Diffusion limit for Random Walk Metropolis algorithm out of stationarity

M. Ottobre and A. M. Stuart

Department of Mathematics
Imperial College London

June 3, 2014

Outline of talk

- Random Walk Metropolis
- Diffusion limit

	product form	non-product form
stationary		
non-stationary		

Outline of talk

- Random Walk Metropolis
- Diffusion limit

	product form	non-product form
stationary		
non-stationary		

Outline of talk

- Random Walk Metropolis
- Diffusion limit

	product form	non-product form
stationary		
non-stationary		

Metropolis - Hastings algorithms

- Generate a chain $\{x_k\}_{k=1}^{\infty}$ satisfying detailed balance condition with respect to target measure π
- At step k,
  1. Propose move
  $$y_{k+1} \sim q(x_k, \cdot)$$
  2. Calculate *acceptance probability*
  $$\alpha_k := \alpha(x_k, y_{k+1}) = \min\left\{1, \frac{\pi(y_{k+1})q(y_{k+1}, x_k)}{\pi(x_k)q(x_k, y_{k+1})}\right\}$$
  3. Update position
  $$x_{k+1} = \begin{cases} y_{k+1} & \text{with probability } \alpha_k \\ x_k & \text{with probability } 1-\alpha_k \end{cases}$$

Metropolis - Hastings algorithms

- Generate a chain $\{x_k\}_{k=1}^{\infty}$ satisfying detailed balance condition with respect to target measure π
- At step k,
  1. Propose move
  $$y_{k+1} \sim q(x_k, \cdot)$$
  2. Calculate *acceptance probability*
  $$\alpha_k := \alpha(x_k, y_{k+1}) = \min\left\{1, \frac{\pi(y_{k+1})q(y_{k+1}, x_k)}{\pi(x_k)q(x_k, y_{k+1})}\right\}$$
  3. Update position
  $$x_{k+1} = \begin{cases} y_{k+1} & \text{with probability } \alpha_k \\ x_k & \text{with probability } 1-\alpha_k \end{cases}$$

Metropolis - Hastings algorithms

- Generate a chain $\{x_k\}_{k=1}^{\infty}$ satisfying detailed balance condition with respect to target measure π
- At step k,
  1. Propose move
  $$y_{k+1} \sim q(x_k, \cdot)$$
  2. Calculate *acceptance probability*
  $$\alpha_k := \alpha(x_k, y_{k+1}) = \min\left\{1, \frac{\pi(y_{k+1})q(y_{k+1}, x_k)}{\pi(x_k)q(x_k, y_{k+1})}\right\}$$
  3. Update position
  $$x_{k+1} = \begin{cases} y_{k+1} & \text{with probability } \alpha_k \\ x_k & \text{with probability } 1-\alpha_k \end{cases}$$

Metropolis - Hastings algorithms

- Generate a chain $\{x_k\}_{k=1}^{\infty}$ satisfying detailed balance condition with respect to target measure π
- At step k,
  1. Propose move
  $$y_{k+1} \sim q(x_k, \cdot)$$
  2. Calculate *acceptance probability*
  $$\alpha_k := \alpha(x_k, y_{k+1}) = \min\left\{1, \frac{\pi(y_{k+1})q(y_{k+1}, x_k)}{\pi(x_k)q(x_k, y_{k+1})}\right\}$$
  3. Update position
  $$x_{k+1} = \begin{cases} y_{k+1} & \text{with probability } \alpha_k \\ x_k & \text{with probability } 1-\alpha_k \end{cases}$$

Random Walk Metropolis

- RWM proposal:

$$y = x + \sigma\,\xi, \qquad \xi \sim \mathcal{N}(0, Id_N).$$

- Being the proposal $q(x, y)$ symmetric, the acceptance probability

$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\}$$

reduces to

$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}.$$

- Notation

x^N	vector on $\mathbb{R}^N$
$x^{i,N}$	i-th component
$x_k^{i,N}$	k-th step

- RWM proposal in $\mathbb{R}^N$:

$$y^N = x^N + \sqrt{\sigma_N}\,\xi^N, \qquad \xi^N \sim \mathcal{N}(0, Id_N).$$

- Problem: efficiency of the algorithm depends crucially on the scaling of the proposal variance.

- Notation

x^N	vector on $\mathbb{R}^N$
$x^{i,N}$	i-th component
$x_k^{i,N}$	k-th step

- RWM proposal in $\mathbb{R}^N$:

$$y^N = x^N + \sqrt{\sigma_N}\,\xi^N, \qquad \xi^N \sim \mathcal{N}(0, Id_N).$$

- Problem: efficiency of the algorithm depends crucially on the scaling of the proposal variance.

- Notation

$$x^N \qquad \text{vector on } \mathbb{R}^N$$
$$x^{i,N} \qquad i\text{-th component}$$
$$x_k^{i,N} \qquad k\text{-th step}$$

- RWM proposal in $\mathbb{R}^N$:

$$y^N = x^N + \sqrt{\sigma_N}\,\xi^N, \qquad \xi^N \sim \mathcal{N}(0, Id_N).$$

- Problem: efficiency of the algorithm depends crucially on the scaling of the proposal variance.

G. O. Roberts, A. Gelman, and W. R. Gilks. *Weak convergence and optimal scaling of random walk Metropolis algorithms.* Ann. Appl. Probab., 1997.

$$\sigma_N = \frac{2\ell^2}{N}$$

1. If target measure is in product form $\pi^N(x^N) = \prod_{i=1}^N e^{-V(x^{i,N})}$
2. Algorithm started in stationarity: $x_0^N \sim \pi^N$

- Consider the interpolant

$$z^{(N)}(t) = (Nt - k)x_{k+1}^N + (k + 1 - Nt)x_k^N, \quad \frac{k}{N} \le t < \frac{k+1}{N}.$$

- Then each component of $z^{(N)}(t)$ converges weakly to

$$dZ_t = -h_\ell V'(Z_t)dt + \sqrt{2h_\ell}\, dW_t$$

G. O. Roberts, A. Gelman, and W. R. Gilks. *Weak convergence and optimal scaling of random walk Metropolis algorithms.* Ann. Appl. Probab., 1997.

$$\sigma_N = \frac{2\ell^2}{N}$$

1. If target measure is in product form $\pi^N(x^N) = \prod_{i=1}^N e^{-V(x^{i,N})}$
2. Algorithm started in stationarity: $x_0^N \sim \pi^N$

- Consider the interpolant

$$z^{(N)}(t) = (Nt-k)x_{k+1}^N + (k+1-Nt)x_k^N, \quad \frac{k}{N} \le t < \frac{k+1}{N}.$$

- Then each component of $z^{(N)}(t)$ converges weakly to

$$dZ_t = -h_\ell V'(Z_t)dt + \sqrt{2h_\ell}\, dW_t$$

G. O. Roberts, A. Gelman, and W. R. Gilks. *Weak convergence and optimal scaling of random walk Metropolis algorithms.* Ann. Appl. Probab., 1997.

$$\sigma_N = \frac{2\ell^2}{N}$$

1. If target measure is in product form $\pi^N(x^N) = \prod_{i=1}^N e^{-V(x^{i,N})}$
2. Algorithm started in stationarity: $x_0^N \sim \pi^N$

- Consider the interpolant

$$z^{(N)}(t) = (Nt - k)x_{k+1}^N + (k + 1 - Nt)x_k^N, \quad \frac{k}{N} \leq t < \frac{k+1}{N}.$$

- Then each component of $z^{(N)}(t)$ converges weakly to

$$dZ_t = -h_\ell V'(Z_t)dt + \sqrt{2h_\ell}\, dW_t$$

G. O. Roberts, A. Gelman, and W. R. Gilks. *Weak convergence and optimal scaling of random walk Metropolis algorithms.* Ann. Appl. Probab., 1997.

$$\sigma_N = \frac{2\ell^2}{N}$$

1. If target measure is in product form $\pi^N(x^N) = \prod_{i=1}^N e^{-V(x^{i,N})}$
2. Algorithm started in stationarity: $x_0^N \sim \pi^N$

- Consider the interpolant

$$z^{(N)}(t) = (Nt - k)x_{k+1}^N + (k + 1 - Nt)x_k^N, \quad \frac{k}{N} \le t < \frac{k+1}{N}.$$

- Then each component of $z^{(N)}(t)$ converges weakly to

$$dZ_t = -h_\ell V'(Z_t)dt + \sqrt{2h_\ell}\, dW_t$$

In non-product form however still stationary

1. A. Beskos, G. Roberts, A. Stuart. *Optimal scalings of Metropolis-Hastings algorithms for non-product targets in high dimensions.* Ann. Appl. Prob, 2009.
2. L. Breyer, M. Piccioni, and S. Scarlatti. Optimal scaling of MALA for nonlinear regression. Ann. Appl. Prob, 2004.
3. L. Breyer and G. Roberts. From Metropolis to diffusions: Gibbs states and optimal scaling. SPA, 2000.
4. J. Mattingly, N. Pillai, A. Stuart. *Diffusion Limits of the Random Walk Metropolis Algorithm in High Dimensions.* Ann. Appl. Prob, 2011.

⋮

- Out of stationarity for target i.i.d. Gaussians:

  O. F. Christensen, G. O. Roberts and J. S. Rosenthal. Scaling limits for the transient phase of local Metropolis-Hastings algorithms. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2005.

- Out of stationarity with target in product form:

  B. Jourdain, T. Lelievre, and B. Miasojedow. *Optimal scaling for the transient phase of the random walk metropolis algorithm: the mean field limit.* Bernoulli, 2013.

Jourdain et al.

1. If target measure is in product form $\pi^N(x^N) = \prod_{i=1}^N \exp\{-V(x^{i,N})\}$
2. Algorithm started out of stationarity

- Then each component of the chain converges to

$$dZ_t = -D_\ell(t)V'(Z_t)dt + \sqrt{G_\ell(t)}d\beta_t$$

where

$$D_\ell(t) := \mathcal{D}(\,\mathbb{E}(V'(Z_t))^2, \mathbb{E}(V''(Z_t))\,)$$
$$G_\ell(t) := \mathcal{G}(\,\mathbb{E}(V'(Z_t))^2, \mathbb{E}(V''(Z_t))\,)$$

- Notice that in stationarity

$$D_\ell(t) = G_\ell(t) = h_\ell\,.$$

Jourdain et al.

1. If target measure is in product form $\pi^N(x^N) = \prod_{i=1}^N \exp\{-V(x^{i,N})\}$
2. Algorithm started out of stationarity

- Then each component of the chain converges to

$$dZ_t = -D_\ell(t)V'(Z_t)dt + \sqrt{G_\ell(t)}d\beta_t$$

where

$$D_\ell(t) := \mathcal{D}(\,\mathbb{E}(V'(Z_t))^2, \mathbb{E}(V''(Z_t))\,)$$
$$G_\ell(t) := \mathcal{G}(\,\mathbb{E}(V'(Z_t))^2, \mathbb{E}(V''(Z_t))\,)$$

- Notice that in stationarity

$$D_\ell(t) = G_\ell(t) = h_\ell\,.$$

Jourdain et al.

1. If target measure is in product form $\pi^N(x^N) = \prod_{i=1}^N \exp\{-V(x^{i,N})\}$
2. Algorithm started out of stationarity

- Then each component of the chain converges to

$$dZ_t = -D_\ell(t) V'(Z_t) dt + \sqrt{G_\ell(t)} d\beta_t$$

where

$$D_\ell(t) := \mathcal{D}(\, \mathbb{E}(V'(Z_t))^2, \mathbb{E}(V''(Z_t))\,)$$
$$G_\ell(t) := \mathcal{G}(\, \mathbb{E}(V'(Z_t))^2, \mathbb{E}(V''(Z_t))\,)$$

- Notice that in stationarity

$$D_\ell(t) = G_\ell(t) = h_\ell\,.$$

Example

- Conditioned diffusion

$$dx_t = -\nabla V(x_t)dt + \sqrt{2\beta^{-1}}dW_t, \quad x(t) \in \mathbb{R}^N$$

$$x(0) = x^-, \quad x(1) = x^+ .$$

- Bridge diffusion measure, $x \in L^2([0,1], \mathbb{R}^N)$

$$d\pi(x) \propto e^{-\Psi(x)} d\pi_0(x),$$

$$\Psi(x) = \int_0^1 G(x(t))dt, \quad G(x) = \frac{\beta}{4}\|\nabla V(x)\|^2 - \frac{1}{2}\Delta V(x) .$$

Example

- Conditioned diffusion

$$dx_t = -\nabla V(x_t)dt + \sqrt{2\beta^{-1}}dW_t, \quad x(t) \in \mathbb{R}^N$$

$$x(0) = x^-, \quad x(1) = x^+ .$$

- Bridge diffusion measure, $x \in L^2([0,1], \mathbb{R}^N)$

$$d\pi(x) \propto e^{-\Psi(x)} d\pi_0(x),$$

$$\Psi(x) = \int_0^1 G(x(t))dt, \quad G(x) = \frac{\beta}{4}\|\nabla V(x)\|^2 - \frac{1}{2}\Delta V(x) .$$

Target measures

- Desired target measure on $\mathcal{H}$

$$d\pi(x) \propto e^{-\Psi(x)} d\pi_0(x), \qquad \pi_0 \sim \mathcal{N}(0, \mathcal{C}), \; x \in \mathcal{H}, \\ \mathcal{C}\varphi_j = \lambda_j^2 \varphi_j \, .$$

- Projection of π on $\mathbb{R}^N = span\{\varphi_1, \dots, \varphi_N\}$

$$d\pi^N \propto e^{-\Psi^N(x^N)} d\pi_0^N, \qquad \pi_0^N \sim \mathcal{N}(0, \mathcal{C}^N), \; x^N \in \mathbb{R}^N \\ \mathcal{C}^N := diag\{\lambda_1^2, \dots, \lambda_N^2\} \, .$$

- π^N has density with respect to Lebesgue measure

$$e^{-\Psi^N(x)} e^{-\frac{1}{2}\langle (\mathcal{C}^N)^{-1} x^N, x^N \rangle} = e^{-\Psi^N(x)} e^{-\frac{1}{2}\sum_{i=1}^N \frac{|x_k^{i,N}|^2}{\lambda_i^2}}$$

Target measures

- Desired target measure on $\mathcal{H}$

$$d\pi(x) \propto e^{-\Psi(x)} d\pi_0(x), \qquad \pi_0 \sim \mathcal{N}(0, \mathcal{C}),\ x \in \mathcal{H},$$
$$\mathcal{C}\varphi_j = \lambda_j^2 \varphi_j\,.$$

- Projection of π on $\mathbb{R}^N = span\{\varphi_1, \dots, \varphi_N\}$

$$d\pi^N \propto e^{-\Psi^N(x^N)} d\pi_0^N, \qquad \pi_0^N \sim \mathcal{N}(0, \mathcal{C}^N),\ x^N \in \mathbb{R}^N$$
$$\mathcal{C}^N := diag\{\lambda_1^2, \dots, \lambda_N^2\}\,.$$

- π^N has density with respect to Lebesgue measure

$$e^{-\Psi^N(x)} e^{-\frac{1}{2}\langle (\mathcal{C}^N)^{-1} x^N, x^N \rangle} = e^{-\Psi^N(x)} e^{-\frac{1}{2}\sum_{i=1}^N \frac{|x_k^{i,N}|^2}{\lambda_i^2}}$$

Target measures

- Desired target measure on $\mathcal{H}$

$$d\pi(x) \propto e^{-\Psi(x)} d\pi_0(x), \qquad \pi_0 \sim \mathcal{N}(0, \mathcal{C}), \;\; x \in \mathcal{H},$$
$$\mathcal{C}\varphi_j = \lambda_j^2 \varphi_j \,.$$

- Projection of π on $\mathbb{R}^N = span\{\varphi_1, \dots, \varphi_N\}$

$$d\pi^N \propto e^{-\Psi^N(x^N)} d\pi_0^N, \qquad \pi_0^N \sim \mathcal{N}(0, \mathcal{C}^N), \;\; x^N \in \mathbb{R}^N$$
$$\mathcal{C}^N := diag\{\lambda_1^2, \dots, \lambda_N^2\} \,.$$

- π^N has density with respect to Lebesgue measure

$$e^{-\Psi^N(x)} e^{-\frac{1}{2}\langle (\mathcal{C}^N)^{-1} x^N, x^N \rangle} = e^{-\Psi^N(x)} e^{-\frac{1}{2}\sum_{i=1}^N \frac{|x_k^{i,N}|^2}{\lambda_i^2}}$$

Algorithm

- Proposal

$$y_{k+1}^N = x_k^N + \sqrt{\frac{2\ell^2 \mathcal{C}^N}{N}}\, \xi^N, \qquad \xi^N \sim \mathcal{N}(0, Id_N)$$

or in components

$$y_{k+1}^{i,N} = x_k^{i,N} + \sqrt{\frac{2\ell^2}{N}}\, \lambda_i\, \xi^{i,N}, \qquad \xi^{i,N} \sim \mathcal{N}(0, 1)$$

- Acceptance probability

$$\alpha_k = 1 \wedge \left\{ \frac{\pi^N(y_{k+1}^N)}{\pi^N(x_k^N)} \right\} = 1 \wedge \exp(Q_k), \quad \text{where}$$

$$Q_k := \frac{1}{2} \sum_{i=1}^{N} \frac{|x_k^{i,N}|^2}{\lambda_i^2} - \frac{1}{2} \sum_{i=1}^{N} \frac{|y_{k+1}^{i,N}|^2}{\lambda_i^2} + \Psi(x_k^N) - \Psi(y_{k+1}^N),$$

Algorithm

- Proposal

$$y_{k+1}^N = x_k^N + \sqrt{\frac{2\ell^2 \mathcal{C}^N}{N}}\, \xi^N, \qquad \xi^N \sim \mathcal{N}(0, Id_N)$$

or in components

$$y_{k+1}^{i,N} = x_k^{i,N} + \sqrt{\frac{2\ell^2}{N}}\, \lambda_i\, \xi^{i,N}, \qquad \xi^{i,N} \sim \mathcal{N}(0, 1)$$

- Acceptance probability

$$\alpha_k = 1 \wedge \left\{ \frac{\pi^N(y_{k+1}^N)}{\pi^N(x_k^N)} \right\} = 1 \wedge \exp(Q_k), \quad \text{where}$$

$$Q_k := \frac{1}{2} \sum_{i=1}^N \frac{|x_k^{i,N}|^2}{\lambda_i^2} - \frac{1}{2} \sum_{i=1}^N \frac{|y_{k+1}^{i,N}|^2}{\lambda_i^2} + \Psi(x_k^N) - \Psi(y_{k+1}^N),$$

In stationarity

1. Target $\quad d\pi^N \propto e^{-\Psi^N(x^N)} d\pi_0^N$
2. Start the chain with $x_0^N \sim \pi^N$

- Then the diffusion limit is given by

$$dz(t) = \tilde{h}_\ell[-z(t) - \mathcal{C}\nabla\Psi(z(t))]\, dt + \sqrt{2\tilde{h}_\ell}\, dW(t)$$

where $W(t)$ is a $\mathcal{C}$-Brownian motion in $\mathcal{H}$.

Out of Stationarity

1. Define the continuous interpolant of the chain x_k^N

$$z^{(N)}(t) = (Nt - k)x_{k+1}^N + (k + 1 - Nt)x_k^N, \quad \frac{k}{N} \leq t < \frac{k+1}{N}.$$

2. Assume

$$S_0 := \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \frac{\left|x_0^{j,N}\right|^2}{\lambda_j^2} < \infty.$$

- Then $\quad z_t^{(N)} \Rightarrow z_t, \quad z_t$ solution of the SDE

$$dz(t) = [-z(t) - \mathcal{C}\nabla\Psi(z(t))]d_\ell(S(t))\,dt + \sqrt{g_\ell(S(t))}\,dW(t),$$

where $S(t) \in \mathbb{R}_+$ solves the ODE

$$dS(t) = A_\ell(S(t))\,dt, \qquad S(0) = S_0\,.$$

Compare limit in stationarity and out of stationarity:

$$\text{target:} \qquad d\pi^N \propto e^{-\Psi^N(x^N)} d\pi_0^N$$

- If the chain is started out of stationarity

$$dz(t) = [-z(t) - \mathcal{C}\nabla\Psi(z(t))] d_\ell(S(t))\, dt + \sqrt{g_\ell(S(t))}\, dW(t),$$

with

$$dS(t) = A_\ell(S(t))\, dt, \quad \text{and} \quad S(t) \to 1.$$

- If $x_0^N \sim \pi^N$ then diffusion limit is

$$dz(t) = [-z(t) - \mathcal{C}\nabla\Psi(z(t))] d_\ell(1)\, dt + \sqrt{g_\ell(1)}\, dW(t)$$

Compare limit in stationarity and out of stationarity:

$$\text{target:} \qquad d\pi^N \propto e^{-\Psi^N(x^N)} d\pi_0^N$$

- If the chain is started out of stationarity

$$dz(t) = [-z(t) - \mathcal{C}\nabla\Psi(z(t))] d_\ell(S(t))\, dt + \sqrt{g_\ell(S(t))}\, dW(t),$$

with

$$dS(t) = A_\ell(S(t))\, dt, \quad \text{and} \quad S(t) \to 1.$$

- If $x_0^N \sim \pi^N$ then diffusion limit is

$$dz(t) = [-z(t) - \mathcal{C}\nabla\Psi(z(t))] d_\ell(1)\, dt + \sqrt{g_\ell(1)}\, dW(t)$$

Compare limit in stationarity and out of stationarity:

$$\text{target:} \qquad d\pi^N \propto e^{-\Psi^N(x^N)} d\pi_0^N$$

- If the chain is started out of stationarity

$$dz(t) = [-z(t) - \mathcal{C}\nabla\Psi(z(t))] d_\ell(S(t))\, dt + \sqrt{g_\ell(S(t))}\, dW(t),$$

with

$$dS(t) = A_\ell(S(t))\, dt, \quad \text{and} \quad S(t) \to 1.$$

- If $x_0^N \sim \pi^N$ then diffusion limit is

$$dz(t) = [-z(t) - \mathcal{C}\nabla\Psi(z(t))] d_\ell(1)\, dt + \sqrt{g_\ell(1)}\, dW(t)$$

- Consider the Markov chain in $\mathbb{R}$ (for every fixed N)

$$S_k^N = \frac{1}{N}\sum_{j=1}^{N}\frac{\left|x_k^{j,N}\right|^2}{\lambda_j^2}$$

- Diffusion limit for this chain: consider the continuous interpolant

$$S^{(N)}(t) = (Nt-k)S_{k+1}^N + (k+1-Nt)S_k^N, \quad \frac{k}{N} \le t < \frac{k+1}{N}.$$

- Then $\quad S^{(N)}(t) \Rightarrow S(t) \quad$ solution of the ODE

$$dS(t) = A_\ell(S(t))\,dt, \qquad S(0) = S_0\,.$$

- Consider the Markov chain in $\mathbb{R}$ (for every fixed N)

$$S_k^N = \frac{1}{N}\sum_{j=1}^{N}\frac{\left|x_k^{j,N}\right|^2}{\lambda_j^2}$$

- Diffusion limit for this chain: consider the continuous interpolant

$$S^{(N)}(t) = (Nt-k)S_{k+1}^N + (k+1-Nt)S_k^N, \quad \frac{k}{N} \le t < \frac{k+1}{N}.$$

- Then $\qquad S^{(N)}(t) \Rightarrow S(t) \qquad$ solution of the ODE

$$dS(t) = A_\ell(S(t))\,dt, \qquad S(0) = S_0\,.$$

- Consider the Markov chain in $\mathbb{R}$ (for every fixed N)

$$S_k^N = \frac{1}{N}\sum_{j=1}^{N}\frac{\left|x_k^{j,N}\right|^2}{\lambda_j^2}$$

- Diffusion limit for this chain: consider the continuous interpolant

$$S^{(N)}(t) = (Nt-k)S_{k+1}^N + (k+1-Nt)S_k^N, \quad \frac{k}{N} \le t < \frac{k+1}{N}.$$

- Then $\qquad S^{(N)}(t) \Rightarrow S(t) \qquad$ solution of the ODE

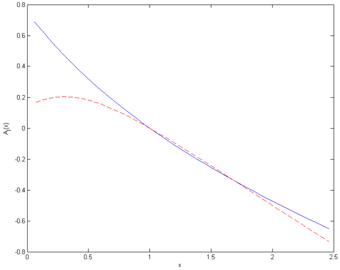
$$dS(t) = A_\ell(S(t))\,dt, \qquad S(0) = S_0\,.$$

Remarks

- In stationarity

$$S_k^N = \frac{1}{N}\sum_{j=1}^{N} \frac{\left|x_k^{j,N}\right|^2}{\lambda_j^2} \stackrel{N\to\infty}{\longrightarrow} 1 \qquad \text{for every fixed } k \geq 1.$$

- Compare $\quad S(t) \longrightarrow 1$



$$dS(t) = A_\ell(S(t))\,dt$$

S_k^N appears in the analysis for the acceptance probability

$$Q(x_k,\xi_{k+1}) = -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{i=1}^N \frac{x_k^{i,N}\xi_{k+1}^{i,N}}{\lambda_i} + \Psi(x_k^N) - \Psi(y_{k+1}^N)$$

$$= -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N, \xi_{k+1}^N\rangle + \Psi(x_k^N) - \Psi(y_k^N)$$

$$= -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N + \mathcal{C}^{1/2}\nabla\Psi^N(x_k), \xi_{k+1}^N\rangle$$

$$+ \Psi(x_k^N) - \Psi(y_k^N) + \langle \mathcal{C}^{1/2}\nabla\Psi^N, \xi_{k+1}^N\rangle$$

$$\simeq -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N, \xi_{k+1}^N\rangle$$

$$\simeq -\ell^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{i=1}^N \frac{x_k^{i,N}\xi_{k+1}^{i,N}}{\lambda_i} \sim \mathcal{N}(-\ell^2, 2\ell^2 S_k^N)\,.$$

S_k^N appears in the analysis for the acceptance probability

$$
\begin{aligned}
Q(x_k, \xi_{k+1}) &= -\frac{\ell^2}{N} \sum_{i=1}^{N} \left| \xi_{k+1}^{i,N} \right|^2 - \sqrt{\frac{2\ell^2}{N}} \sum_{i=1}^{N} \frac{x_k^{i,N} \xi_{k+1}^{i,N}}{\lambda_i} + \Psi(x_k^N) - \Psi(y_{k+1}^N) \\
&= -\frac{\ell^2}{N} \sum_{i=1}^{N} \left| \xi_{k+1}^{i,N} \right|^2 - \sqrt{\frac{2\ell^2}{N}} \langle \mathcal{C}^{-1/2} x_k^N, \xi_{k+1}^N \rangle + \Psi(x_k^N) - \Psi(y_k^N) \\
&= -\frac{\ell^2}{N} \sum_{i=1}^{N} \left| \xi_{k+1}^{i,N} \right|^2 - \sqrt{\frac{2\ell^2}{N}} \langle \mathcal{C}^{-1/2} x_k^N + \mathcal{C}^{1/2} \nabla \Psi^N(x_k), \xi_{k+1}^N \rangle \\
&\quad + \Psi(x_k^N) - \Psi(y_k^N) + \langle \mathcal{C}^{1/2} \nabla \Psi^N, \xi_{k+1}^N \rangle \\
&\simeq -\frac{\ell^2}{N} \sum_{i=1}^{N} \left| \xi_{k+1}^{i,N} \right|^2 - \sqrt{\frac{2\ell^2}{N}} \langle \mathcal{C}^{-1/2} x_k^N, \xi_{k+1}^N \rangle \\
&\simeq -\ell^2 - \sqrt{\frac{2\ell^2}{N}} \sum_{i=1}^{N} \frac{x_k^{i,N} \xi_{k+1}^{i,N}}{\lambda_i} \sim \mathcal{N}(-\ell^2, 2\ell^2 S_k^N) \, .
\end{aligned}
$$

S_k^N appears in the analysis for the acceptance probability

$$Q(x_k,\xi_{k+1}) = -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{i=1}^N \frac{x_k^{i,N}\xi_{k+1}^{i,N}}{\lambda_i} + \Psi(x_k^N) - \Psi(y_{k+1}^N)$$

$$= -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N, \xi_{k+1}^N\rangle + \Psi(x_k^N) - \Psi(y_k^N)$$

$$= -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N + \mathcal{C}^{1/2}\nabla\Psi^N(x_k), \xi_{k+1}^N\rangle$$

$$+ \Psi(x_k^N) - \Psi(y_k^N) + \langle \mathcal{C}^{1/2}\nabla\Psi^N, \xi_{k+1}^N\rangle$$

$$\simeq -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N, \xi_{k+1}^N\rangle$$

$$\simeq -\ell^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{i=1}^N \frac{x_k^{i,N}\xi_{k+1}^{i,N}}{\lambda_i} \sim \mathcal{N}(-\ell^2, 2\ell^2 S_k^N)\,.$$

S_k^N appears in the analysis for the acceptance probability

$$
\begin{aligned}
Q(x_k, \xi_{k+1}) &= -\frac{\ell^2}{N} \sum_{i=1}^{N} \left| \xi_{k+1}^{i,N} \right|^2 - \sqrt{\frac{2\ell^2}{N}} \sum_{i=1}^{N} \frac{x_k^{i,N} \xi_{k+1}^{i,N}}{\lambda_i} + \Psi(x_k^N) - \Psi(y_{k+1}^N) \\
&= -\frac{\ell^2}{N} \sum_{i=1}^{N} \left| \xi_{k+1}^{i,N} \right|^2 - \sqrt{\frac{2\ell^2}{N}} \langle \mathcal{C}^{-1/2} x_k^N, \xi_{k+1}^N \rangle + \Psi(x_k^N) - \Psi(y_k^N) \\
&= -\frac{\ell^2}{N} \sum_{i=1}^{N} \left| \xi_{k+1}^{i,N} \right|^2 - \sqrt{\frac{2\ell^2}{N}} \langle \mathcal{C}^{-1/2} x_k^N + \mathcal{C}^{1/2} \nabla \Psi^N(x_k), \xi_{k+1}^N \rangle \\
&+ \Psi(x_k^N) - \Psi(y_k^N) + \langle \mathcal{C}^{1/2} \nabla \Psi^N, \xi_{k+1}^N \rangle \\
&\simeq -\frac{\ell^2}{N} \sum_{i=1}^{N} \left| \xi_{k+1}^{i,N} \right|^2 - \sqrt{\frac{2\ell^2}{N}} \langle \mathcal{C}^{-1/2} x_k^N, \xi_{k+1}^N \rangle \\
&\simeq -\ell^2 - \sqrt{\frac{2\ell^2}{N}} \sum_{i=1}^{N} \frac{x_k^{i,N} \xi_{k+1}^{i,N}}{\lambda_i} \sim \mathcal{N}(-\ell^2, 2\ell^2 S_k^N) .
\end{aligned}
$$

S_k^N appears in the analysis for the acceptance probability

$$
\begin{aligned}
Q(x_k, \xi_{k+1}) &= -\frac{\ell^2}{N}\sum_{i=1}^{N}\left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{i=1}^{N}\frac{x_k^{i,N}\xi_{k+1}^{i,N}}{\lambda_i} + \Psi(x_k^N) - \Psi(y_{k+1}^N) \\
&= -\frac{\ell^2}{N}\sum_{i=1}^{N}\left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N, \xi_{k+1}^N\rangle + \Psi(x_k^N) - \Psi(y_k^N) \\
&= -\frac{\ell^2}{N}\sum_{i=1}^{N}\left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N + \mathcal{C}^{1/2}\nabla\Psi^N(x_k), \xi_{k+1}^N\rangle \\
&+ \Psi(x_k^N) - \Psi(y_k^N) + \langle \mathcal{C}^{1/2}\nabla\Psi^N, \xi_{k+1}^N\rangle \\
&\simeq -\frac{\ell^2}{N}\sum_{i=1}^{N}\left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2}x_k^N, \xi_{k+1}^N\rangle \\
&\simeq -\ell^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{i=1}^{N}\frac{x_k^{i,N}\xi_{k+1}^{i,N}}{\lambda_i} \sim \mathcal{N}(-\ell^2, 2\ell^2 S_k^N)\,.
\end{aligned}
$$

References

1. B. Jourdain, T. Lelievre, and B. Miasojedow. *Optimal scaling for the transient phase of the random walk metropolis algorithm:the mean field limit.* Bernoulli, 2013.

3. J.C. Mattingly, N.S. Pillai, and A.M. Stuart. *Diffusion Limits of the Random Walk Metropolis Algorithm in High Dimensions*. Ann. Appl. Prob., 2011.

4. C. P. Robert and G. Casella. Introducing Monte Carlo methods with R. Use R!. Springer, New York, 2010.

5. G. O. Roberts, A. Gelman, and W. R. Gilks. *Weak convergence and optimal scaling of random walk Metropolis algorithms.* Ann. Appl. Probab., 1997.

6. L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 1998.