M]$$  \hspace{1cm} (1)

$$f_j(x_t, \theta) \tau \gg 1, \quad \forall j \in [1, M]$$  \hspace{1cm} (2)

• Conditions (1) and (2) can be satisfied if $x_i \gg 1$.
• (1) implies that the number of transitions to states $j$ are independently Poisson distributed with mean $f_j(x_t, \theta) \tau$.
• (2) implies that the number of transitions can be reasonably approximated by a Normal distribution.
Diffusion Approximation

• Informal derivation, $\tau$-leaping:
• Choose $\tau > 0$ such that:

$$f_j(x_{t'}, \theta) \approx f_j(x_t, \theta), \quad \forall t' \in [t, t + \tau], \forall j \in [1, M]$$ (1)

$$f_j(x_t, \theta) \tau \gg 1, \quad \forall j \in [1, M]$$ (2)

• Conditions (1) and (2) can be satisfied if $x_i \gg 1$.
• (1) implies that the number of transitions to states $j$ are independently Poisson distributed with mean $f_j(x_t, \theta) \tau$.
• (2) implies that the number of transitions can be reasonably approximated by a Normal distribution.

Langevin Equation

$$dx_t = Sf(x_t, \theta)dt + \frac{1}{\sqrt{\Omega}}S\sqrt{\text{diag}(f(x_t, \theta))}dB_t$$ (3)
Diffusion Approximation

- Informal derivation, $\tau$-leaping:
- Choose $\tau > 0$ such that:

\[ f_j(x_{t'}, \theta) \approx f_j(x_t, \theta), \quad \forall t' \in [t, t + \tau], \forall j \in [1, M] \quad (1) \]

\[ f_j(x_t, \theta) \tau \gg 1, \quad \forall j \in [1, M] \quad (2) \]

- Conditions (1) and (2) can be satisfied if $x_i \gg 1$.
- (1) implies that the number of transitions to states $j$ are independently Poisson distributed with mean $f_j(x_t, \theta)\tau$.
- (2) implies that the number of transitions can be reasonably approximated by a Normal distribution.

Langevin Equation

\[ dx_t = Sf(x_t, \theta)dt + \frac{1}{\sqrt{\Omega}}S\sqrt{\text{diag}(f(x_t, \theta))}dB_t \quad (3) \]
Linear Noise Approximation

• Let \( x = \phi + \frac{1}{\sqrt{\Omega}} \xi \) where \( \phi \) are deterministic and \( \xi \) stochastic variables.
Linear Noise Approximation

- Let $x = \phi + \frac{1}{\sqrt{\Omega}} \xi$ where $\phi$ are deterministic and $\xi$ stochastic variables.
- Rewrite the transition functions and Taylor expand around $\phi$
Linear Noise Approximation

- Let $x = \phi + \frac{1}{\sqrt{\Omega}} \xi$ where $\phi$ are deterministic and $\xi$ stochastic variables.
- Rewrite the transition functions and Taylor expand around $\phi$.
- Replace in the diffusion approximation retaining $O(1)$ for $d\phi$.

$$d\phi_t = Sf(\phi_t, \theta)dt$$
Linear Noise Approximation

- Let \( x = \phi + \frac{1}{\sqrt{\Omega}} \xi \) where \( \phi \) are deterministic and \( \xi \) stochastic variables.
- Rewrite the transition functions and Taylor expand around \( \phi \)
- Replace in the diffusion approximation retaining \( O(1) \) for \( d\phi \).

\[
d\phi_t = Sf(\phi_t, \theta)dt
\]

- Neglect any terms higher than \( O\left( \frac{1}{\sqrt{\Omega}} \right) \) for \( d\xi \)

\[
d\xi_t = SJf(\phi, \theta)\xi dt + S\sqrt{\text{diag}(f(\phi_t, \theta))}dB_t
\]
Linear Noise Approximation

• Let \( x = \phi + \frac{1}{\sqrt{\Omega}} \xi \) where \( \phi \) are deterministic and \( \xi \) stochastic variables.

• Rewrite the transition functions and Taylor expand around \( \phi \)

• Replace in the diffusion approximation retaining \( O(1) \) for \( d\phi \).

\[ d\phi_t = Sf(\phi_t, \theta)dt \]

• Neglect any terms higher than \( O(\frac{1}{\sqrt{\Omega}}) \) for \( d\xi \)

\[ d\xi_t = SJ_f(\phi, \theta)\xi dt + S\sqrt{\text{diag}(f(\phi, \theta))} dB_t \]

• which is a linear SDE with solution

\[ \xi_t = \Phi(t_0, t) \left( \xi_0 + \int_{t_0}^{t} \Phi(s, t)^{-1} S\sqrt{\text{diag}(f(\phi_s, \theta))} dB_s \right) \]

• where \( \Phi(t_0, s) \) the solution to

\[ d\Phi(t_0, s) = SJ_f(\phi_s, \theta)\Phi(t_0, s)ds, \Phi(t_0, t_0) = I \]
Likelihood for the Linear Noise Approximation

- \( x^{(TS)} = \{ x_{t_1}, \ldots, x_{t_n} \}^T \) an \( nN \) vector observed sample path.
Likelihood for the Linear Noise Approximation

- \( x^{(TS)} = \{x_{t_1}, \ldots, x_{t_n}\}^T \) an \( nN \) vector observed sample path.
- From the Markov property we have that

\[
p(x^{(TS)}|\theta) = \prod_{i=1}^{n} p(x_{t_i}|x_{t_{i-1}}, \theta)p(x_{t_0}|\theta) \propto \mathcal{N}(\mu(\theta), \Sigma(\theta))
\]
Likelihood for the Linear Noise Approximation

- \( x^{(TS)} = \{ x_{t_1}, \ldots, x_{t_n} \}^T \) an \( nN \) vector observed sample path.
- From the Markov property we have that

\[
p(x^{(TS)}|\theta) = \prod_{i=1}^{n} p(x_{t_i}|x_{t_{i-1}}, \theta) p(x_{t_0}|\theta) \propto \mathcal{N}(\mu(\theta), \Sigma(\theta))
\]

- \( \mu(\theta) = (\phi_{t_1}, \ldots, \phi_{t_n})^T \) a \( nN \) vector with solutions of the MRE.
Likelihood for the Linear Noise Approximation

- $\mathbf{x}^{(TS)} = \{ \mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_n} \}^T$ an $nN$ vector observed sample path.
- From the Markov property we have that

$$ p(\mathbf{x}^{(TS)} | \theta) = \prod_{i=1}^{n} p(\mathbf{x}_{t_i} | \mathbf{x}_{t_{i-1}}, \theta) p(\mathbf{x}_{t_0} | \theta) \propto \mathcal{N}(\mu(\theta), \Sigma(\theta)) $$

- $\mu(\theta) = (\phi_{t_1}, \ldots, \phi_{t_n})^T$ a $nN$ vector with solutions of the MRE.
- Metric tensor is expected Fisher Information

$$ Fl(\theta)_{m,n} = \frac{\partial \mu(\theta)^T}{\partial \theta_m} \Sigma^{-1}(\theta) \frac{\partial \mu(\theta)}{\partial \theta_n} + \frac{1}{2} \text{tr} \left( \Sigma^{-1}(\theta) \frac{\partial \Sigma(\theta)}{\partial \theta_m} \Sigma^{-1}(\theta) \frac{\partial \Sigma(\theta)}{\partial \theta_n} \right) $$
Likelihood for the Linear Noise Approximation

- \( x^{(TS)} = \{x_{t_1}, \ldots, x_{t_n}\}^T \) an \( nN \) vector observed sample path.
- From the Markov property we have that
  \[
p(x^{(TS)}|\theta) = \prod_{i=1}^{n} p(x_{t_i}|x_{t_{i-1}}, \theta)p(x_{t_0}|\theta) \propto \mathcal{N}(\mu(\theta), \Sigma(\theta))
  \]
- \( \mu(\theta) = (\phi_{t_1}, \ldots, \phi_{t_n})^T \) a \( nN \) vector with solutions of the MRE.
- Metric tensor is expected Fisher Information
  \[
  Fl(\theta)_{m,n} = \frac{\partial \mu(\theta)}{\partial \theta_m}^T \Sigma^{-1}(\theta) \frac{\partial \mu(\theta)}{\partial \theta_n} + \frac{1}{2} \text{tr} \left( \Sigma^{-1}(\theta) \frac{\partial \Sigma(\theta)}{\partial \theta_m} \Sigma^{-1}(\theta) \frac{\partial \Sigma(\theta)}{\partial \theta_n} \right)
  \]
Likelihood for the Linear Noise Approximation

- $x^{(TS)} = \{x_{t_1}, \ldots, x_{t_n}\}^T$ an nN vector observed sample path.
- From the Markov property we have that
  $$p(x^{(TS)}|\theta) = \prod_{i=1}^n p(x_{t_i}|x_{t_{i-1}}, \theta)p(x_{t_0}|\theta) \propto \mathcal{N}(\mu(\theta), \Sigma(\theta))$$
- $\mu(\theta) = (\phi_{t_1}, \ldots, \phi_{t_n})^T$ a nN vector with solutions of the MRE.
- Metric tensor is expected Fisher Information
  $$F_l(\theta)_{m,n} = \frac{\partial \mu(\theta)^T}{\partial \theta_m} \Sigma^{-1}(\theta) \frac{\partial \mu(\theta)}{\partial \theta_n} + \frac{1}{2} \text{tr} \left( \Sigma^{-1}(\theta) \frac{\partial \Sigma(\theta)}{\partial \theta_m} \Sigma^{-1}(\theta) \frac{\partial \Sigma(\theta)}{\partial \theta_n} \right)$$
- For computation we augment the MRE for $\phi$ with the lower triangular elements of $V$ and solve the augmented system with forward sensitivity analysis.
Single gene expression model

- Single gene expression autoregulation.
Single gene expression model

- Single gene expression autoregulation.
- System state $(R(t), P(t))^T$ models the population of RNA and protein.
Single gene expression model

- Single gene expression autoregulation.
- System state \((R(t), P(t))^T\) models the population of RNA and protein.
- \(k_R(P, t) = (b_0 \exp(-b_1(t - b_2)^2) + b_3)/(1 + (P/H)^{n_H})\)
- \(H = b_3 k_P/(2\gamma_R\gamma_P), \quad n_H = 1/2\)
Single gene expression model

- Single gene expression autoregulation.
- System state \((R(t), P(t))^T\) models the population of RNA and protein.
- \(k_R(P, t) = (b_0 \exp(-b_1(t - b_2)^2) + b_3)/(1 + (P/H)^{n_H})\)
- \(H = b_3 k_P/(2\gamma_R\gamma_P)\), \(n_H = 1/2\)

\[
S = \begin{pmatrix}
1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1 \\
\end{pmatrix}
\]
Single gene expression model

- Single gene expression autoregulation.
- System state \((R(t), P(t))^T\) models the population of RNA and protein.
- \(k_R(P, t) = (b_0 \exp(-b_1(t - b_2)^2) + b_3)/(1 + (P/H)^n_H)\)
- \(H = b_3 k_P / (2\gamma_R \gamma_P)\), \(n_H = 1/2\)

\[
S = \begin{pmatrix}
1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1 \\
\end{pmatrix}
\]

- Deterministic MRE \(\phi = (\phi_R, \phi_P)^T\)
  \[
d\phi_R/dt = k_R(\phi_P, t) - \gamma_R \phi_R \\
d\phi_P/dt = k_P \phi_R - \gamma_P \phi_P
\]
Single gene expression model

- Single gene expression autoregulation.
- System state \((R(t), P(t))^T\) models the population of RNA and protein.
- \(k_R(P, t) = (b_0 \exp(-b_1(t - b_2)^2) + b_3)/(1 + (P/H)^{n_H})\)
- \(H = b_3 k_P/(2\gamma_R\gamma_P), \quad n_H = 1/2\)

\[
S = \begin{pmatrix}
1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1
\end{pmatrix}
\]

- Deterministic MRE \(\phi = (\phi_R, \phi_P)^T\)
  \[
d\phi_R/dt = k_R(\phi_P, t) - \gamma_R\phi_R \\
d\phi_P/dt = k_P\phi_R - \gamma_P\phi_P
\]

- Parameters \(\theta = (\gamma_R, \gamma_P, k_P, b_0, b_1, b_2, b_3)^T\)
Simulated Data

- Simulated data generated with SSA.
- 10 independent sample paths for each time point.
- Parameters set to

\[
\begin{array}{ccccccc}
\gamma_R & \gamma_P & k_P & b_0 & b_1 & b_2 & b_3 \\
0.44 & 0.52 & 10.0 & 15.0 & 0.40 & 7.0 & 3.0
\end{array}
\]
Trace Plots

RMHMC

SMMALA

CWMH
### Effective Sample Size

10,000 posterior samples

<table>
<thead>
<tr>
<th>Method</th>
<th>$\gamma_R$</th>
<th>$\gamma_P$</th>
<th>$k_P$</th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMHMC</td>
<td>6532</td>
<td>6593</td>
<td>6614</td>
<td>5112</td>
<td>5384</td>
<td>6595</td>
<td>6642</td>
</tr>
<tr>
<td>SMMALA</td>
<td>2990</td>
<td>3270</td>
<td>3454</td>
<td>3124</td>
<td>3164</td>
<td>3316</td>
<td>3195</td>
</tr>
<tr>
<td>CWMH</td>
<td>201</td>
<td>71</td>
<td>73</td>
<td>465</td>
<td>339</td>
<td>420</td>
<td>239</td>
</tr>
</tbody>
</table>
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC

- Promising capability of methodology
- Ongoing development
- Potential bottleneck at metric tensor and Christoffel symbols
- Theoretical analysis of convergence
- Orthogonal Foliations for Hierarchical models
- Investigate alternative manifold structures
- Design and effect of metric and connection
- Optimality of Hamiltonian flows as local geodesics
- Alternative transition kernels

- No silver bullet or cure all - new powerful methodology for MC toolkit
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
- Promising capability of methodology
- Ongoing development
  - Potential bottleneck at metric tensor and Christoffel symbols
  - Theoretical analysis of convergence
- Orthogonal Foliations for Hierarchical models
- Investigate alternative manifold structures
- Design and effect of metric and connection
- Optimality of Hamiltonian flows as local geodesics
- Alternative transition kernels
  - No silver bullet or cure all - new powerful methodology for MC toolkit
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
  - Promising capability of methodology
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
    - Promising capability of methodology
- Ongoing development
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
  - Promising capability of methodology

- Ongoing development
  - Potential bottleneck at metric tensor and Christoffel symbols

- Theoretical analysis of convergence
- Orthogonal Foliations for Hierarchical models
- Investigate alternative manifold structures
- Design and effect of metric and connection
- Optimality of Hamiltonian flows as local geodesics
- Alternative transition kernels

No silver bullet or cure all - new powerful methodology for MC toolkit
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
  - Promising capability of methodology

- Ongoing development
  - Potential bottleneck at metric tensor and Christoffel symbols
  - Theoretical analysis of convergence
Conclusion and Discussion

• Geometry of statistical models harnessed in Monte Carlo methods
  • Diffusions that respect structure and curvature of space - Manifold MALA
  • Geodesic flows on model manifold - RMHMC - generalisation of HMC
  • Assessed on correlated & high-dimensional latent variable models
  • Promising capability of methodology

• Ongoing development
  • Potential bottleneck at metric tensor and Christoffel symbols
  • Theoretical analysis of convergence
  • Orthogonal Foliations for Hierarchical models
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
  - Promising capability of methodology

- Ongoing development
  - Potential bottleneck at metric tensor and Christoffel symbols
  - Theoretical analysis of convergence
  - Orthogonal Foliations for Hierarchical models
  - Investigate alternative manifold structures

- Design and effect of metric and connection
- Optimality of Hamiltonian flows as local geodesics
- Alternative transition kernels

No silver bullet or cure all - new powerful methodology for MC toolkit
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
  - Promising capability of methodology
- Ongoing development
  - Potential bottleneck at metric tensor and Christoffel symbols
  - Theoretical analysis of convergence
  - Orthogonal Foliations for Hierarchical models
  - Investigate alternative manifold structures
  - Design and effect of metric and connection
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
  - Promising capability of methodology

- Ongoing development
  - Potential bottleneck at metric tensor and Christoffel symbols
  - Theoretical analysis of convergence
  - Orthogonal Foliations for Hierarchical models
  - Investigate alternative manifold structures
  - Design and effect of metric and connection
  - Optimality of Hamiltonian flows as local geodesics

- Alternative transition kernels
- No silver bullet or cure all - new powerful methodology for MC toolkit
Conclusion and Discussion

- Geometry of statistical models harnessed in Monte Carlo methods
  - Diffusions that respect structure and curvature of space - Manifold MALA
  - Geodesic flows on model manifold - RMHMC - generalisation of HMC
  - Assessed on correlated & high-dimensional latent variable models
  - Promising capability of methodology

- Ongoing development
  - Potential bottleneck at metric tensor and Christoffel symbols
  - Theoretical analysis of convergence
  - Orthogonal Foliations for Hierarchical models
  - Investigate alternative manifold structures
  - Design and effect of metric and connection
  - Optimality of Hamiltonian flows as local geodesics
  - Alternative transition kernels

No silver bullet or cure all - new powerful methodology for MC toolkit
Conclusion and Discussion

• Geometry of statistical models harnessed in Monte Carlo methods
  • Diffusions that respect structure and curvature of space - Manifold MALA
  • Geodesic flows on model manifold - RMHMC - generalisation of HMC
  • Assessed on correlated & high-dimensional latent variable models
  • Promising capability of methodology

• Ongoing development
  • Potential bottleneck at metric tensor and Christoffel symbols
  • Theoretical analysis of convergence
  • Orthogonal Foliations for Hierarchical models
  • Investigate alternative manifold structures
  • Design and effect of metric and connection
  • Optimality of Hamiltonian flows as local geodesics
  • Alternative transition kernels

• No silver bullet or cure all - new powerful methodology for MC toolkit
Funding Acknowledgment

- EPSRC Advanced Research Fellowship and BBSRC project grant, Microsoft Research Cambridge