MCMC Sampling for Intractable MJP Models of Chemical Kinetics via the Linear Noise Approximation

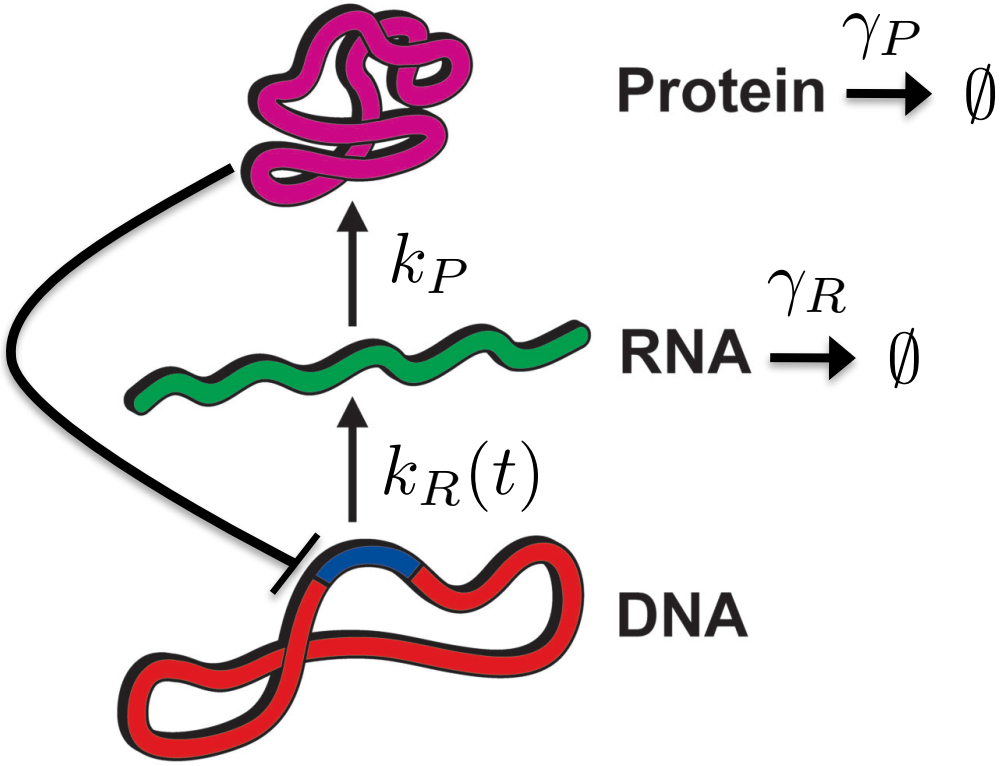
Mark A. Girolami

Department of Statistical Science
University College London

June 2012

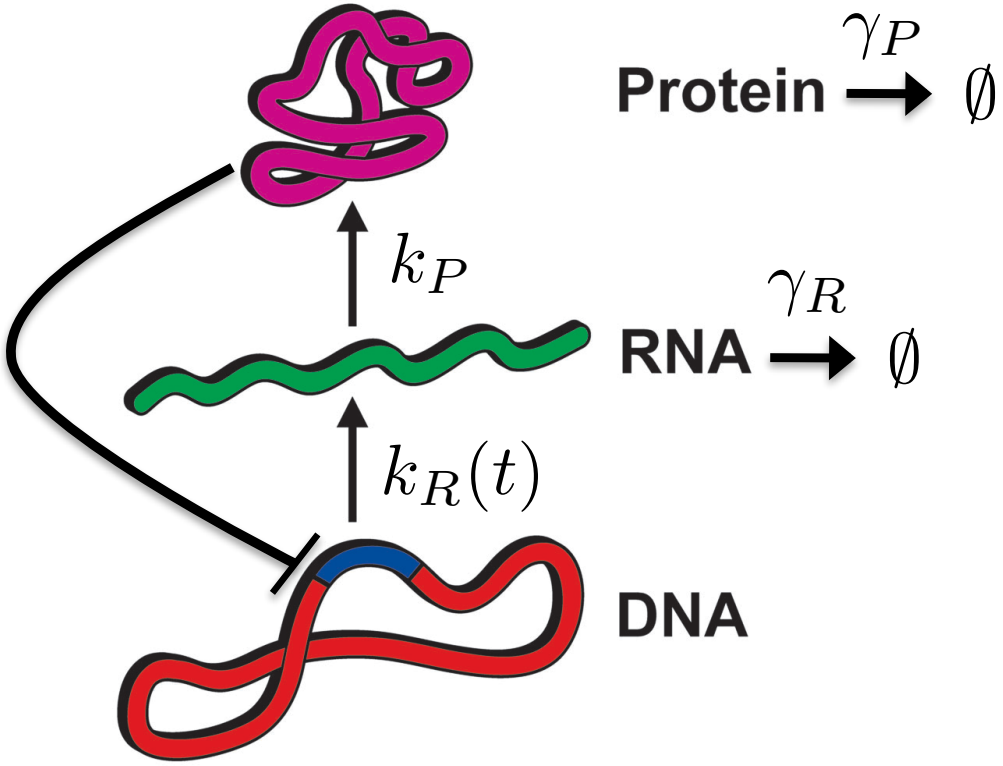
Outline

Single gene expression model



- Gene autoregulation.

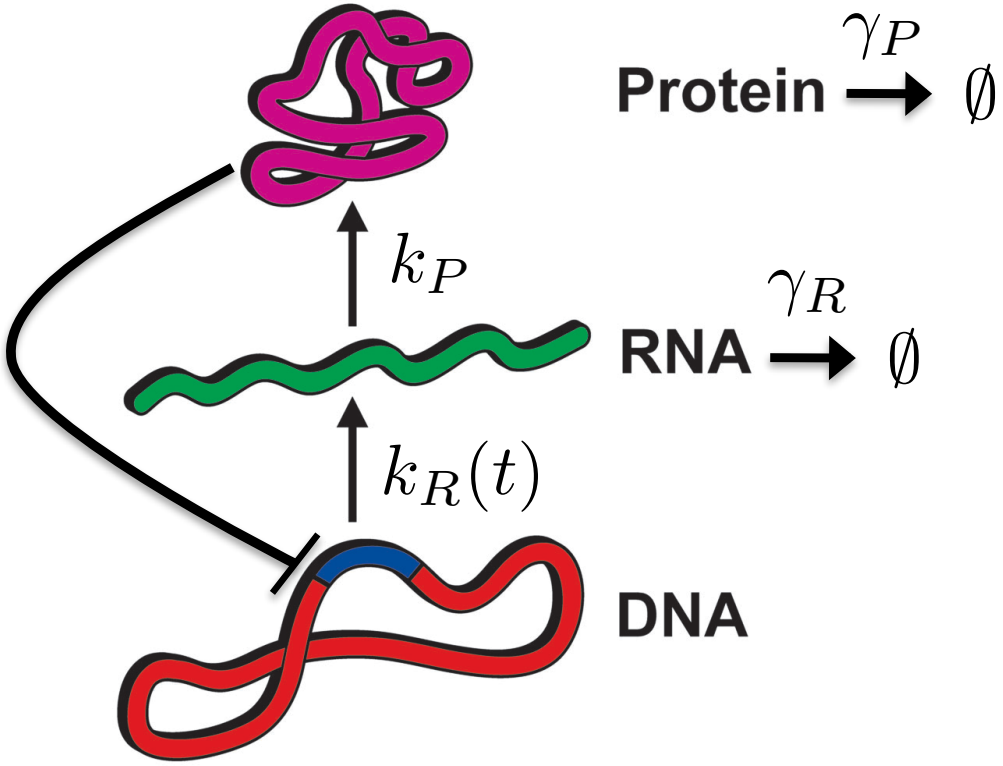
Single gene expression model



- Gene autoregulation.

$$
\begin{aligned}
R1 : DNA &\xrightarrow{k_R(t)} DNA + R \\
R2 : R &\xrightarrow{\gamma_R} \emptyset \\
R3 : R &\xrightarrow{k_P} R + P \\
R4 : P &\xrightarrow{\gamma_P} \emptyset.
\end{aligned}
$$

Single gene expression model

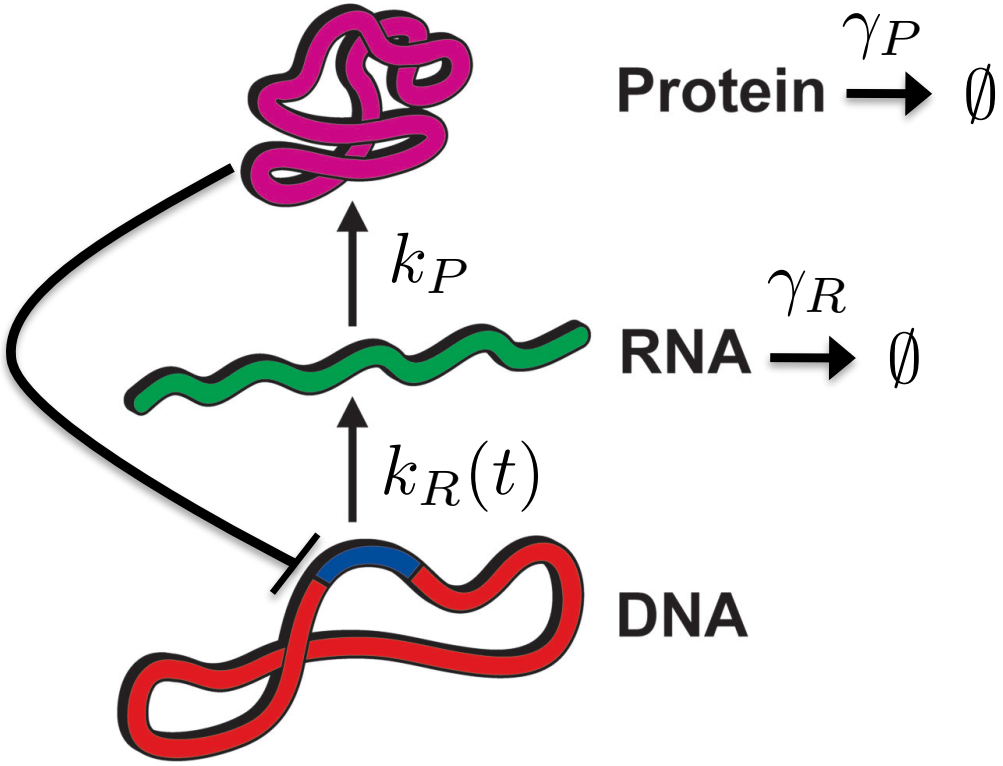


- Gene autoregulation.

$$
\begin{aligned}
R1 : DNA &\xrightarrow{k_R(t)} DNA + R \\
R2 : R &\xrightarrow{\gamma_R} \emptyset \\
R3 : R &\xrightarrow{k_P} R + P \\
R4 : P &\xrightarrow{\gamma_P} \emptyset.
\end{aligned}
$$

- System state $R(t), P(t))$

Single gene expression model



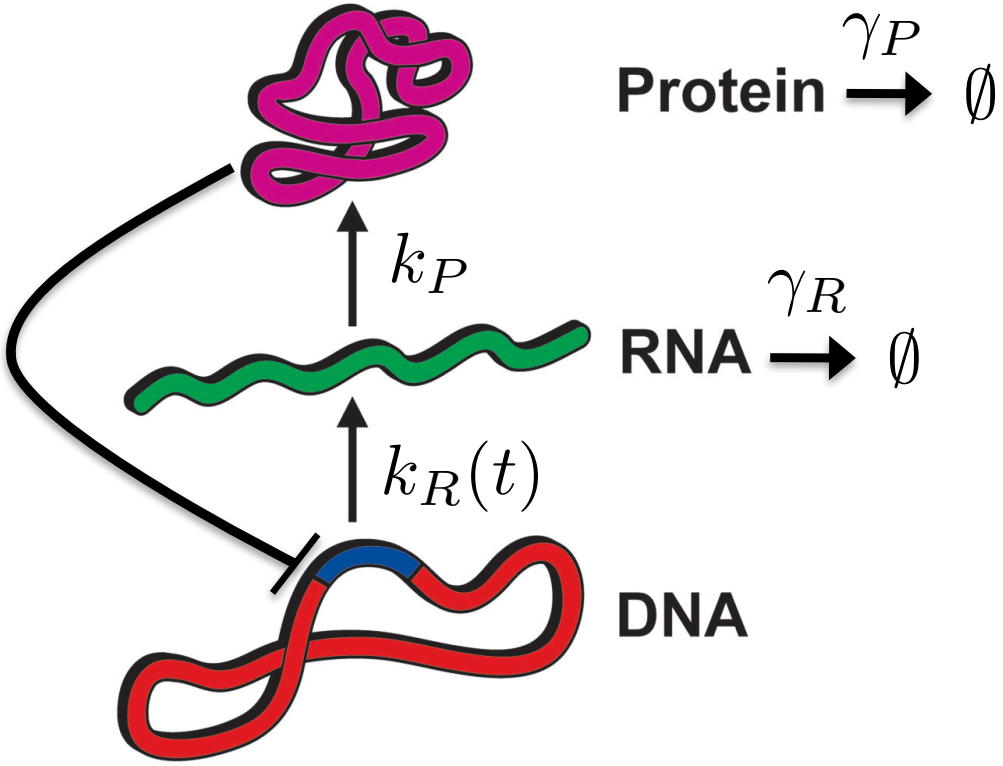
- Gene autoregulation.

$$
\begin{aligned}
R1 : DNA &\xrightarrow{k_R(t)} DNA + R \\
R2 : R &\xrightarrow{\gamma_R} \emptyset \\
R3 : R &\xrightarrow{k_P} R + P \\
R4 : P &\xrightarrow{\gamma_P} \emptyset.
\end{aligned}
$$

- System state $R(t), P(t))$

$$
\boldsymbol{S} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.
$$

Single gene expression model



- Gene autoregulation.

$$
\begin{aligned}
R1 : DNA &\xrightarrow{k_R(t)} DNA + R \\
R2 : R &\xrightarrow{\gamma_R} \emptyset \\
R3 : R &\xrightarrow{k_P} R + P \\
R4 : P &\xrightarrow{\gamma_P} \emptyset.
\end{aligned}
$$

- System state $R(t), P(t))$

$$
\boldsymbol{S} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.
$$

- $\boldsymbol{\theta} = (\gamma_R, \gamma_P, k_P, b_0, b_1, b_2, b_3)$

Motivation

- Continuous stochastic process on a discrete state space.

Motivation

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

Motivation

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

Motivation

- 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, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_0) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$

Motivation

- 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, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_0) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$
- $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{N,t})^T$ system state at time t for N random variables.

Motivation

- 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, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$
- $\boldsymbol{x}_t = (x_{1,t},\ldots,x_{N,t})^T$ system state at time t for N random variables.
- State change vectors $\boldsymbol{s}_j = (s_{1,j},\ldots,s_{N,j})^T,\ j \in \{1,\ldots,M\}$

Motivation

- 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, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$
- $\boldsymbol{x}_t = (x_{1,t},\ldots,x_{N,t})^T$ system state at time t for N random variables.
- State change vectors $\boldsymbol{s}_j = (s_{1,j},\ldots,s_{N,j})^T,\ j \in \{1,\ldots,M\}$
- Transition rates, $f_j(\boldsymbol{x},\boldsymbol{\theta})dt$ for state change j in the interval $[t, t+dt)$.

Motivation

- 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, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$
- $\boldsymbol{x}_t = (x_{1,t},\ldots,x_{N,t})^T$ system state at time t for N random variables.
- State change vectors $\boldsymbol{s}_j = (s_{1,j},\ldots,s_{N,j})^T,\ j \in \{1,\ldots,M\}$
- Transition rates, $f_j(\boldsymbol{x},\boldsymbol{\theta})dt$ for state change j in the interval $[t, t+dt)$.
- Master Equation

$$\frac{dp(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)}{dt} = \sum_{j=1}^{M}\left[f_j(\boldsymbol{x}-\boldsymbol{s}_j,\boldsymbol{\theta},t)p(\boldsymbol{x}-\boldsymbol{s}_j,t|\boldsymbol{x}_0,t_0) - f_j(\boldsymbol{x},\boldsymbol{\theta},t)p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right].$$

Motivation

- 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, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$
- $\boldsymbol{x}_t = (x_{1,t},\ldots,x_{N,t})^T$ system state at time t for N random variables.
- State change vectors $\boldsymbol{s}_j = (s_{1,j},\ldots,s_{N,j})^T,\ j \in \{1,\ldots,M\}$
- Transition rates, $f_j(\boldsymbol{x},\boldsymbol{\theta})dt$ for state change j in the interval $[t, t+dt)$.
- Master Equation

$$\frac{dp(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)}{dt} = \sum_{j=1}^{M}\left[f_j(\boldsymbol{x}-\boldsymbol{s}_j,\boldsymbol{\theta},t)p(\boldsymbol{x}-\boldsymbol{s}_j,t|\boldsymbol{x}_0,t_0) - f_j(\boldsymbol{x},\boldsymbol{\theta},t)p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right].$$

- Intractible, simulate trajectories by a Stochastic Simulation Algorithm.

Motivation

- 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, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$
- $\boldsymbol{x}_t = (x_{1,t},\ldots,x_{N,t})^T$ system state at time t for N random variables.
- State change vectors $\boldsymbol{s}_j = (s_{1,j},\ldots,s_{N,j})^T,\ j \in \{1,\ldots,M\}$
- Transition rates, $f_j(\boldsymbol{x},\boldsymbol{\theta})dt$ for state change j in the interval $[t, t+dt)$.
- Master Equation

$$\frac{dp(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)}{dt} = \sum_{j=1}^{M}\left[f_j(\boldsymbol{x}-\boldsymbol{s}_j,\boldsymbol{\theta},t)p(\boldsymbol{x}-\boldsymbol{s}_j,t|\boldsymbol{x}_0,t_0) - f_j(\boldsymbol{x},\boldsymbol{\theta},t)p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right].$$

- Intractible, simulate trajectories by a Stochastic Simulation Algorithm.
- Parameter inference scheme Boys, Wilkinson, Kirkwood (2006)

Diffusion Approximation

- Employed by Golightly and Wilkinson (2005, 2008)

Diffusion Approximation

- Employed by Golightly and Wilkinson (2005, 2008)
- Ito diffusion associated with Fokker-Planck equation

Diffusion Approximation

- Employed by Golightly and Wilkinson (2005, 2008)
- Ito diffusion associated with Fokker-Planck equation
- Informal derivation, τ-leaping:

Diffusion Approximation

- Employed by Golightly and Wilkinson (2005, 2008)
- Ito diffusion associated with Fokker-Planck equation
- Informal derivation, τ-leaping:
- Choose $\tau > 0$ such that:

$$f_j(\boldsymbol{x}_{t'}, \boldsymbol{\theta}) \approx f_j(\boldsymbol{x}_t, \boldsymbol{\theta}), \quad \forall t' \in [t, t+\tau],\ \forall j \in [1, M] \tag{1}$$

$$f_j(\boldsymbol{x}_t, \boldsymbol{\theta})\tau \gg 1, \qquad \forall j \in [1, M] \tag{2}$$

Diffusion Approximation

- Employed by Golightly and Wilkinson (2005, 2008)
- Ito diffusion associated with Fokker-Planck equation
- Informal derivation, τ-leaping:
- Choose $\tau > 0$ such that:

$$f_j(\boldsymbol{x}_{t'}, \boldsymbol{\theta}) \approx f_j(\boldsymbol{x}_t, \boldsymbol{\theta}), \quad \forall t' \in [t, t+\tau],\ \forall j \in [1, M] \tag{1}$$

$$f_j(\boldsymbol{x}_t, \boldsymbol{\theta})\tau \gg 1, \qquad \forall j \in [1, M] \tag{2}$$

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

Diffusion Approximation

- Employed by Golightly and Wilkinson (2005, 2008)
- Ito diffusion associated with Fokker-Planck equation
- Informal derivation, τ-leaping:
- Choose $\tau > 0$ such that:

$$f_j(\boldsymbol{x}_{t'}, \boldsymbol{\theta}) \approx f_j(\boldsymbol{x}_t, \boldsymbol{\theta}), \quad \forall t' \in [t, t+\tau],\ \forall j \in [1, M] \tag{1}$$

$$f_j(\boldsymbol{x}_t, \boldsymbol{\theta})\tau \gg 1, \qquad \forall j \in [1, M] \tag{2}$$

- 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 $f_j(\boldsymbol{x}_t, \boldsymbol{\theta})\tau$.

Diffusion Approximation

- Employed by Golightly and Wilkinson (2005, 2008)
- Ito diffusion associated with Fokker-Planck equation
- Informal derivation, τ-leaping:
- Choose $\tau > 0$ such that:

$$f_j(\boldsymbol{x}_{t'}, \boldsymbol{\theta}) \approx f_j(\boldsymbol{x}_t, \boldsymbol{\theta}), \quad \forall t' \in [t, t+\tau],\ \forall j \in [1, M] \tag{1}$$

$$f_j(\boldsymbol{x}_t, \boldsymbol{\theta})\tau \gg 1, \qquad \forall j \in [1, M] \tag{2}$$

- 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 $f_j(\boldsymbol{x}_t, \boldsymbol{\theta})\tau$.
- (2) implies that the number of transitions can be reasonably approximated by a Normal distribution.

Diffusion Approximation

- Employed by Golightly and Wilkinson (2005, 2008)
- Ito diffusion associated with Fokker-Planck equation
- Informal derivation, τ-leaping:
- Choose $\tau > 0$ such that:

$$f_j(\boldsymbol{x}_{t'}, \boldsymbol{\theta}) \approx f_j(\boldsymbol{x}_t, \boldsymbol{\theta}), \quad \forall t' \in [t, t+\tau],\ \forall j \in [1, M] \tag{1}$$

$$f_j(\boldsymbol{x}_t, \boldsymbol{\theta})\tau \gg 1, \qquad \forall j \in [1, M] \tag{2}$$

- 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 $f_j(\boldsymbol{x}_t, \boldsymbol{\theta})\tau$.
- (2) implies that the number of transitions can be reasonably approximated by a Normal distribution.
- Langevin Equation

$$d\boldsymbol{x}_t = \boldsymbol{S}\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{\theta})dt + \frac{1}{\sqrt{\Omega}}\boldsymbol{S}\sqrt{\mathrm{diag}(\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{\theta}))}dB_t \tag{3}$$

Diffusion Approximation

- Inference proceeds employing Euler-Maryuama discretision

Diffusion Approximation

- Inference proceeds employing Euler-Maryuama discretision
- Augmentation of observed values with 'missing values' Roberts and Stramer, 2001, Golightly and Wilkinson, 2005

Diffusion Approximation

- Inference proceeds employing Euler-Maryuama discretision
- Augmentation of observed values with 'missing values' Roberts and Stramer, 2001, Golightly and Wilkinson, 2005
- Sample a skeleton path then parameters of interest

Diffusion Approximation

- Inference proceeds employing Euler-Maryuama discretision
- Augmentation of observed values with 'missing values' Roberts and Stramer, 2001, Golightly and Wilkinson, 2005
- Sample a skeleton path then parameters of interest
- Issues of efficiency given the coupling of discrete path sampled and parameters

Diffusion Approximation

- Inference proceeds employing Euler-Maryuama discretision
- Augmentation of observed values with 'missing values' Roberts and Stramer, 2001, Golightly and Wilkinson, 2005
- Sample a skeleton path then parameters of interest
- Issues of efficiency given the coupling of discrete path sampled and parameters
- If system is close to thermodynamic limit further approximation valid

Diffusion Approximation

- Inference proceeds employing Euler-Maryuama discretision
- Augmentation of observed values with 'missing values' Roberts and Stramer, 2001, Golightly and Wilkinson, 2005
- Sample a skeleton path then parameters of interest
- Issues of efficiency given the coupling of discrete path sampled and parameters
- If system is close to thermodynamic limit further approximation valid
- Linear Noise Approximation (LNA) of van Kampen 1976

Diffusion Approximation

- Inference proceeds employing Euler-Maryuama discretision
- Augmentation of observed values with 'missing values' Roberts and Stramer, 2001, Golightly and Wilkinson, 2005
- Sample a skeleton path then parameters of interest
- Issues of efficiency given the coupling of discrete path sampled and parameters
- If system is close to thermodynamic limit further approximation valid
- Linear Noise Approximation (LNA) of van Kampen 1976
- Employed extensively in chemical physics and genome research

Diffusion Approximation

- Inference proceeds employing Euler-Maryuama discretision
- Augmentation of observed values with 'missing values' Roberts and Stramer, 2001, Golightly and Wilkinson, 2005
- Sample a skeleton path then parameters of interest
- Issues of efficiency given the coupling of discrete path sampled and parameters
- If system is close to thermodynamic limit further approximation valid
- Linear Noise Approximation (LNA) of van Kampen 1976
- Employed extensively in chemical physics and genome research
- May provide intermediate scheme for MCMC based inference

Linear Noise Approximation

- **Assume** that $\boldsymbol{x} = \phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}$

Linear Noise Approximation

- **Assume** that $\boldsymbol{x} = \phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}$
- ϕ are deterministic and $\boldsymbol{\xi}$ stochastic variables.

Linear Noise Approximation

- **Assume** that $\boldsymbol{x} = \phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}$
- ϕ are deterministic and $\boldsymbol{\xi}$ stochastic variables.

$$f_j(\boldsymbol{x}, \boldsymbol{\theta}) = f_j\left(\phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}, \boldsymbol{\theta}\right) = f_j(\phi, \boldsymbol{\theta}) + \frac{1}{\sqrt{\Omega}}\sum_{i=1}^{N}\frac{\partial f_j(\phi, \boldsymbol{\theta})}{\partial \phi_i}\xi_i + O(\Omega^{-1})$$

Linear Noise Approximation

- **Assume** that $\boldsymbol{x} = \phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}$
- ϕ are deterministic and $\boldsymbol{\xi}$ stochastic variables.

$$f_j(\boldsymbol{x}, \boldsymbol{\theta}) = f_j\left(\phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}, \boldsymbol{\theta}\right) = f_j(\phi, \boldsymbol{\theta}) + \frac{1}{\sqrt{\Omega}}\sum_{i=1}^{N}\frac{\partial f_j(\phi, \boldsymbol{\theta})}{\partial \phi_i}\xi_i + O(\Omega^{-1})$$

- Replace in the diffusion retain $O(1)$ terms for $d\phi$.

$$d\phi_t = \boldsymbol{S}\boldsymbol{f}(\phi_t, \boldsymbol{\theta})dt$$

Linear Noise Approximation

- **Assume** that $\boldsymbol{x} = \phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}$
- ϕ are deterministic and $\boldsymbol{\xi}$ stochastic variables.

$$f_j(\boldsymbol{x}, \boldsymbol{\theta}) = f_j\left(\phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}, \boldsymbol{\theta}\right) = f_j(\phi, \boldsymbol{\theta}) + \frac{1}{\sqrt{\Omega}}\sum_{i=1}^{N}\frac{\partial f_j(\phi, \boldsymbol{\theta})}{\partial \phi_i}\xi_i + O(\Omega^{-1})$$

- Replace in the diffusion retain $O(1)$ terms for $d\phi$.

$$d\phi_t = \boldsymbol{S}\boldsymbol{f}(\phi_t, \boldsymbol{\theta})dt$$

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

$$d\boldsymbol{\xi}_t = \boldsymbol{S}\boldsymbol{J}_{\boldsymbol{f}}(\phi, \boldsymbol{\theta})\boldsymbol{\xi}dt + \boldsymbol{S}\sqrt{\mathrm{diag}(\boldsymbol{f}(\phi_t, \boldsymbol{\theta}))}dB_t$$

Linear Noise Approximation

- **Assume** that $\boldsymbol{x} = \phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}$
- ϕ are deterministic and $\boldsymbol{\xi}$ stochastic variables.

$$f_j(\boldsymbol{x}, \boldsymbol{\theta}) = f_j\left(\phi + \frac{1}{\sqrt{\Omega}}\boldsymbol{\xi}, \boldsymbol{\theta}\right) = f_j(\phi, \boldsymbol{\theta}) + \frac{1}{\sqrt{\Omega}}\sum_{i=1}^{N}\frac{\partial f_j(\phi, \boldsymbol{\theta})}{\partial \phi_i}\xi_i + O(\Omega^{-1})$$

- Replace in the diffusion retain $O(1)$ terms for $d\phi$.

$$d\phi_t = \boldsymbol{S}\boldsymbol{f}(\phi_t, \boldsymbol{\theta})dt$$

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

$$d\boldsymbol{\xi}_t = \boldsymbol{S}\boldsymbol{J}_{\boldsymbol{f}}(\phi, \boldsymbol{\theta})\boldsymbol{\xi}dt + \boldsymbol{S}\sqrt{\text{diag}(\boldsymbol{f}(\phi_t, \boldsymbol{\theta}))}dB_t$$

- which is a linear SDE with analytic solution

$$\boldsymbol{\xi}_t = \boldsymbol{\Phi}(t_0, t)\left(\boldsymbol{\xi}_0 + \int_{t0}^{t}\boldsymbol{\Phi}(s, t)^{-1}\boldsymbol{S}\sqrt{\text{diag}(\boldsymbol{f}(\phi_s, \boldsymbol{\theta}))}dB_s\right)$$

- where $\boldsymbol{\Phi}(t0, s)$ the solution to

$$d\boldsymbol{\Phi}(t0, s) = \boldsymbol{S}\boldsymbol{J}_{\boldsymbol{f}}(\phi_s, \boldsymbol{\theta})\boldsymbol{\Phi}(t0, s)ds, \;\; \boldsymbol{\Phi}(t0, t0) = \boldsymbol{I}$$

Likelihood for the Linear Noise Approximation

$$p(\boldsymbol{x}^{(TS)}|\boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

Likelihood for the Linear Noise Approximation

$$p(\boldsymbol{x}^{(TS)}|\boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

- $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\phi_{t1}, \ldots, \phi_{tn})^T$ a nN vector with solutions of the MRE.

Likelihood for the Linear Noise Approximation

$$p(\boldsymbol{x}^{(TS)}|\boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

- $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\phi_{t1}, \ldots, \phi_{tn})^T$ a nN vector with solutions of the MRE.
- $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ a $nN \times nN$ block matrix with blocks $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j}$ $N \times N$

Likelihood for the Linear Noise Approximation

$$p(\boldsymbol{x}^{(TS)}|\boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

- $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\phi_{t1}, \ldots, \phi_{tn})^T$ a nN vector with solutions of the MRE.
- $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ a $nN \times nN$ block matrix with blocks $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j}$ $N \times N$
- $$\boldsymbol{x}_t \sim \mathcal{N}(\phi_t, \boldsymbol{V}_t), \ \frac{d\boldsymbol{V}_t}{dt} = \boldsymbol{S}\boldsymbol{J}_{\boldsymbol{f}}(\phi_t, \boldsymbol{\theta})\boldsymbol{V}_t + \boldsymbol{V}_t\boldsymbol{J}_{\boldsymbol{f}}(\phi_t, \boldsymbol{\theta})^T\boldsymbol{S}^T + \boldsymbol{S}\mathrm{diag}(\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{\theta}))\boldsymbol{S}^T$$

Likelihood for the Linear Noise Approximation

$$p(\boldsymbol{x}^{(TS)}|\boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

- $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\phi_{t1}, \ldots, \phi_{tn})^T$ a nN vector with solutions of the MRE.
- $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ a $nN \times nN$ block matrix with blocks $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j}$ $N \times N$
- $$\boldsymbol{x}_t \sim \mathcal{N}(\phi_t, \boldsymbol{V}_t), \ \frac{d\boldsymbol{V}_t}{dt} = \boldsymbol{S}\boldsymbol{J}_{\boldsymbol{f}}(\phi_t, \boldsymbol{\theta})\boldsymbol{V}_t + \boldsymbol{V}_t\boldsymbol{J}_{\boldsymbol{f}}(\phi_t, \boldsymbol{\theta})^T\boldsymbol{S}^T + \boldsymbol{S}\text{diag}(\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{\theta}))\boldsymbol{S}^T$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j} = \begin{cases} \boldsymbol{V}_{t_i} & \text{if } i = j \\ cov(\boldsymbol{x}_{t_i}, \boldsymbol{x}_{t_j}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j-1}\boldsymbol{\Phi}(t_{j-1}, t_j)^T & \end{cases}$$

Likelihood for the Linear Noise Approximation

$$p(\boldsymbol{x}^{(TS)}|\boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

- $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\phi_{t1}, \ldots, \phi_{tn})^T$ a nN vector with solutions of the MRE.
- $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ a $nN \times nN$ block matrix with blocks $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j}$ $N \times N$
- $$\boldsymbol{x}_t \sim \mathcal{N}(\phi_t, \boldsymbol{V}_t), \ \frac{d\boldsymbol{V}_t}{dt} = \boldsymbol{S}\boldsymbol{J}_{\boldsymbol{f}}(\phi_t, \boldsymbol{\theta})\boldsymbol{V}_t + \boldsymbol{V}_t\boldsymbol{J}_{\boldsymbol{f}}(\phi_t, \boldsymbol{\theta})^T\boldsymbol{S}^T + \boldsymbol{S}\text{diag}(\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{\theta}))\boldsymbol{S}^T$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j} = \begin{cases} \boldsymbol{V}_{t_i} & \text{if } i = j \\ cov(\boldsymbol{x}_{t_i}, \boldsymbol{x}_{t_j}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j-1}\boldsymbol{\Phi}(t_{j-1}, t_j)^T & \end{cases}$$

- Fisher Information

$$FI(\boldsymbol{\theta})_{m,n} = \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})^T}{\partial \theta_m}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_n} + \frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_m}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_n}\right)$$

Likelihood for the Linear Noise Approximation

$$p(\boldsymbol{x}^{(TS)}|\boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

- $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\phi_{t1}, \ldots, \phi_{tn})^T$ a nN vector with solutions of the MRE.
- $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ a $nN \times nN$ block matrix with blocks $\boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j}$ $N \times N$
- $$\boldsymbol{x}_t \sim \mathcal{N}(\phi_t, \boldsymbol{V}_t), \ \frac{d\boldsymbol{V}_t}{dt} = \boldsymbol{S}\boldsymbol{J}_{\boldsymbol{f}}(\phi_t, \boldsymbol{\theta})\boldsymbol{V}_t + \boldsymbol{V}_t\boldsymbol{J}_{\boldsymbol{f}}(\phi_t, \boldsymbol{\theta})^T\boldsymbol{S}^T + \boldsymbol{S}\text{diag}(\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{\theta}))\boldsymbol{S}^T$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j} = \begin{cases} \boldsymbol{V}_{t_i} & \text{if } i = j \\ cov(\boldsymbol{x}_{t_i}, \boldsymbol{x}_{t_j}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{i,j-1}\boldsymbol{\Phi}(t_{j-1}, t_j)^T & \end{cases}$$

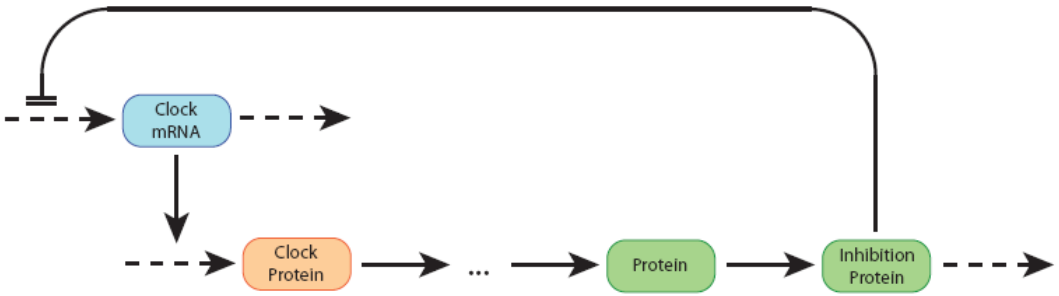
- Fisher Information

$$FI(\boldsymbol{\theta})_{m,n} = \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})^T}{\partial \theta_m}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_n} + \frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_m}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_n}\right)$$

- Augment the MRE for ϕ with the lower triangular elements of $\boldsymbol{V}$ and solve the augmented system with forward sensitivity analysis.

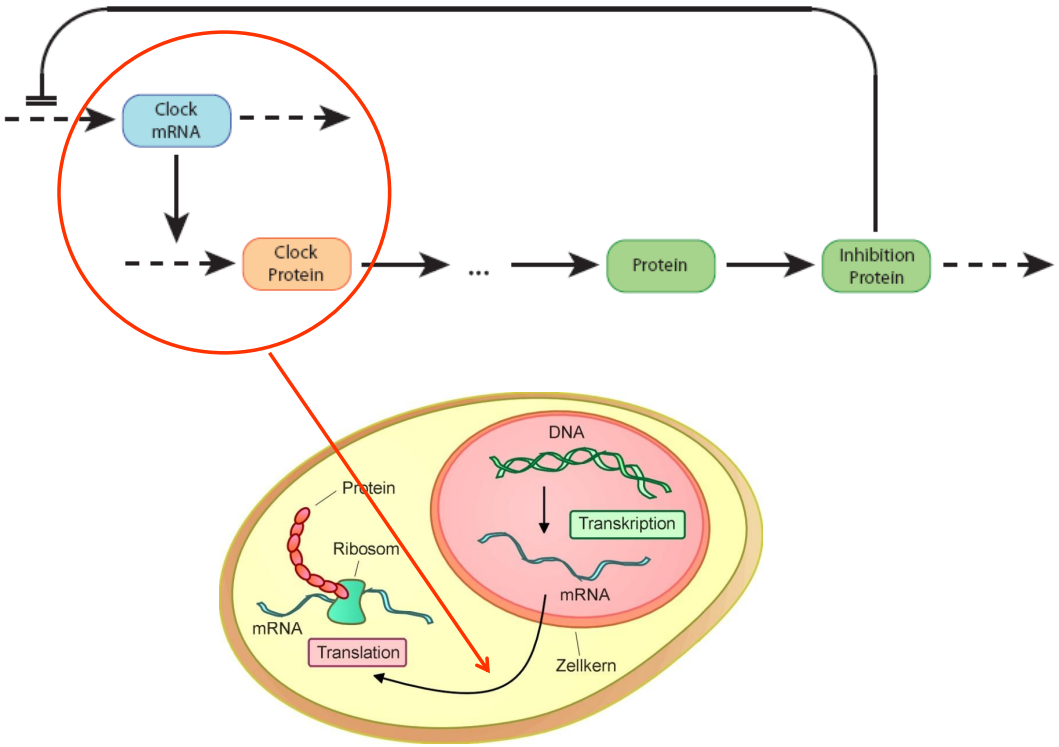
Mechanistic Models

Mechanistic modelling



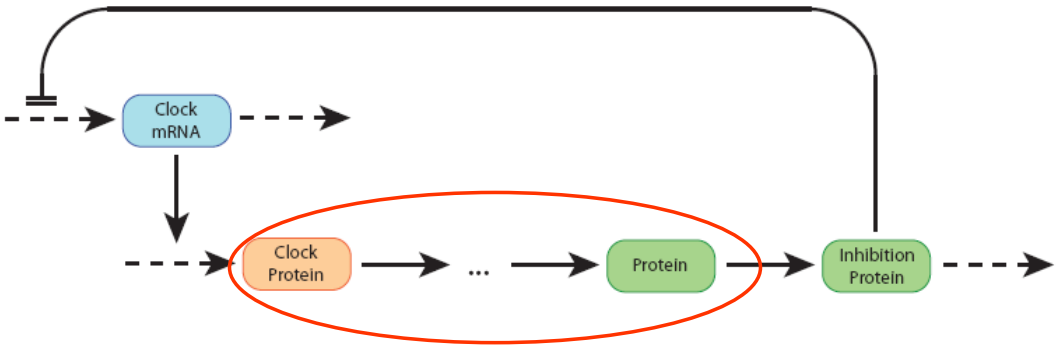
Mechanistic Models

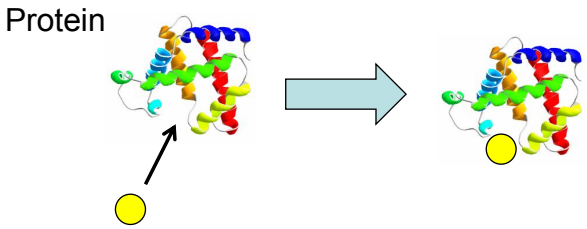
Mechanistic modelling



Mechanistic Models

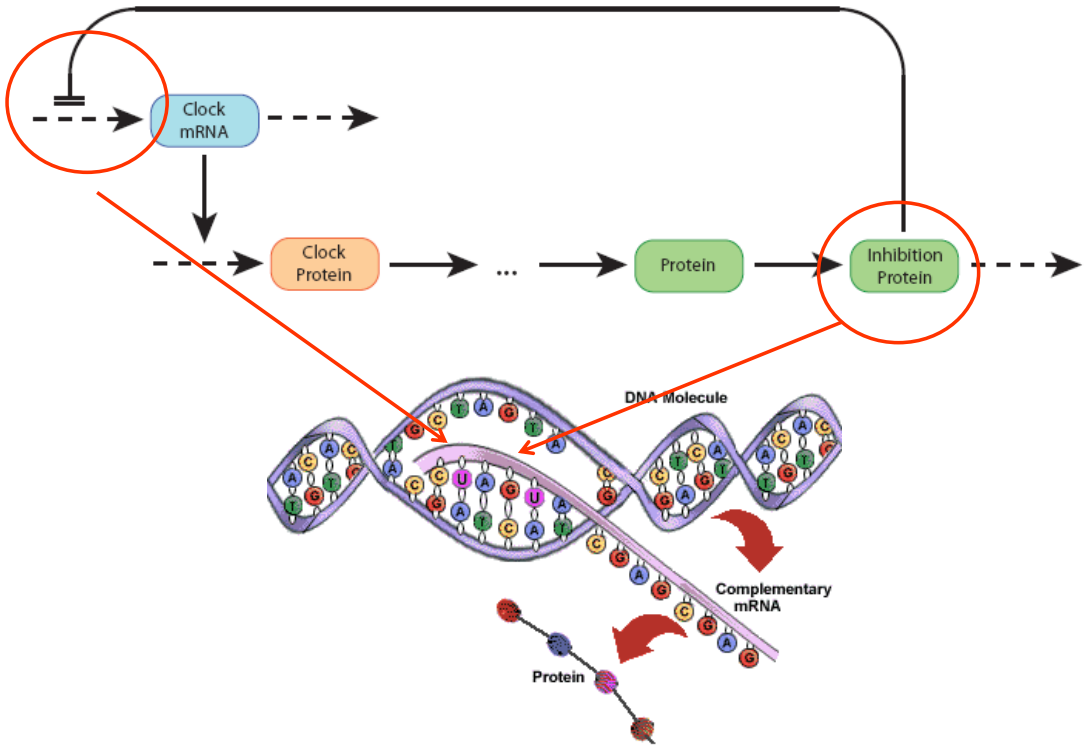
Mechanistic modelling





Mechanistic Models

Mechanistic modelling



Mechanistic Models

Mechanistic modelling

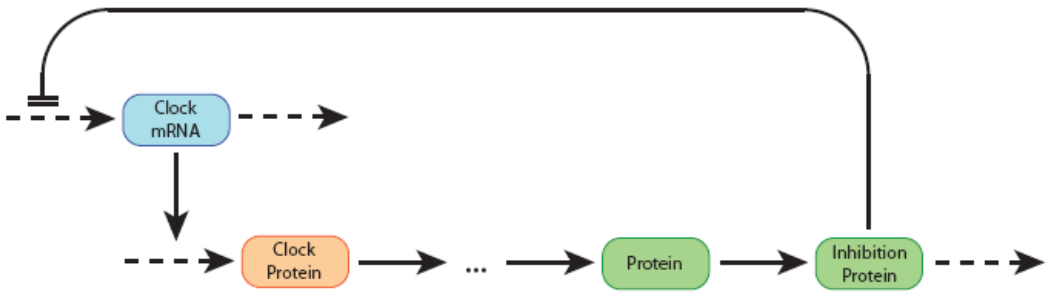


Figure: A diagram of the Goodwin Circadian Oscillator model network.

$$
\begin{aligned}
\frac{dx_1}{dt} &= \frac{k_1}{1 + x_n^{\rho}} - m_1 x_1 \\
\frac{dx_2}{dt} &= k_2 x_1 - m_2 x_2 \\
&\vdots \\
\frac{dx_n}{dt} &= k_n x_{n-1} - m_n x_n
\end{aligned}
$$

Mechanistic Models

Mechanistic modelling

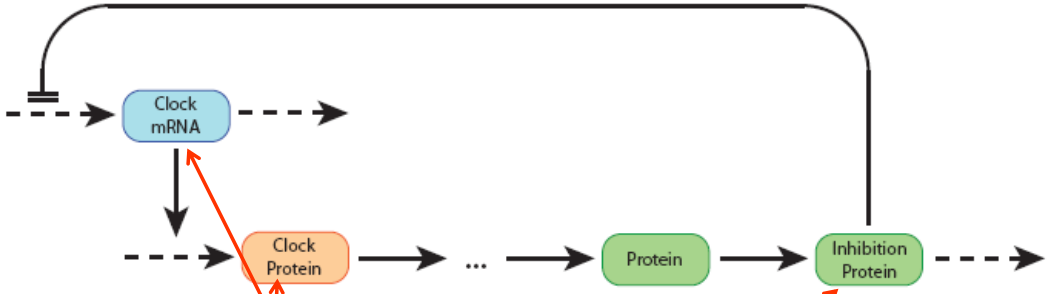


Figure: A diagram of the Goodwin Circadian Oscillator model network.

$$\frac{dx_1}{dt} = \frac{k_1}{1 + x_n^p} - m_1 x_1$$

$$\frac{dx_2}{dt} = k_2 x_1 - m_2 x_2$$

$$\vdots$$

$$\frac{dx_n}{dt} = k_n x_{n-1} - m_n x_n$$

Mechanistic Models

Mechanistic modelling

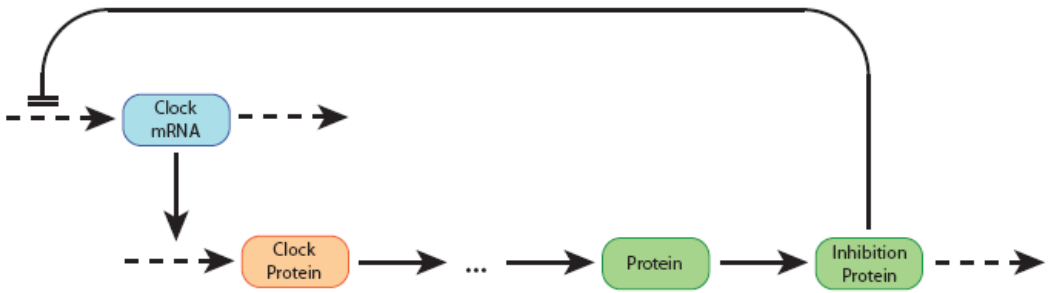


Figure: A diagram of the Goodwin Circadian Oscillator model network.

$$
\begin{aligned}
\frac{dx_1}{dt} &= \frac{k_1}{1 + x_n^{\rho}} - m_1 x_1 \\
\frac{dx_2}{dt} &= k_2 x_1 - m_2 x_2 \\
&\vdots \\
\frac{dx_n}{dt} &= k_n x_{n-1} - m_n x_n
\end{aligned}
$$

Mechanistic Models

Mechanistic modelling

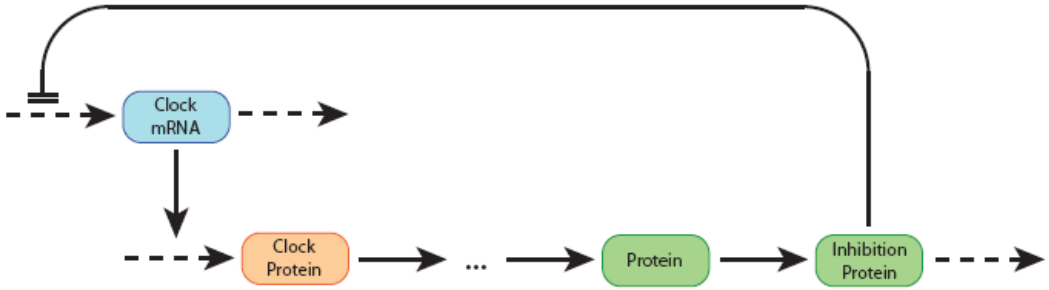


Figure: A diagram of the Goodwin Circadian Oscillator model network.

Dependence on kinetic parameters → $\boldsymbol{\theta}$

$$
\begin{aligned}
\frac{dx_1}{dt} &= \frac{k_1}{1 + x_n^{\rho}} - m_1 x_1 \\
\frac{dx_2}{dt} &= k_2 x_1 - m_2 x_2 \\
&\vdots \\
\frac{dx_n}{dt} &= k_n x_{n-1} - m_n x_n
\end{aligned}
$$

Mechanistic Models

Necessary condition for stable oscillations

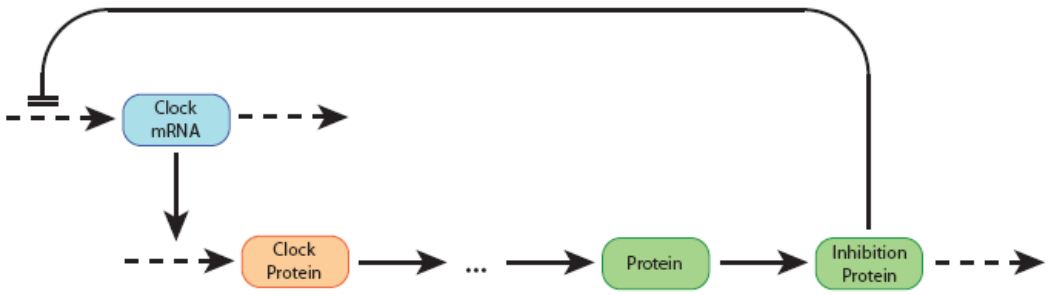


Figure: A diagram of the Goodwin Circadian Oscillator model network.

$$
\begin{aligned}
\frac{dx_1}{dt} &= \frac{k_1}{1 + x_n^{\rho}} - m_1 x_1 \\
\frac{dx_2}{dt} &= k_2 x_1 - m_2 x_2 \\
&\vdots \\
\frac{dx_n}{dt} &= k_n x_{n-1} - m_n x_n
\end{aligned}
$$

> 2

> 8

Mechanistic Models

Example: log likelihood landscape

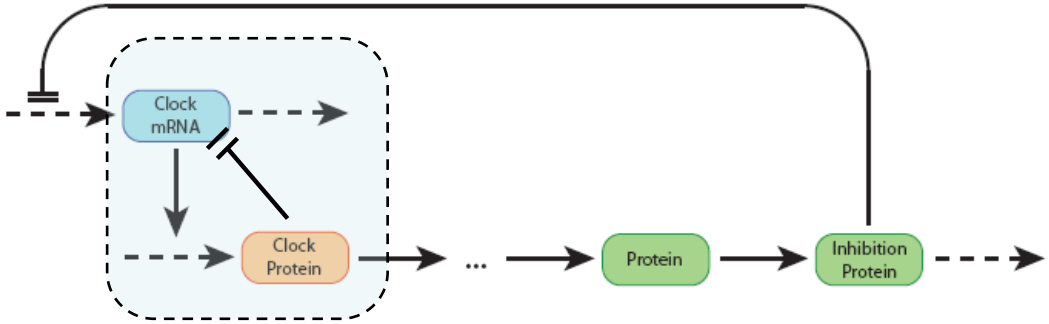


Figure: A diagram of the Goodwin Circadian Oscillator model network.

$$
\begin{aligned}
\frac{dx_1}{dt} &= \frac{k_1}{1 + x_n^{\rho}} - m_1 x_1 \\
\frac{dx_2}{dt} &= k_2 x_1 - m_2 x_2 \\
&\vdots \\
\frac{dx_n}{dt} &= k_n x_{n-1} - m_n x_n
\end{aligned}
$$

Mechanistic Models

Example: log likelihood landscape

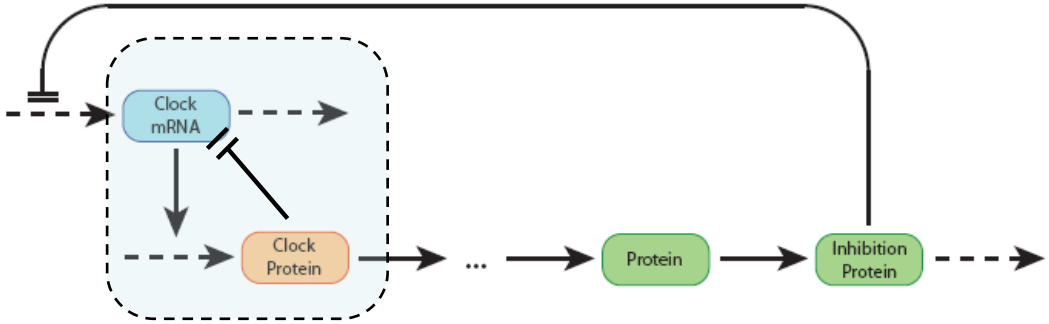


Figure: A diagram of the Goodwin Circadian Oscillator model network.

$$\frac{dx_1}{dt} = \frac{k_1}{1 + x_n^{\rho}} - m_1 x_1$$

$$\frac{dx_2}{dt} = k_2 x_1 - m_2 x_2$$

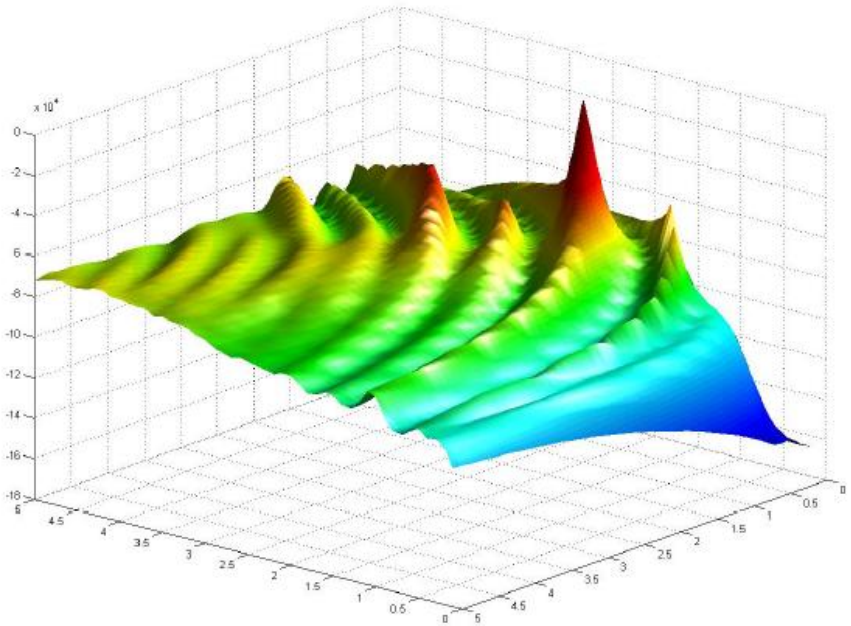
$$\vdots$$

$$\frac{dx_n}{dt} = k_n x_{n-1} - m_n x_n$$

2 parameters free,
others fixed

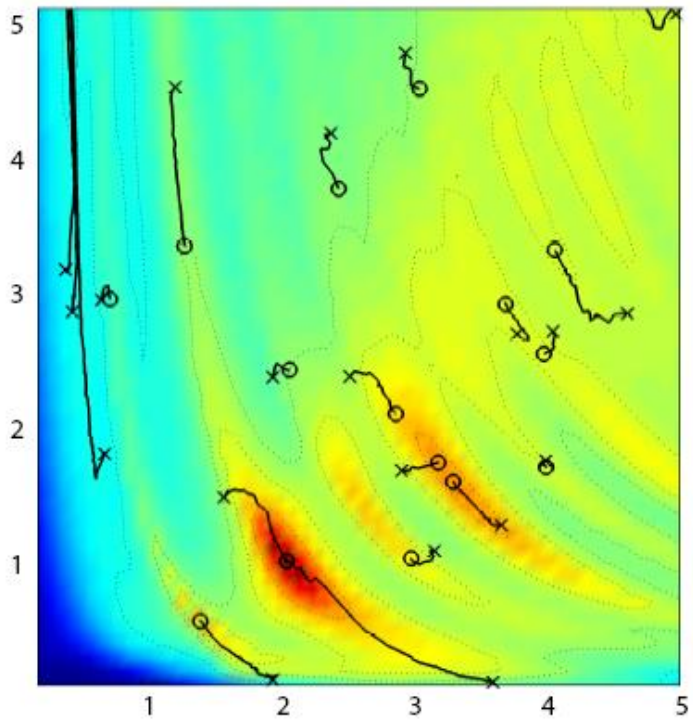
Mechanistic Models

Posterior landscape $P(\boldsymbol{\theta}|\mathcal{M}, \mathcal{D})$



Mechanistic Models

MCMC trajectories



Journal of the
Royal Statistical Society
SERIES B
Statistical
Methodology

J. R. Statist. Soc. B (2011)
73, *Part* 2, *pp.* 123–214

Riemann manifold Langevin and Hamiltonian Monte Carlo methods

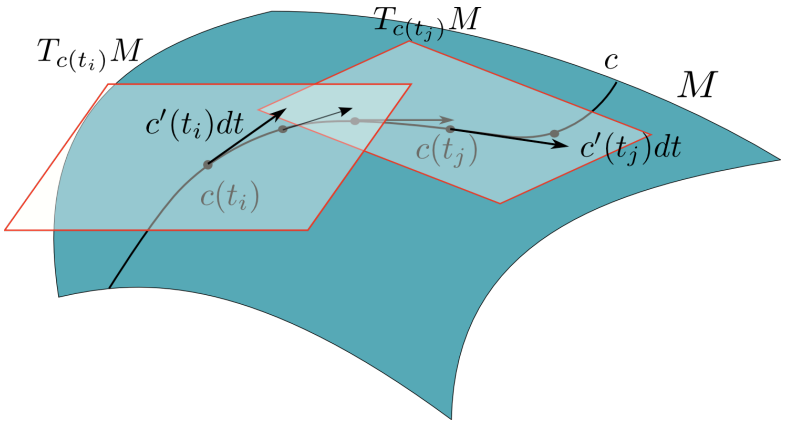
Mark Girolami and Ben Calderhead

University College London, UK

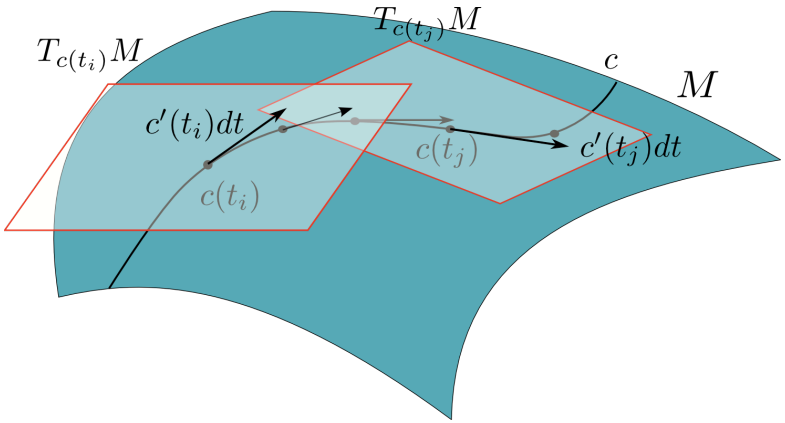
[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, October 13th, 2010*, Professor D. M. Titterington *in the Chair*]

Summary. The paper proposes Metropolis adjusted Langevin and Hamiltonian Monte Carlo sampling methods defined on the Riemann manifold to resolve the shortcomings of existing Monte Carlo algorithms when sampling from target densities that may be high dimensional and exhibit strong correlations. The methods provide fully automated adaptation mechanisms that circumvent the costly pilot runs that are required to tune proposal densities for Metropolis–Hastings or indeed Hamiltonian Monte Carlo and Metropolis adjusted Langevin algorithms. This allows for highly efficient sampling even in very high dimensions where different scalings may be required for the transient and stationary phases of the Markov chain. The methodology proposed exploits the Riemann geometry of the parameter space of statistical models and thus automat-

Geometric Concepts in MCMC

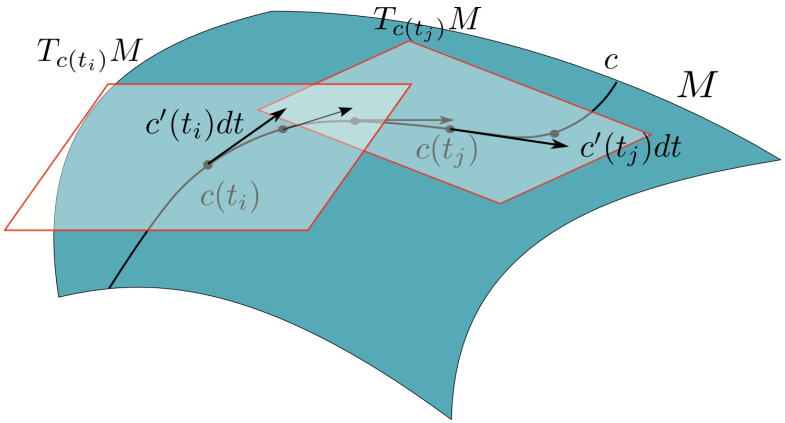


Geometric Concepts in MCMC



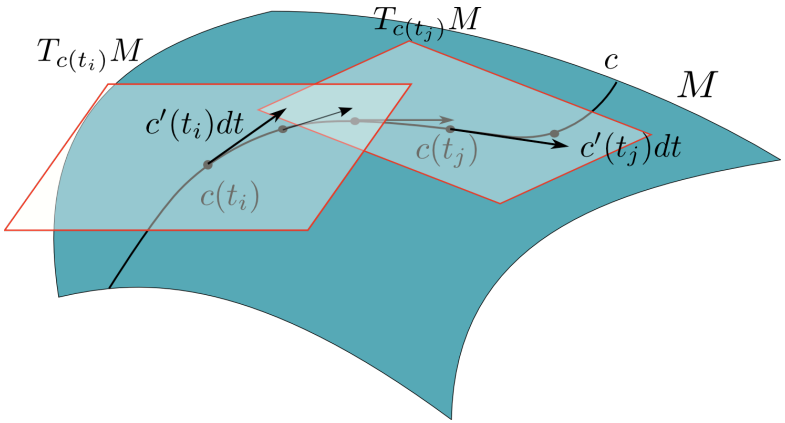
- Tangent space - local metric defined by $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta} = g_{kl}\delta\theta_k\delta\theta_l$

Geometric Concepts in MCMC



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

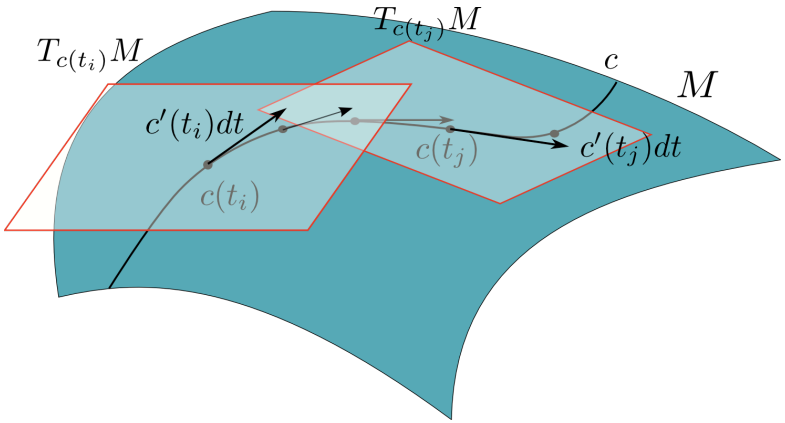
Geometric Concepts in MCMC



- Tangent space - local metric defined by $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta} = g_{kl}\delta\theta_k\delta\theta_l$
- 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

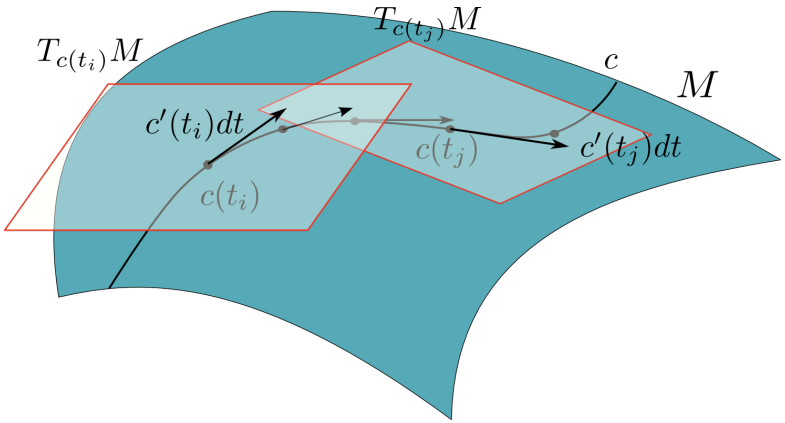


- Tangent space - local metric defined by $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta} = g_{kl}\delta\theta_k\delta\theta_l$
- Christoffel symbols - characterise Levi-Civita connection on manifold

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

- Geodesics - shortest path between two points on manifold

Geometric Concepts in MCMC



- Tangent space - local metric defined by $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta} = g_{kl}\delta\theta_k\delta\theta_l$
- 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. $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\mu, \sigma)^{\mathsf{T}}$

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. $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\mu, \sigma)^{\mathsf{T}}$
- Metric is Expected Fisher Information

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

Illustration of Geometric Concepts

- Consider Normal density $p(x|\mu, \sigma) = \mathcal{N}_x(\mu, \sigma)$
- Local inner product on tangent space defined by metric tensor, i.e. $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\mu, \sigma)^{\mathsf{T}}$
- Metric is Expected Fisher Information

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

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

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. $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\mu, \sigma)^{\mathsf{T}}$
- Metric is Expected Fisher Information

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

- Components of connection $\partial_\mu \mathbf{G} = \mathbf{0}$ and $\partial_\sigma \mathbf{G} = -\operatorname{diag}(2\sigma^{-3}, 4\sigma^{-3})$
- Metric on tangent space

$$\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta} = \frac{(\delta\mu^2 + 2\delta\sigma^2)}{\sigma^2}$$

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. $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\mu, \sigma)^{\mathsf{T}}$
- Metric is Expected Fisher Information

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

- Components of connection $\partial_\mu \mathbf{G} = \mathbf{0}$ and $\partial_\sigma \mathbf{G} = -\operatorname{diag}(2\sigma^{-3}, 4\sigma^{-3})$
- Metric on tangent space

$$\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta} = \frac{(\delta\mu^2 + 2\delta\sigma^2)}{\sigma^2}$$

- 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. $\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\mu, \sigma)^{\mathsf{T}}$
- Metric is Expected Fisher Information

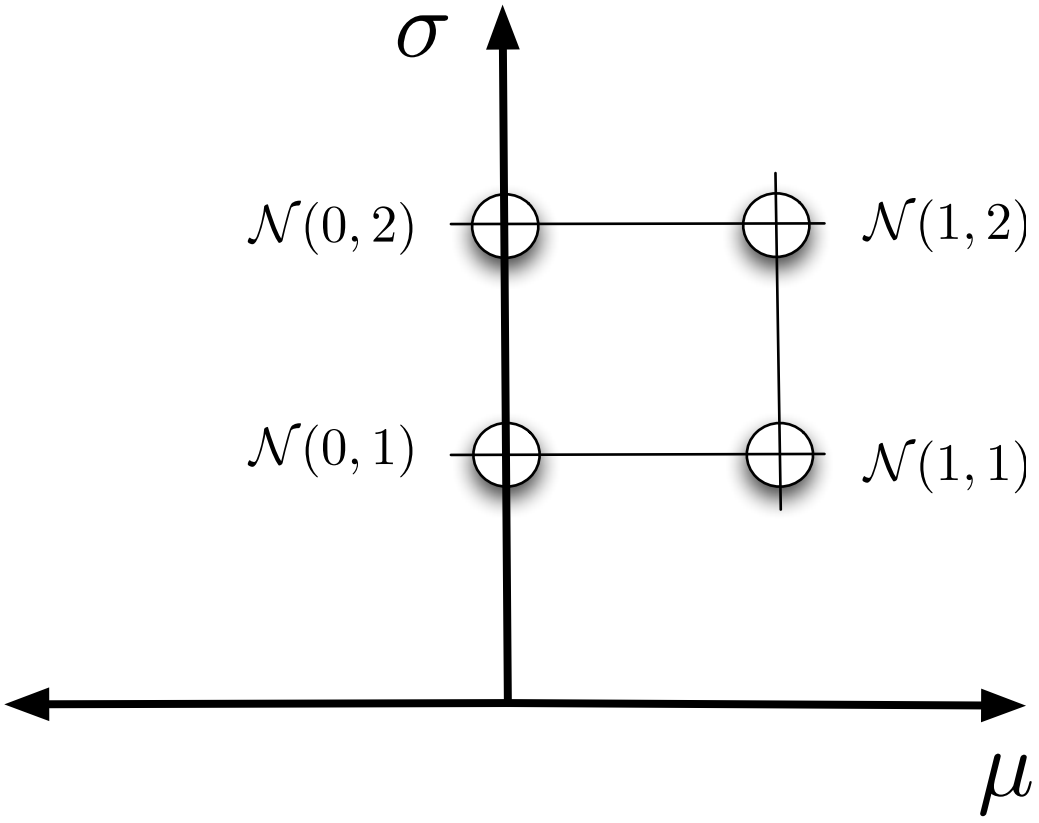
$$\mathbf{G}(\mu, \sigma) = \begin{bmatrix} \sigma^{-2} & 0 \\ 0 & 2\sigma^{-2} \end{bmatrix}$$

- Components of connection $\partial_\mu \mathbf{G} = \mathbf{0}$ and $\partial_\sigma \mathbf{G} = -\operatorname{diag}(2\sigma^{-3}, 4\sigma^{-3})$
- Metric on tangent space

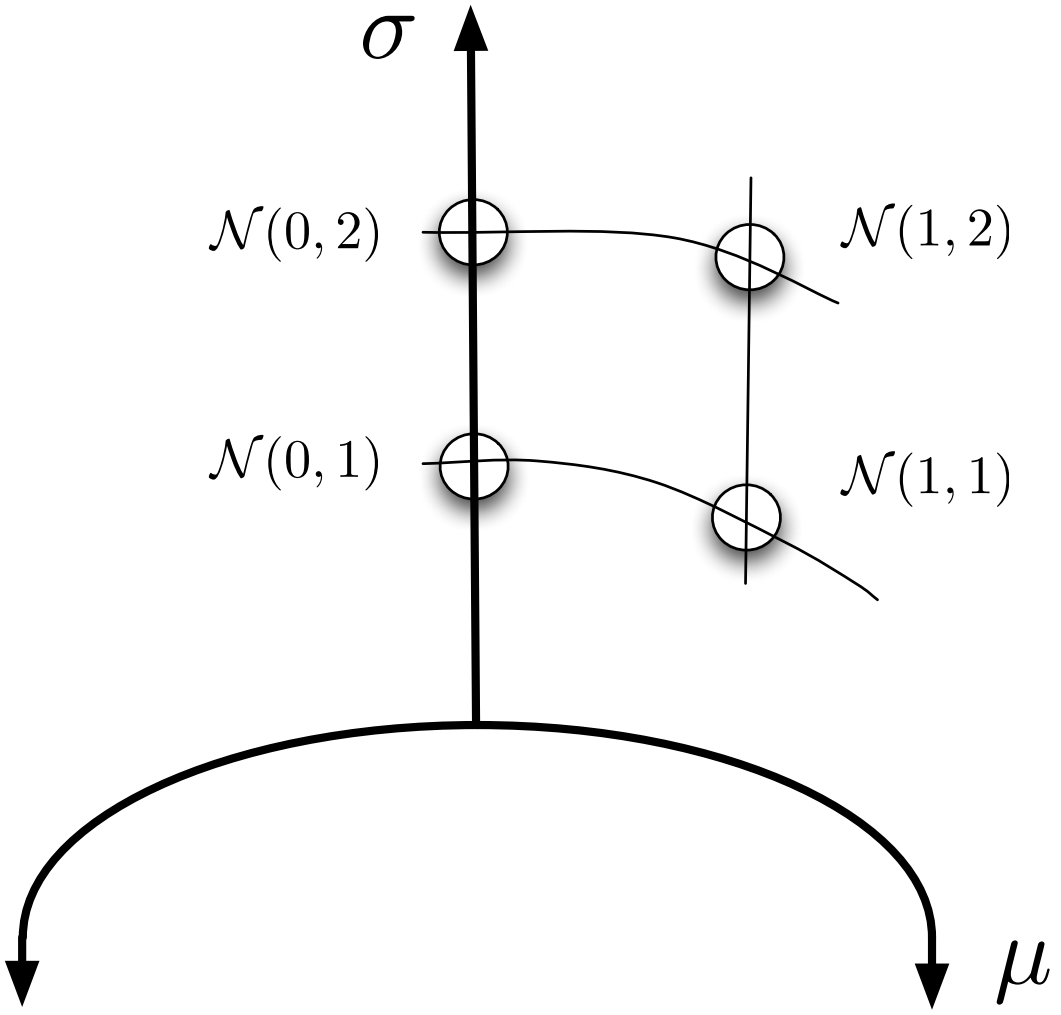
$$\delta\boldsymbol{\theta}^{\mathsf{T}}\mathbf{G}(\boldsymbol{\theta})\delta\boldsymbol{\theta} = \frac{(\delta\mu^2 + 2\delta\sigma^2)}{\sigma^2}$$

- Metric tensor for univariate Normal defines a Hyperbolic Space
- Consider densities $\mathcal{N}(0, 1)$ & $\mathcal{N}(1, 1)$ and $\mathcal{N}(0, 2)$ & $\mathcal{N}(1, 2)$

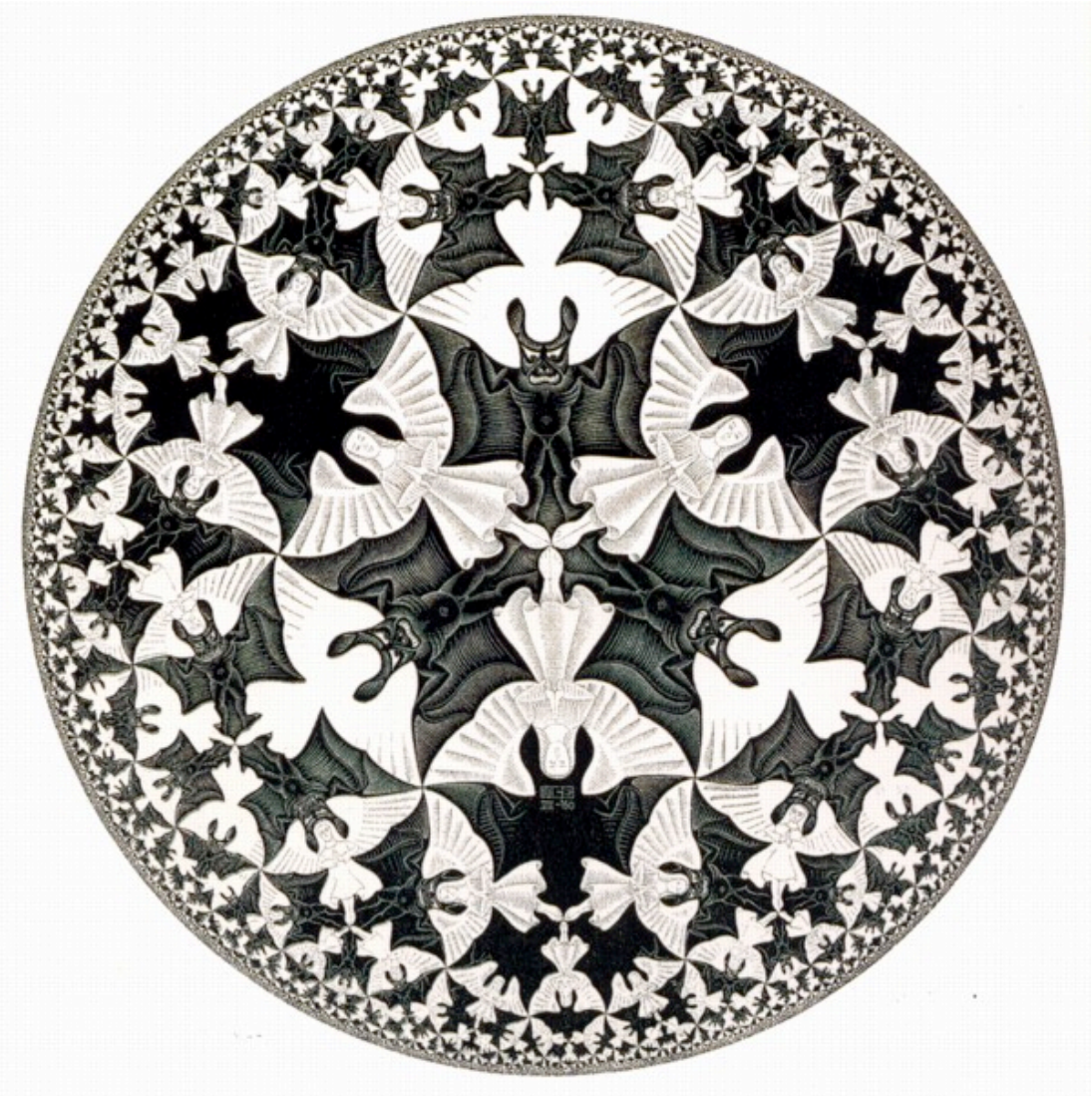
Normal Density - Euclidean Parameter space



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

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

Langevin Diffusion on Riemannian manifold

- Discretised Langevin diffusion on manifold defines proposal mechanism

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

- Manifold with constant curvature then proposal mechanism reduces to

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

Langevin Diffusion on Riemannian manifold

- Discretised Langevin diffusion on manifold defines proposal mechanism

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

- Manifold with constant curvature then proposal mechanism reduces to

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

- MALA proposal with preconditioning

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

Langevin Diffusion on Riemannian manifold

- Discretised Langevin diffusion on manifold defines proposal mechanism

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

- Manifold with constant curvature then proposal mechanism reduces to

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

- MALA proposal with preconditioning

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

- Proposal and acceptance probability

$$p_p(\theta'|\theta) = \mathcal{N}(\theta'|\boldsymbol{\mu}(\theta,\epsilon), \epsilon^2\mathbf{G}^{-1}(\theta))$$

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

Langevin Diffusion on Riemannian manifold

- Discretised Langevin diffusion on manifold defines proposal mechanism

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

- Manifold with constant curvature then proposal mechanism reduces to

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

- MALA proposal with preconditioning

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

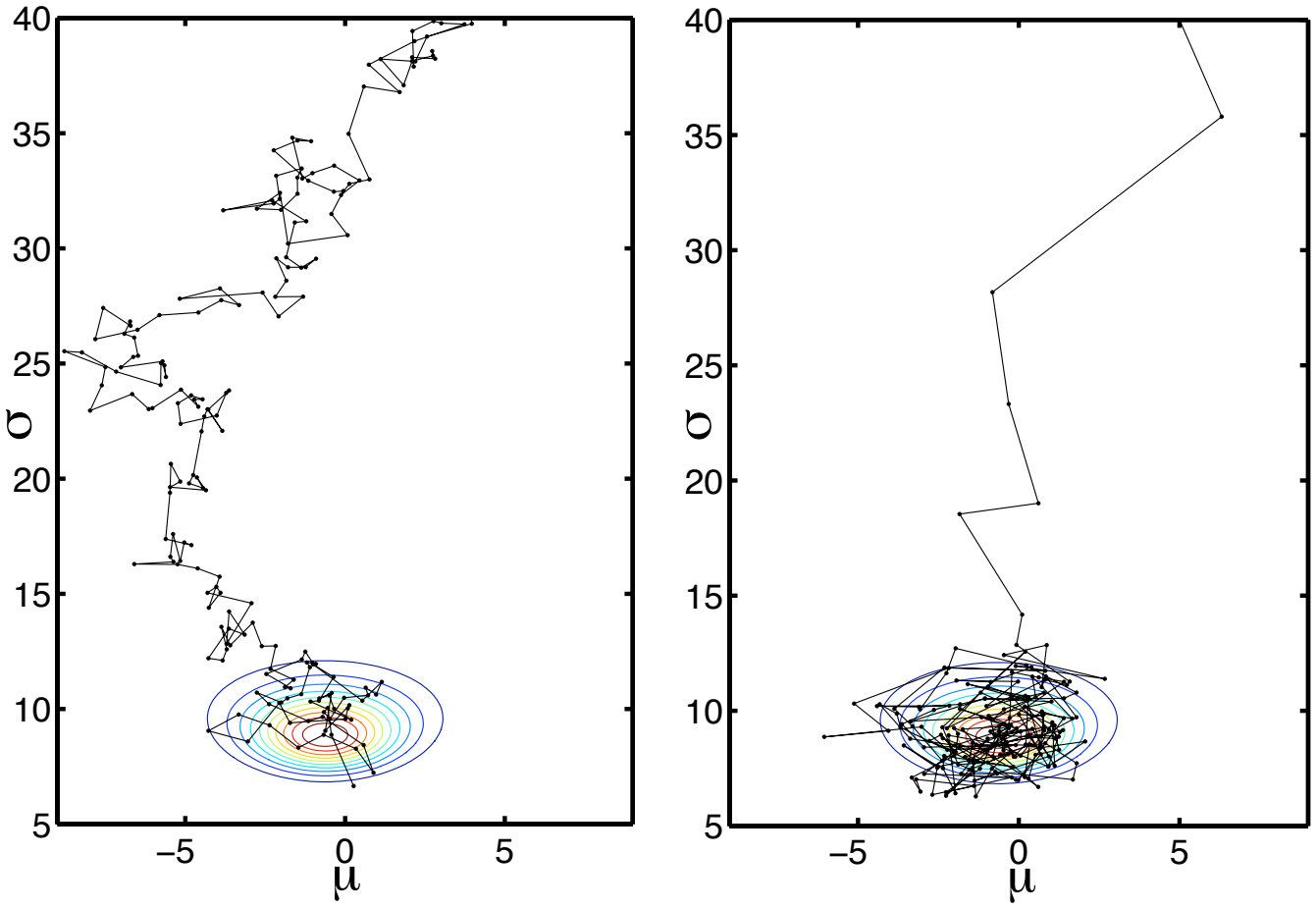
- Proposal and acceptance probability

$$p_p(\theta'|\theta) = \mathcal{N}(\theta'|\boldsymbol{\mu}(\theta,\epsilon), \epsilon^2\mathbf{G}^{-1}(\theta))$$

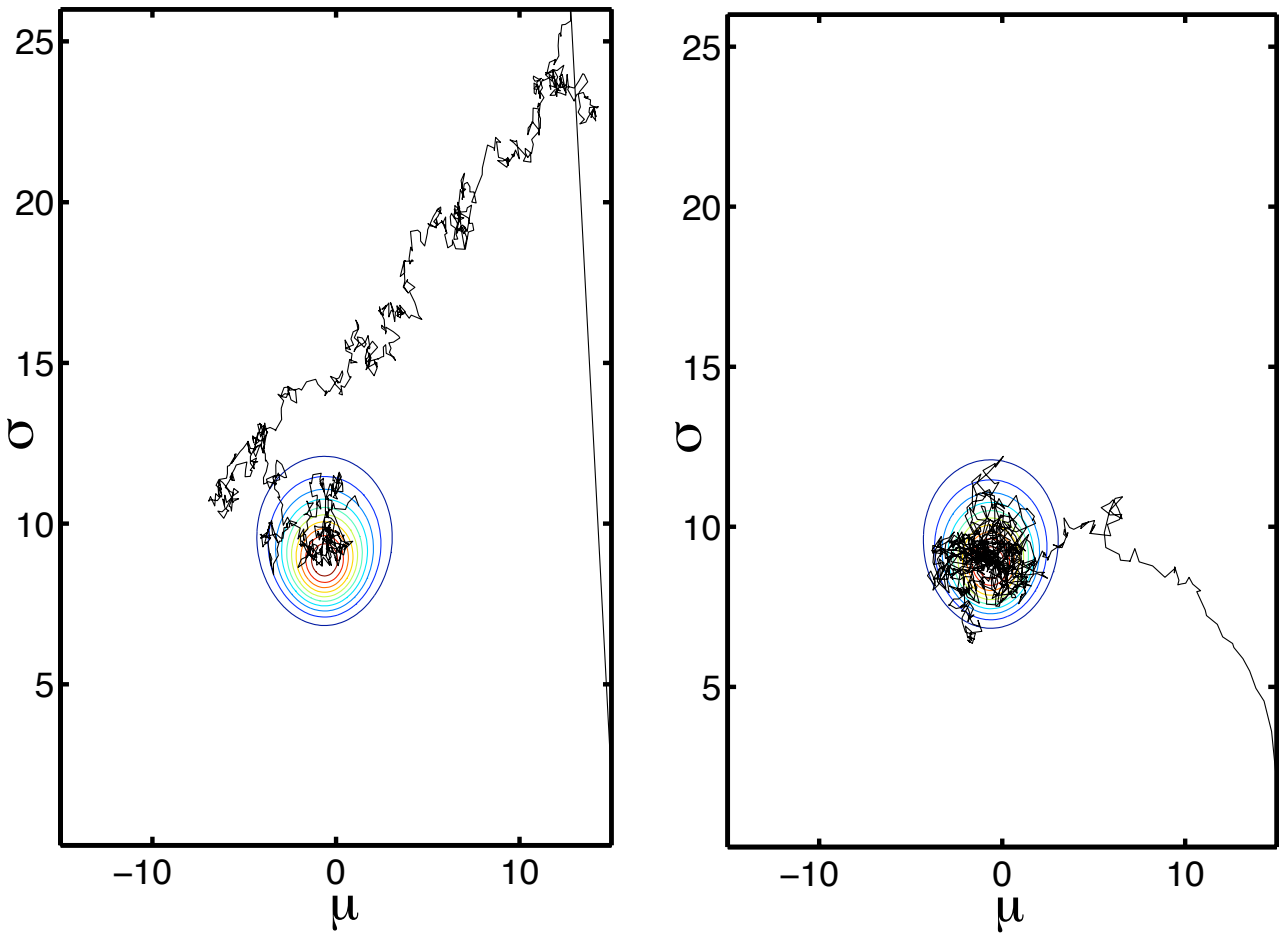
$$p_a(\theta'|\theta) = \min\left[1, \frac{\pi(\theta')p_p(\theta|\theta')}{\pi(\theta)p_p(\theta'|\theta)}\right]$$

- Proposal mechanism diffuses approximately along the manifold

Langevin Diffusion on Riemannian manifold



Langevin Diffusion on Riemannian manifold



Simplified Manifold MALA

- For $\boldsymbol{\theta} \in \mathbb{R}^D$ with density $\pi(\boldsymbol{\theta})$, $\mathcal{L}(\boldsymbol{\theta}) \equiv \log \pi(\boldsymbol{\theta})$
- Define proposal distribution

$$q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{t-1}) = \mathcal{N}\left(\boldsymbol{\theta}^* | \boldsymbol{\mu}(\boldsymbol{\theta}^{t-1}, \epsilon), \epsilon^2 \mathbf{G}^{-1}(\boldsymbol{\theta}^{t-1})\right)$$

Simplified Manifold MALA

- For $\boldsymbol{\theta} \in \mathbb{R}^D$ with density $\pi(\boldsymbol{\theta})$, $\mathcal{L}(\boldsymbol{\theta}) \equiv \log \pi(\boldsymbol{\theta})$
- Define proposal distribution

$$q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1}) = \mathcal{N}\left(\boldsymbol{\theta}^*|\boldsymbol{\mu}(\boldsymbol{\theta}^{t-1}, \epsilon), \epsilon^2 \mathbf{G}^{-1}(\boldsymbol{\theta}^{t-1})\right)$$

Where

$$\boldsymbol{\mu}(\boldsymbol{\theta}, \epsilon) = \boldsymbol{\theta} + \frac{\epsilon^2}{2}\left(\mathbf{G}^{-1}(\boldsymbol{\theta})\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})\right), \quad \mathbf{G}^{-1}(\boldsymbol{\theta}) = FI(\boldsymbol{\theta})$$

Simplified Manifold MALA

- For $\boldsymbol{\theta} \in \mathbb{R}^D$ with density $\pi(\boldsymbol{\theta})$, $\mathcal{L}(\boldsymbol{\theta}) \equiv \log \pi(\boldsymbol{\theta})$
- Define proposal distribution

$$q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1}) = \mathcal{N}\left(\boldsymbol{\theta}^*|\boldsymbol{\mu}(\boldsymbol{\theta}^{t-1}, \epsilon), \epsilon^2 \mathbf{G}^{-1}(\boldsymbol{\theta}^{t-1})\right)$$

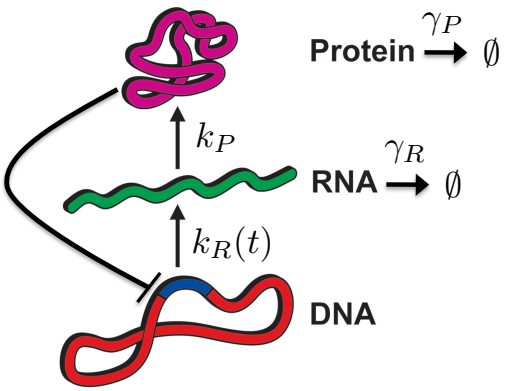
Where

$$\boldsymbol{\mu}(\boldsymbol{\theta}, \epsilon) = \boldsymbol{\theta} + \frac{\epsilon^2}{2}\left(\mathbf{G}^{-1}(\boldsymbol{\theta})\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})\right), \quad \mathbf{G}^{-1}(\boldsymbol{\theta}) = FI(\boldsymbol{\theta})$$

- Accept $\boldsymbol{\theta}^*$ with probability

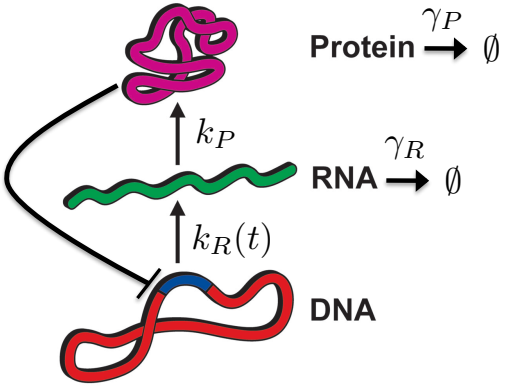
$$\min\left\{1, \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{t-1})}\frac{q(\boldsymbol{\theta}^{t-1}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1})}\right\}$$

Single gene expression model



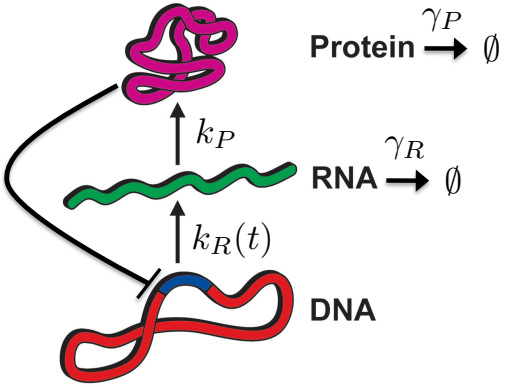
- Single gene expression model with negative feedback.

Single gene expression model



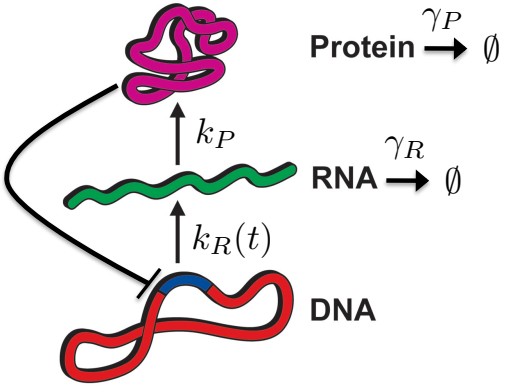
- Single gene expression model with negative feedback.
- System state $(R(t), P(t))^T$ models the population of RNA and protein.

Single gene expression model



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

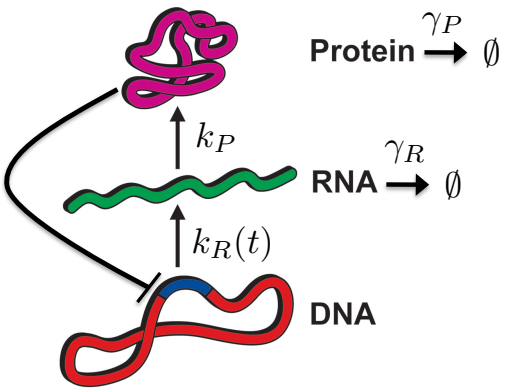
Single gene expression model



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

$$\boldsymbol{S} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Single gene expression model



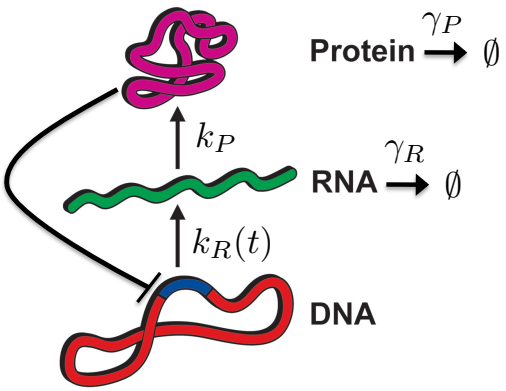
- Single gene expression model with negative feedback.
- System state $(R(t), P(t))^T$ models the population of RNA and protein.
- $k_R(P, t) = (b_0 \exp(-b_1(t - b_2)^2) + b_3)/(1 + (P/H)^{n_H})$
- $H = b_3 k_P/(2\gamma_R\gamma_P), \quad n_H = 1/2$

$$\boldsymbol{S} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

- Deterministic MRE $\phi = (\phi_R, \phi_P)^T$

$$\begin{aligned} d\phi_R/dt &= k_R(\phi_P, t) - \gamma_R\phi_R \\ d\phi_P/dt &= k_P\phi_R - \gamma_P\phi_P \end{aligned}$$

Single gene expression model



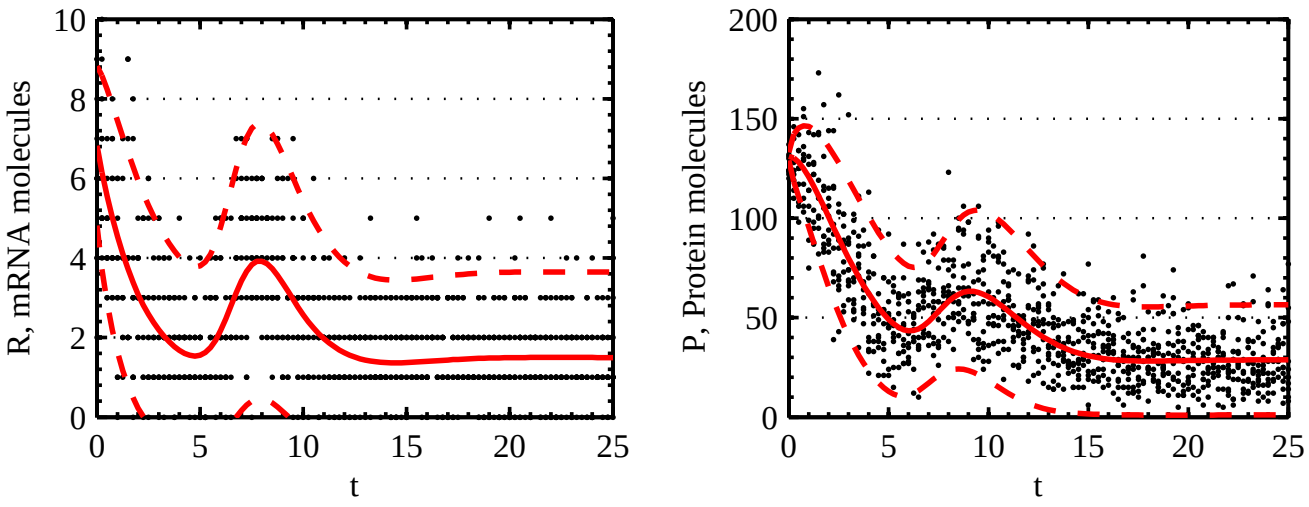
- Single gene expression model with negative feedback.
- System state $(R(t), P(t))^T$ models the population of RNA and protein.
- $k_R(P, t) = (b_0 \exp(-b_1(t - b_2)^2) + b_3)/(1 + (P/H)^{n_H})$
- $H = b_3 k_P/(2\gamma_R\gamma_P), \quad n_H = 1/2$

$$\boldsymbol{S} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

- Deterministic MRE $\phi = (\phi_R, \phi_P)^T$

$$\begin{aligned} d\phi_R/dt &= k_R(\phi_P, t) - \gamma_R\phi_R \\ d\phi_P/dt &= k_P\phi_R - \gamma_P\phi_P \end{aligned}$$

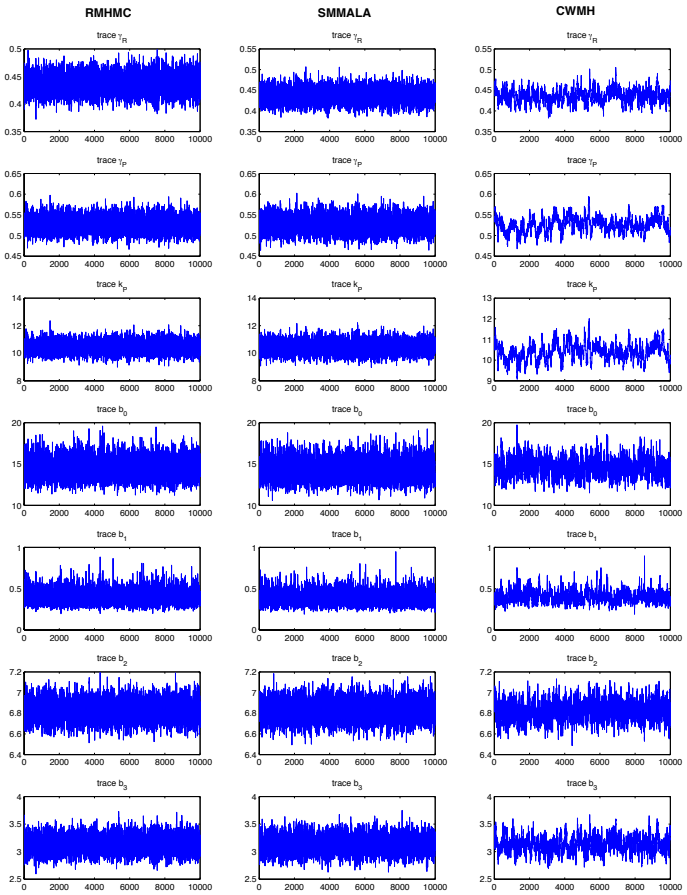
Simulated Data



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

γ_R	γ_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

Trace Plots


RMHMC
SMMALA
CWMH
trace γ_R
trace γ_P
trace k_P
trace b_0
trace b_1
trace b_2
trace b_3

Effective Sample Size

10,000 posterior samples

	γ_R	γ_P	k_P	b_0	b_1	b_2	b_3
RMHMC	6532	6593	6614	5112	5384	6595	6642
SMMALA	2990	3270	3454	3124	3164	3316	3195
CWMH	201	71	73	465	339	420	239

Effects of System Size

- Unstable monomer, S_1, can dimerise to an unstable dimer, S_2, then converted to a stable form, S_3.

Effects of System Size

- Unstable monomer, S_1, can dimerise to an unstable dimer, S_2, then converted to a stable form, S_3.
- The reaction set for this system is

$$
\begin{aligned}
R1 : S_1 &\xrightarrow{c_1} \emptyset \\
R2 : 2S_1 &\xrightarrow{c_2\Omega^{-1}} S_2 \\
R3 : S_2 &\xrightarrow{c_3} 2S_1 \\
R4 : S_2 &\xrightarrow{c_4} S_3
\end{aligned}
$$

Effects of System Size

- Unstable monomer, S_1, can dimerise to an unstable dimer, S_2, then converted to a stable form, S_3.
- The reaction set for this system is

$$
\begin{aligned}
R1 : S_1 &\xrightarrow{c_1} \emptyset \\
R2 : 2S_1 &\xrightarrow{c_2\Omega^{-1}} S_2 \\
R3 : S_2 &\xrightarrow{c_3} 2S_1 \\
R4 : S_2 &\xrightarrow{c_4} S_3
\end{aligned}
$$

- The propensity functions $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) = [c_1 S_1(t), c_2\Omega^{-1} S_1(t)(S_1(t) - 1)/2, c_3 S_2(t), c_4 S_3(t)]^T$

Effects of System Size

- Unstable monomer, S_1, can dimerise to an unstable dimer, S_2, then converted to a stable form, S_3.
- The reaction set for this system is

$$
\begin{aligned}
R1 : S_1 &\xrightarrow{c_1} \emptyset \\
R2 : 2S_1 &\xrightarrow{c_2\Omega^{-1}} S_2 \\
R3 : S_2 &\xrightarrow{c_3} 2S_1 \\
R4 : S_2 &\xrightarrow{c_4} S_3
\end{aligned}
$$

- The propensity functions $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) = [c_1 S_1(t), c_2\Omega^{-1} S_1(t)(S_1(t) - 1)/2, c_3 S_2(t), c_4 S_3(t)]^T$
- Corresponding state change matrix is

$$
\boldsymbol{S} = \begin{pmatrix} -1 & -2 & 2 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

Effects of System Size

- Unstable monomer, S_1, can dimerise to an unstable dimer, S_2, then converted to a stable form, S_3.
- The reaction set for this system is

$$
\begin{aligned}
R1 : S_1 &\xrightarrow{c_1} \emptyset \\
R2 : 2S_1 &\xrightarrow{c_2\Omega^{-1}} S_2 \\
R3 : S_2 &\xrightarrow{c_3} 2S_1 \\
R4 : S_2 &\xrightarrow{c_4} S_3
\end{aligned}
$$

- The propensity functions $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) = [c_1 S_1(t), c_2\Omega^{-1} S_1(t)(S_1(t) - 1)/2, c_3 S_2(t), c_4 S_3(t)]^T$
- Corresponding state change matrix is

$$
\boldsymbol{S} = \begin{pmatrix} -1 & -2 & 2 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

- Assume initial conditions known $S_1(t_0) = 5\Omega$, $S_2(t_0) = S_3(t_0) = 0$, $t_0 = 0$. Reaction rate parameters to $c_1 = 1$, $\hat{c}_2 = 2\Omega^{-1}$, $c_3 = 0.5$ and $c_4 = 0.04$

Effects of System Size

min. ESS vs. Ω

Ω	M.H.	SMMALA	RMHMC
1	121 (3.6)	150 (3.9)	245 (0.06)
2	226 (6.7)	2163 (57.2)	4775 (1.3)
5	132 (3.9)	3539 (93.6)	4618 (1.2)
10	180 (5.3)	3397 (89.8)	5954 (1.6)
100	214 (6.4)	3725 (98.5)	6066 (1.7)

Posterior mean and SD. vs. Ω

Ω	c_1	$\hat{c}_2$	c_3	c_4
1	0.88 (0.031)	1.72 (0.253)	0.39 (0.039)	0.003 (0.002)
2	1.3 (0.041)	0.69 (0.066)	0.35 (0.016)	0.014 (0.002)
5	0.93 (0.019)	0.39 (0.028)	0.48 (0.025)	0.034 (0.002)
10	1.0 (0.015)	0.18 (0.008)	0.47 (0.015)	0.037 (0.001)
100	0.99 (0.004)	0.01 (0.0002)	0.52 (0.004)	0.039 (0.0003)

Failure Modes

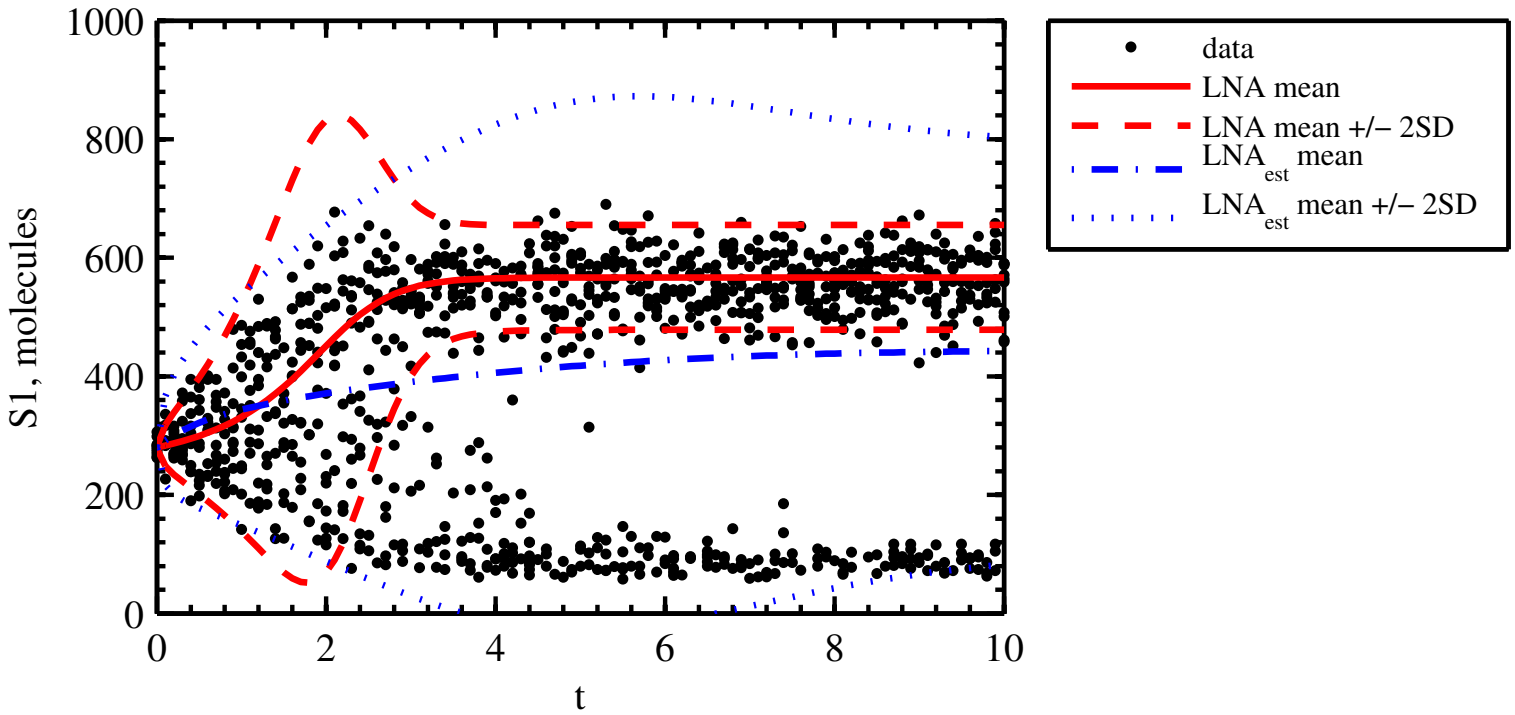


Figure: Simulated time point data using SSA for the Schlögl reaction set and LNA predictions. Dots correspond to simulated data. The bold and dashed red lines correspond to the LNA prediction for the means and standard deviations using the true parameters. Doted blue lines correspond the LNA predictions using the posterior means for the rate parameters. (Online version in colour.)

Conclusions

- MJP common tool to describe many phenomena in physical and life sciences.

Conclusions

- MJP common tool to describe many phenomena in physical and life sciences.
- LNA provides a useful approximation in appropriate operational regimes.

Conclusions

- MJP common tool to describe many phenomena in physical and life sciences.
- LNA provides a useful approximation in appropriate operational regimes.
- Decouples deterministic and stochastic characteristics of model.

Conclusions

- MJP common tool to describe many phenomena in physical and life sciences.
- LNA provides a useful approximation in appropriate operational regimes.
- Decouples deterministic and stochastic characteristics of model.
- Statistical inference remains a formidable challenge over such models.

Conclusions

- MJP common tool to describe many phenomena in physical and life sciences.
- LNA provides a useful approximation in appropriate operational regimes.
- Decouples deterministic and stochastic characteristics of model.
- Statistical inference remains a formidable challenge over such models.
- Exploitation of schoolboy differential geometry in MCMC provides effective inference tool.

Conclusions

- MJP common tool to describe many phenomena in physical and life sciences.
- LNA provides a useful approximation in appropriate operational regimes.
- Decouples deterministic and stochastic characteristics of model.
- Statistical inference remains a formidable challenge over such models.
- Exploitation of schoolboy differential geometry in MCMC provides effective inference tool.
- Phil.Trans paper describes a number of larger scale scenarios.

Conclusions

- MJP common tool to describe many phenomena in physical and life sciences.
- LNA provides a useful approximation in appropriate operational regimes.
- Decouples deterministic and stochastic characteristics of model.
- Statistical inference remains a formidable challenge over such models.
- Exploitation of schoolboy differential geometry in MCMC provides effective inference tool.
- Phil.Trans paper describes a number of larger scale scenarios.
- Ongoing work with Sherlock, Golightly.