Generalized pCN Metropolis algorithms for Bayesian inference in Hilbert spaces

Oliver Ernst, Daniel Rudolf, Björn Sprungk

Data Assimilation and Inverse Problems (U Warwick)
23rd February, 2016

TECHNISCHE UNIVERSITÄT
CHEMNITZ
FAKULTÄT FÜR MATHEMATIK

Outline

1. Problem setting and Metropolis-Hastings algorithms in Hilbert spaces

2. The gpCN-Metropolis algorithm

3. Convergence

Groundwater flow modelling:

- PDE model for groundwater pressure head p, e.g.,

$$-\nabla \cdot (\mathrm{e}^{\kappa(x)} \nabla p(x)) = 0$$

- Noisy observations of κ and p at locations x_j, $j = 1, \dots, J$
- Interested in functional f of flux $\boldsymbol{u}(x) = -\mathrm{e}^{\kappa(x)} \nabla p(x)$ (e.g., breakthrough time of pollutants)

UQ approach:

- Model unknown coefficient κ as (Gaussian) random function
- Employ observational data to fit stochastic model for κ
- Compute expectations or probabilities for resulting random f

Groundwater flow modelling:

- PDE model for groundwater pressure head p, e.g.,

$$-\nabla \cdot (\mathrm{e}^{\kappa(x)} \nabla p(x)) = 0$$

- Noisy observations of κ and p at locations x_j, $j = 1, \ldots, J$
- Interested in functional f of flux $\boldsymbol{u}(x) = -\mathrm{e}^{\kappa(x)} \nabla p(x)$ (e.g., breakthrough time of pollutants)

UQ approach:

- Model unknown coefficient κ as (Gaussian) random function
- Employ observational data to fit stochastic model for κ
- Compute expectations or probabilities for resulting random f

- Random function in Hilbert space $\mathscr{H}$ with CONS $\{\phi_m\}_{m\in\mathbb{N}}$

$$\kappa(x,\omega) = \sum_{m\geq 1} \xi_m(\omega)\ \phi_m(x),$$

where $\boldsymbol{\xi} = (\xi_m)_{m\in\mathbb{N}}$ random vector in ℓ^2

- Employ measurements of κ to fit Gaussian prior μ_0 for κ resp. $\boldsymbol{\xi}$:

$$\boldsymbol{\xi} \sim \mu_0 = N(0, C_0) \qquad \text{on } \ell^2$$

- Conditioning prior μ_0 on noisy observations $\boldsymbol{p} \in \mathbb{R}^J$ of p

$$\boldsymbol{p} = G(\boldsymbol{\xi}) + \varepsilon, \qquad (\boldsymbol{\xi}, \varepsilon) \sim \mu_0 \otimes N(0, \Sigma)$$

results in posterior via Bayes' rule [Stuart, 2010]

$$\mu(\mathrm{d}\boldsymbol{\xi}) \propto \exp\left(-\Phi(\boldsymbol{\xi})\right)\ \mu_0(\mathrm{d}\boldsymbol{\xi}), \qquad \Phi(\boldsymbol{\xi}) = \frac{1}{2}|\boldsymbol{p} - G(\boldsymbol{\xi})|^2_{\Sigma^{-1}}.$$

- Random function in Hilbert space $\mathscr{H}$ with CONS $\{\phi_m\}_{m\in\mathbb{N}}$

$$\kappa(x,\omega) = \sum_{m\geq 1} \xi_m(\omega)\, \phi_m(x),$$

where $\boldsymbol{\xi} = (\xi_m)_{m\in\mathbb{N}}$ random vector in ℓ^2

- Employ measurements of κ to fit Gaussian prior μ_0 for κ resp. $\boldsymbol{\xi}$:

$$\boldsymbol{\xi} \sim \mu_0 = N(0, C_0) \qquad \text{on } \ell^2$$

- Conditioning prior μ_0 on noisy observations $\boldsymbol{p} \in \mathbb{R}^J$ of p

$$\boldsymbol{p} = G(\boldsymbol{\xi}) + \varepsilon, \qquad (\boldsymbol{\xi}, \varepsilon) \sim \mu_0 \otimes N(0, \Sigma)$$

results in posterior via Bayes' rule [Stuart, 2010]

$$\mu(\mathrm{d}\boldsymbol{\xi}) \propto \exp\left(-\Phi(\boldsymbol{\xi})\right)\, \mu_0(\mathrm{d}\boldsymbol{\xi}), \qquad \Phi(\boldsymbol{\xi}) = \frac{1}{2}|\boldsymbol{p} - G(\boldsymbol{\xi})|^2_{\Sigma^{-1}}.$$

- Random function in Hilbert space $\mathscr{H}$ with CONS $\{\phi_m\}_{m\in\mathbb{N}}$

$$\kappa(x,\omega) = \sum_{m\geq 1} \xi_m(\omega)\ \phi_m(x),$$

where $\boldsymbol{\xi} = (\xi_m)_{m\in\mathbb{N}}$ random vector in ℓ^2

- Employ measurements of κ to fit Gaussian prior μ_0 for κ resp. $\boldsymbol{\xi}$:

$$\boldsymbol{\xi} \sim \mu_0 = N(0, C_0) \qquad \text{on } \ell^2$$

- Conditioning prior μ_0 on noisy observations $\boldsymbol{p} \in \mathbb{R}^J$ of p

$$\boldsymbol{p} = G(\boldsymbol{\xi}) + \varepsilon, \qquad (\boldsymbol{\xi}, \varepsilon) \sim \mu_0 \otimes N(0, \Sigma)$$

results in posterior via Bayes' rule [Stuart, 2010]

$$\mu(\mathrm{d}\boldsymbol{\xi}) \propto \exp\left(-\Phi(\boldsymbol{\xi})\right)\ \mu_0(\mathrm{d}\boldsymbol{\xi}), \qquad \Phi(\boldsymbol{\xi}) = \frac{1}{2}|\boldsymbol{p} - G(\boldsymbol{\xi})|^2_{\Sigma^{-1}}.$$

- Employ MCMC sampling to compute $\mathbb{E}_\mu[f] = \int_{\mathscr{H}} f(\boldsymbol{\xi})\,\mu(\mathrm{d}\boldsymbol{\xi})$.

- Construct Markov chain $(\boldsymbol{\xi}_k)_{k\in\mathbb{N}}$ in $\mathscr{H} = \ell^2$ with transition kernel

$$M(\boldsymbol{\eta}, A) := \mathbb{P}(\boldsymbol{\xi}_{k+1} \in A | \boldsymbol{\xi}_k = \boldsymbol{\eta})$$

which is μ-reversible:

$$M(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta})\,\mu(\mathrm{d}\boldsymbol{\xi}) = M(\boldsymbol{\eta}, \mathrm{d}\boldsymbol{\xi})\,\mu(\mathrm{d}\boldsymbol{\eta})$$

- Then μ stationary measure of chain and (under suitable conditions)

$$\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{\xi}_k) \xrightarrow{n\to\infty} \mathbb{E}_\mu[f] \qquad \text{a.s.}$$

- μ-reversible Markov chain geometrically ergodic and $\boldsymbol{\xi}_1 \sim \mu$, then

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{\xi}_k) - \mathbb{E}_\mu[f]\right) \quad \xrightarrow{d} \quad N(0,\sigma_f^2),$$

  see [Roberts & Rosenthal, 1997], where

$$\sigma_f^2 = \sum_{k=-\infty}^{\infty} \gamma_f(k), \qquad \gamma_f(k) = \operatorname{Cov}\left(f(\boldsymbol{\xi}_1), f(\boldsymbol{\xi}_{1+k})\right)$$

- Rapid decay of autocovariance function γ_f yields high statistical efficiency.

- Common measure of efficiency is effective sample size:

$$\mathsf{ESS}_f(n) = n\,\frac{\sigma_f^2}{\operatorname{Var}_\mu(f)}.$$

Metropolis-Hastings algorithm for generating state $\boldsymbol{\xi}_{k+1}$ of Markov chain

1. Draw a new state $\boldsymbol{\eta}$ from proposal kernel $P(\boldsymbol{\xi}_k, \mathrm{d}\boldsymbol{\eta})$:

$$\boldsymbol{\eta} \sim P(\boldsymbol{\xi}_k).$$

2. Accept proposal $\boldsymbol{\eta}$ with acceptance probability $\alpha(\boldsymbol{\xi}_k, \boldsymbol{\eta})$, i.e., draw $a \sim \mathrm{Uni}[0, 1]$ and set

$$\boldsymbol{\xi}_{k+1} = \begin{cases} \boldsymbol{\eta}, & a \leq \alpha(\boldsymbol{\xi}_k, \boldsymbol{\eta}), \\ \boldsymbol{\xi}_k, & \text{otherwise.} \end{cases}$$

Transition kernel of resulting chain is called Metropolis kernel and reads

$$M(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) = \alpha(\boldsymbol{\xi}, \boldsymbol{\eta}) P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) + \left[1 - \int \alpha(\boldsymbol{\xi}, \boldsymbol{\zeta})\, P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\zeta})\right] \delta_{\boldsymbol{\xi}}(\mathrm{d}\boldsymbol{\eta}).$$

Metropolis-Hastings algorithm for generating state $\boldsymbol{\xi}_{k+1}$ of Markov chain

1. Draw a new state $\boldsymbol{\eta}$ from proposal kernel $P(\boldsymbol{\xi}_k, \mathrm{d}\boldsymbol{\eta})$:

$$\boldsymbol{\eta} \sim P(\boldsymbol{\xi}_k).$$

2. Accept proposal $\boldsymbol{\eta}$ with acceptance probability $\alpha(\boldsymbol{\xi}_k, \boldsymbol{\eta})$, i.e., draw $a \sim \mathrm{Uni}[0, 1]$ and set

$$\boldsymbol{\xi}_{k+1} = \begin{cases} \boldsymbol{\eta}, & a \leq \alpha(\boldsymbol{\xi}_k, \boldsymbol{\eta}), \\ \boldsymbol{\xi}_k, & \text{otherwise.} \end{cases}$$

Transition kernel of resulting chain is called Metropolis kernel and reads

$$M(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) = \alpha(\boldsymbol{\xi}, \boldsymbol{\eta}) P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) + \left[1 - \int \alpha(\boldsymbol{\xi}, \boldsymbol{\zeta})\, P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\zeta})\right] \, \delta_{\boldsymbol{\xi}}(\mathrm{d}\boldsymbol{\eta}).$$

To ensure μ-reversibility of Metropolis kernel we need to choose

$$\alpha(\boldsymbol{\xi}_k, \boldsymbol{\eta}) = \min\left\{1, \frac{\mathrm{d}\nu^\top}{\mathrm{d}\nu}(\boldsymbol{\xi}_k, \boldsymbol{\eta})\right\},$$

where $\nu(\mathrm{d}\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) := P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta})\, \mu(\mathrm{d}\boldsymbol{\xi})$ and $\nu^\top(\mathrm{d}\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) := \nu(\mathrm{d}\boldsymbol{\eta}, \mathrm{d}\boldsymbol{\xi})$.

Since μ and μ_0 are equivalent (Bayes' rule), a μ_0-reversible proposal P,

$$P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta})\, \mu_0(\mathrm{d}\boldsymbol{\xi}) \;=\; P(\boldsymbol{\eta}, \mathrm{d}\boldsymbol{\xi})\, \mu_0(\mathrm{d}\boldsymbol{\eta}),$$

will yield

$$\frac{\mathrm{d}\nu^\top}{\mathrm{d}\nu}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\boldsymbol{\eta}) \left(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\boldsymbol{\xi})\right)^{-1} = \exp\left(\Phi(\boldsymbol{\xi}) - \Phi(\boldsymbol{\eta})\right).$$

To ensure μ-reversibility of Metropolis kernel we need to choose

$$\alpha(\boldsymbol{\xi}_k, \boldsymbol{\eta}) = \min\left\{1, \frac{\mathrm{d}\nu^\top}{\mathrm{d}\nu}(\boldsymbol{\xi}_k, \boldsymbol{\eta})\right\},$$

where $\nu(\mathrm{d}\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) := P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta})\,\mu(\mathrm{d}\boldsymbol{\xi})$ and $\nu^\top(\mathrm{d}\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) := \nu(\mathrm{d}\boldsymbol{\eta}, \mathrm{d}\boldsymbol{\xi})$.

Since μ and μ_0 are equivalent (Bayes' rule), a μ_0-reversible proposal P,

$$P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta})\,\mu_0(\mathrm{d}\boldsymbol{\xi}) \;=\; P(\boldsymbol{\eta}, \mathrm{d}\boldsymbol{\xi})\,\mu_0(\mathrm{d}\boldsymbol{\eta}),$$

will yield

$$\frac{\mathrm{d}\nu^\top}{\mathrm{d}\nu}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\boldsymbol{\eta})\,\left(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\boldsymbol{\xi})\right)^{-1} = \exp\left(\Phi(\boldsymbol{\xi}) - \Phi(\boldsymbol{\eta})\right).$$

Gaussian random walk-proposal: $P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 C_0)$

- $s \in \mathbb{R}_+$ is a stepsize parameter typically tuned such that acceptance rate
$$\bar{\alpha} := \int_{\mathscr{H}} \int_{\mathscr{H}} \alpha(\boldsymbol{\xi}, \boldsymbol{\eta})\, P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta})\, \mu(\mathrm{d}\boldsymbol{\xi}) \approx 0.234.$$
- This proposal P is not μ_0 reversible.

Preconditioned Crank-Nicolson-proposal: $P(\xi) = N(\sqrt{1-s^2}\xi, s^2 C_0)$

- Derived from a CN-scheme for (the drift of) the SDE
$$\mathrm{d}\xi_t = (0 - \xi_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$
where B_t is C_0-Brownian motion in $\mathscr{H}$, see [Cotter et al., 2013].
- This proposal is μ_0 reversible.

Gaussian random walk-proposal: $P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 C_0)$

- $s \in \mathbb{R}_+$ is a stepsize parameter typically tuned such that acceptance rate
$$\bar{\alpha} := \int_{\mathcal{H}} \int_{\mathcal{H}} \alpha(\boldsymbol{\xi}, \boldsymbol{\eta})\, P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta})\, \mu(\mathrm{d}\boldsymbol{\xi}) \approx 0.234.$$
- This proposal P is not μ_0 reversible.

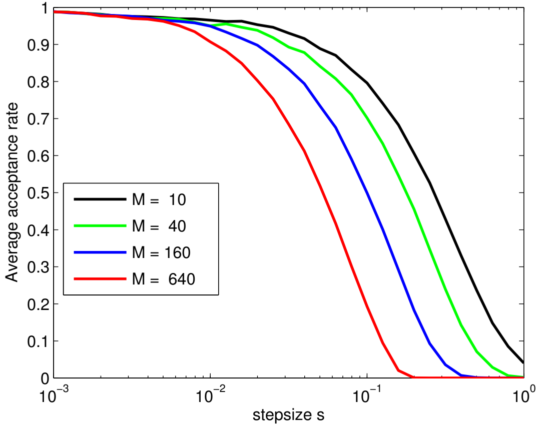
Preconditioned Crank-Nicolson-proposal: $P(\boldsymbol{\xi}) = N(\sqrt{1-s^2}\boldsymbol{\xi}, s^2 C_0)$

- Derived from a CN-scheme for (the drift of) the SDE
$$\mathrm{d}\boldsymbol{\xi}_t = (0 - \boldsymbol{\xi}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$
where B_t is C_0-Brownian motion in $\mathcal{H}$, see [Cotter et al., 2013].
- This proposal is μ_0 reversible.

Problem: Bayesian inference in 2D groundwater flow model.

Acceptance rate vs. stepsize s for different dimensions M of $\boldsymbol{\xi} \in \mathbb{R}^M$.

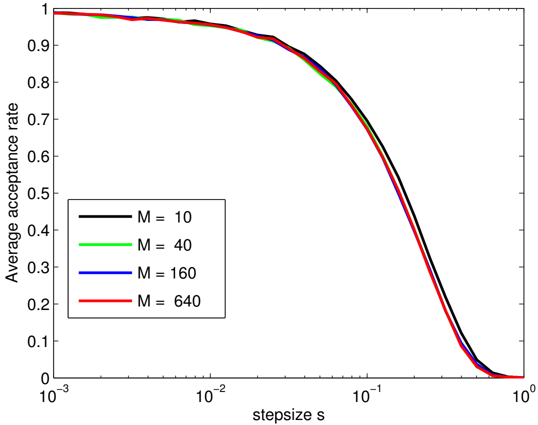
Random walk-proposal $P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 C_0)$



Problem: Bayesian inference in 2D groundwater flow model.

Acceptance rate vs. stepsize s for different dimensions M of $\boldsymbol{\xi} \in \mathbb{R}^M$.

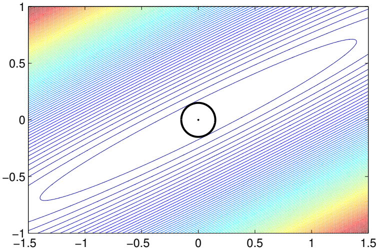
pCN-proposal $P(\boldsymbol{\xi}) = N(\sqrt{1-s^2}\boldsymbol{\xi}, s^2 C_0)$



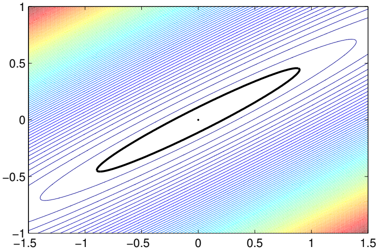
Observation:

Higher statistical efficiency when proposal employs same covariance as target measure μ, see [Tierney, 1994], [Roberts & Rosenthal, 2001], ...

Example: $\mu = N(\mathbf{0}, \boldsymbol{C})$ in 2D, MH with different proposal covariances



$$P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 \boldsymbol{I})$$

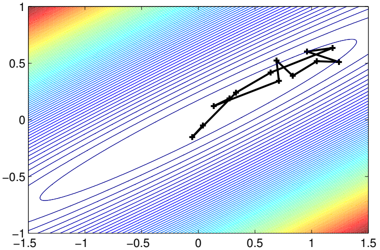


$$P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 \boldsymbol{C})$$

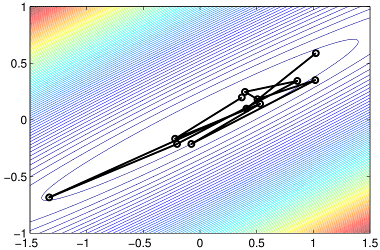
Observation:

Higher statistical efficiency when proposal employs same covariance as target measure μ, see [Tierney, 1994], [Roberts & Rosenthal, 2001], ...

Example: $\mu = N(\mathbf{0}, \boldsymbol{C})$ in 2D, MH with different proposal covariances



$$P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 \boldsymbol{I})$$

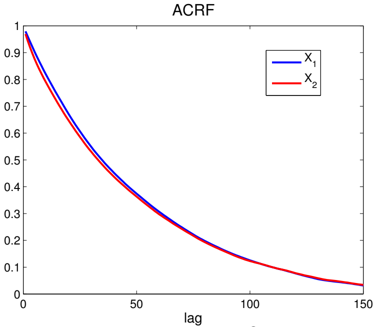


$$P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 \boldsymbol{C})$$

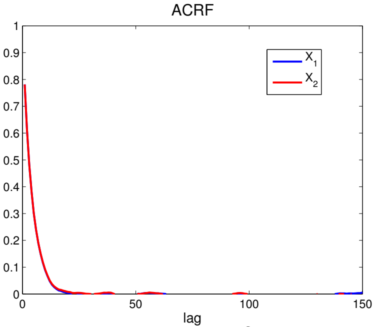
Observation:

Higher statistical efficiency when proposal employs same covariance as target measure μ, see [Tierney, 1994], [Roberts & Rosenthal, 2001], ...

Example: $\mu = N(\mathbf{0}, \boldsymbol{C})$ in 2D, MH with different proposal covariances



$P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 \boldsymbol{I})$



$P(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 \boldsymbol{C})$

- If forward map G were linear,then

$$\mu = N(m, C), \qquad C = (C_0^{-1} + G^* \Sigma^{-1} G)^{-1}.$$

- **Idea:** Linearization of nonlinear G at ξ_0

$$G(\xi) \approx \widetilde{G}(\xi) := G(\xi_0) + J\xi, \qquad J = \nabla G(\xi_0)$$

yields approximation to posterior covariance

$$C \approx \widetilde{C} = (C_0^{-1} + J^* \Sigma^{-1} J)^{-1}.$$

- Good choice for ξ_0 might be maximum a posteriori estimate:

$$\xi_{\mathrm{MAP}} = \operatorname*{argmin}_{\xi} \left(|\boldsymbol{p} - G(\xi)|^2_{\Sigma^{-1}} + \|C_0^{-1/2}\xi\|^2 \right).$$

- If forward map G were linear,then

$$\mu = N(m, C), \qquad C = (C_0^{-1} + G^* \Sigma^{-1} G)^{-1}.$$

- **Idea:** Linearization of nonlinear G at $\boldsymbol{\xi}_0$

$$G(\boldsymbol{\xi}) \approx \widetilde{G}(\boldsymbol{\xi}) := G(\boldsymbol{\xi}_0) + J\boldsymbol{\xi}, \qquad J = \nabla G(\boldsymbol{\xi}_0)$$

yields approximation to posterior covariance

$$C \approx \widetilde{C} = (C_0^{-1} + J^* \Sigma^{-1} J)^{-1}.$$

- Good choice for ξ_0 might be maximum a posteriori estimate:

$$\xi_{\mathrm{MAP}} = \operatorname*{argmin}_{\xi} \left(|\boldsymbol{p} - G(\boldsymbol{\xi})|^2_{\Sigma^{-1}} + \|C_0^{-1/2} \boldsymbol{\xi}\|^2 \right).$$

- If forward map G were linear,then

$$\mu = N(m, C), \qquad C = (C_0^{-1} + G^*\Sigma^{-1}G)^{-1}.$$

- **Idea:** Linearization of nonlinear G at $\boldsymbol{\xi}_0$

$$G(\boldsymbol{\xi}) \approx \widetilde{G}(\boldsymbol{\xi}) := G(\boldsymbol{\xi}_0) + J\boldsymbol{\xi}, \qquad J = \nabla G(\boldsymbol{\xi}_0)$$

yields approximation to posterior covariance

$$C \approx \widetilde{C} = (C_0^{-1} + J^*\Sigma^{-1}J)^{-1}.$$

- Good choice for $\boldsymbol{\xi}_0$ might be maximum a posteriori estimate:

$$\boldsymbol{\xi}_{\mathrm{MAP}} = \operatorname*{argmin}_{\boldsymbol{\xi}} \left(|\boldsymbol{p} - G(\boldsymbol{\xi})|^2_{\Sigma^{-1}} + \|C_0^{-1/2}\boldsymbol{\xi}\|^2 \right).$$

- Class of admissible proposal covariances:

$$C_\Gamma = (C_0^{-1} + \Gamma)^{-1}, \qquad \Gamma \text{ positive, self-adjoint and bounded}$$

- Define generalized pCN-proposal kernel as

$$P_\Gamma(\boldsymbol{\xi}) = N(A_\Gamma \boldsymbol{\xi}, s^2 C_\Gamma),$$

cf. operator weighted proposals [Law, 2013] and [Cui et al., 2014]

- Enforcing μ_0-reversibility of P_Γ – as for pCN-proposal P_0 – yields

$$A_\Gamma = C_0^{1/2} \sqrt{I - s^2 (I + H_\Gamma)^{-1}}\, C_0^{-1/2}, \qquad H_\Gamma := C_0^{1/2} \Gamma C_0^{1/2}.$$

- Boundedness of A_Γ on $\mathscr{H}$ not obvious

- Class of admissible proposal covariances:

$$C_\Gamma = (C_0^{-1} + \Gamma)^{-1}, \qquad \Gamma \text{ positive, self-adjoint and bounded}$$

- Define generalized pCN-proposal kernel as

$$P_\Gamma(\boldsymbol{\xi}) = N(A_\Gamma \boldsymbol{\xi}, s^2 C_\Gamma),$$

cf. operator weighted proposals [Law, 2013] and [Cui et al., 2014]

- Enforcing μ_0-reversibility of P_Γ – as for pCN-proposal P_0 – yields

$$A_\Gamma = C_0^{1/2} \sqrt{I - s^2 (I + H_\Gamma)^{-1}}\, C_0^{-1/2}, \qquad H_\Gamma := C_0^{1/2} \Gamma C_0^{1/2}.$$

- Boundedness of A_Γ on $\mathscr{H}$ not obvious

gpCN Metropolis

The Metropolis algorihtm with the gpCN-proposal kernel

$$P_\Gamma(\boldsymbol{\xi}) = N(A_\Gamma \boldsymbol{\xi}, s^2 C_\Gamma),$$

where $s \in (0,1)$, and the acceptance probability

$$\alpha(\boldsymbol{\xi}, \boldsymbol{\eta}) = \min\{1, \exp(\Phi(\boldsymbol{\xi}) - \Phi(\boldsymbol{\eta}))\}$$

is well-defined and yields a μ-reversible Markov chain in $\mathscr{H}$.

It is called gpCN Metropolis and its Metropolis kernel denoted by M_Γ.

Note, for $\Gamma = 0$ we recover the pCN Metropolis algorithm with kernel M_0.

- **Setting**
  - 1D model: $\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{\kappa(x)}\,\frac{\mathrm{d}p}{\mathrm{d}x}(x)\right) = 0, \quad p(0) = 0,\; p(1) = 2$
  - Observations: $\boldsymbol{p} = \left[\,p(0.2j)\,\right]_{j=1}^{4} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma_\varepsilon^2 I)$
  - Prior: κ Brownian bridge on $[0,1]$, i.e.,

$$\kappa(x, \boldsymbol{\xi}(\omega)) \approx \frac{\sqrt{2}}{\pi}\sum_{m=1}^{M} \xi_m(\omega)\sin(m\pi x), \qquad \xi_m \sim N(0, m^{-2})$$

  - Quantity of interest: $f(\boldsymbol{\xi}) = \int_0^1 \exp(\kappa(x, \boldsymbol{\xi}))\,\mathrm{d}x$

- **Proposals for MH-MCMC**

- **Results**

- **Setting**
- **Proposals for MH-MCMC**

Gaussian random walk (RW): $P_1(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 C_0)$

pCN: $P_2(\boldsymbol{\xi}) = N(\sqrt{1-s^2}\boldsymbol{\xi}, s^2 C_0)$

Gauss-Newton RW (GN-RW): $P_3(\boldsymbol{\xi}) = N(\boldsymbol{\xi}, s^2 C_\Gamma)$

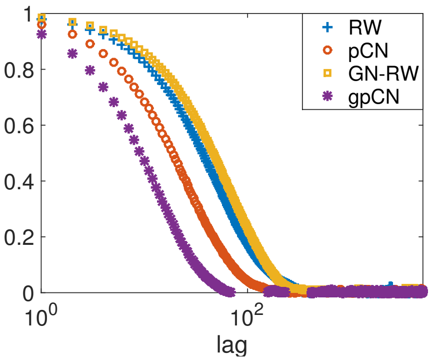
gpCN: $P_4(\boldsymbol{\xi}) = N(A_\Gamma \boldsymbol{\xi}, s^2 C_\Gamma)$

$$\Gamma = \sigma_\varepsilon^{-2} J J^\top, \qquad J = \nabla G(\boldsymbol{\xi}_{\mathrm{MAP}})$$

- **Results**

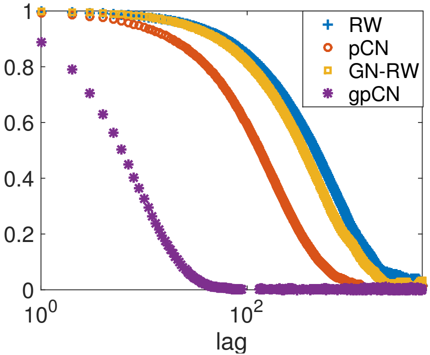
- **Setting**
- **Proposals for MH-MCMC**
- **Results**

Autocorrelation for $M = 50$, $\sigma_\varepsilon = 0.1$



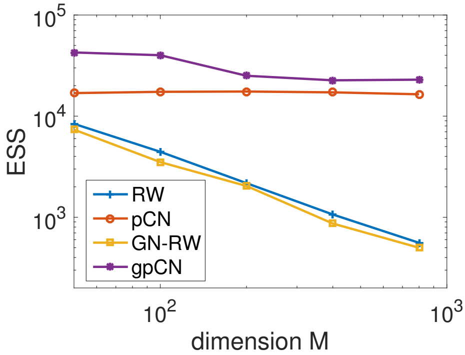
- **Setting**
- **Proposals for MH-MCMC**
- **Results**

Autocorrelation for $M = 400$, $\sigma_\varepsilon = 0.01$



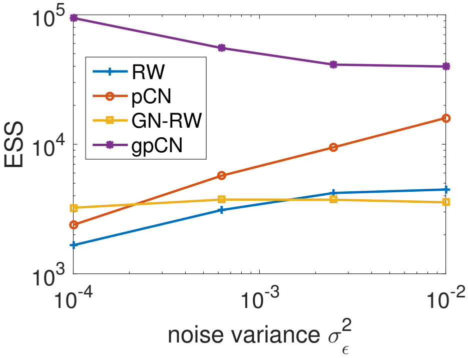
- **Setting**
- **Proposals for MH-MCMC**
- **Results**

Effective sample size vs. dimension



- **Setting**
- **Proposals for MH-MCMC**
- **Results**

Effective sample size vs. noise variance



Consider Markov chain $\{\boldsymbol{\xi}_k\}_{k\in\mathbb{N}_0}$ with μ-reversible transition kernel M.

- Kernel M is L^2_μ-geometrically ergodic if $r > 0$ exists such that

$$\|\mu - \nu M^k\|_{\mathrm{TV}} \le C_\nu \exp(-r\,k) \qquad \forall \nu : \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^2_\mu(\mathscr{H}).$$

- Define associated Markov operator $\mathrm{M} : L^2_\mu \to L^2_\mu$ by

$$\mathrm{M}f(\boldsymbol{\xi}) := \int_{\mathscr{H}} f(\boldsymbol{\eta})\, M(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}).$$

- L^2_μ-spectral gap of operator M: $\quad \mathrm{gap}_\mu(\mathrm{M}) := 1 - \|\mathrm{M} - \mathbb{E}_\mu\|_{L^2_\mu \to L^2_\mu}$

- M is L^2_μ-geometrically ergodic iff $\mathrm{gap}_\mu(\mathrm{M}) > 0$.

- For pCN-Metropolis kernel M_0 an L^2_μ-spectral gap was proven in [Hairer et al., 2014] under certain conditions on $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}$.

- **Our Strategy:**

  Take a comparative approach and relate $\mathrm{gap}_\mu(\mathrm{M}_\Gamma)$ to $\mathrm{gap}_\mu(\mathrm{M}_0)$ in order to prove

$$0 < \mathrm{gap}_\mu(\mathrm{M}_0) \quad \Longrightarrow \quad 0 < \mathrm{gap}_\mu(\mathrm{M}_\Gamma).$$

- **Idea behind:**

  If kernels M_0 and M_Γ do not "differ" too much, then (maybe) they inherit convergence properties from each other.

Theorem (Comparison of spectral gaps)

For $i = 1, 2$ let M_i be μ-reversible Metropolis kernels with proposals P_i and acceptance probability α.

Assume that

1. *the associated Markov operators M_i are positive,*
2. *the Radon-Nikodym derivative $\rho(\boldsymbol{\xi};\boldsymbol{\eta}) := \frac{\mathrm{d}P_1(\boldsymbol{\xi})}{\mathrm{d}P_2(\boldsymbol{\xi})}(\boldsymbol{\eta})$ exists,*
3. *and for a $\beta > 1$ we have*

$$\sup_{\mu(A)\in(0,\frac{1}{2}]} \frac{\int_A \int_{A^c} \rho^\beta(\boldsymbol{\xi};\boldsymbol{\eta})\, P_2(\boldsymbol{\xi},\mathrm{d}\boldsymbol{\eta})\,\mu(\mathrm{d}\boldsymbol{\xi})}{\mu(A)} < \infty,$$

Then

$$\mathrm{gap}_\mu(\mathrm{M}_1)^{2\beta} \leq K_\beta\, \mathrm{gap}_\mu(\mathrm{M}_2)^{\beta-1}.$$

Lemma (Positivity of Markov operators)

The Markov operator M_Γ *associated to gpCN-Metropolis is positive for any admissible* Γ.

Lemma (Absolute continuity, integrability of density)

There exists a density between the pCN- and gpCN-proposal

$$\rho_\Gamma(\boldsymbol{\xi};\boldsymbol{\eta}) := \frac{\mathrm{d}P_0(\boldsymbol{\xi})}{\mathrm{d}P_\Gamma(\boldsymbol{\xi})}(\boldsymbol{\eta})$$

and constants $b, K < \infty$ *such that for* $\beta < 1 + \frac{1}{2\|H_\Gamma\|}$

$$\int_{\mathscr{H}} \rho_\Gamma^\beta(\boldsymbol{\xi};\boldsymbol{\eta})\ P_\Gamma(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) \le K\ \exp\left(\frac{b^2}{2}\|\boldsymbol{\xi}\|^2\right).$$

- For $R > 0$ set $B_R := \{\boldsymbol{\xi} \in \mathscr{H} : \|\boldsymbol{\xi}\| < R\}$ and define

$$\mu_R(\mathrm{d}\boldsymbol{\xi}) := \frac{1}{\mu(B_R)} \mathbf{1}_{B_R}(\boldsymbol{\xi})\, \mu(\mathrm{d}\boldsymbol{\xi}).$$

- For Metropolis kernel M with proposal P and acceptance probability α set

$$M_R(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) := \alpha_R(\boldsymbol{\xi}, \boldsymbol{\eta}) P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) + \left[1 - \int_{\mathscr{H}} \alpha_R(\boldsymbol{\xi}, \boldsymbol{\zeta}) P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\zeta})\right] \delta_{\boldsymbol{\xi}}(\mathrm{d}\boldsymbol{\eta})$$

where $\alpha_R(\boldsymbol{\xi}, \boldsymbol{\eta}) := \mathbf{1}_{B_R}(\boldsymbol{\eta}) \alpha(\boldsymbol{\xi}, \boldsymbol{\eta})$.

Lemma

If M *is self-adjoint and positive on* $L^2_\mu(\mathscr{H})$*, then so is* M_R *on* $L^2_{\mu_R}(\mathscr{H})$ *and*

$$\mathrm{gap}_{\mu_R}(\mathrm{M}_R) \geq \mathrm{gap}_\mu(\mathrm{M}) - \sup_{\boldsymbol{\xi} \in B_R} M(\boldsymbol{\xi}, B_R^c).$$

- For $R > 0$ set $B_R := \{\boldsymbol{\xi} \in \mathscr{H} : \|\boldsymbol{\xi}\| < R\}$ and define

$$\mu_R(\mathrm{d}\boldsymbol{\xi}) := \frac{1}{\mu(B_R)} \mathbf{1}_{B_R}(\boldsymbol{\xi})\, \mu(\mathrm{d}\boldsymbol{\xi}).$$

- For Metropolis kernel M with proposal P and acceptance probability α set

$$M_R(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) := \alpha_R(\boldsymbol{\xi}, \boldsymbol{\eta}) P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) + \left[1 - \int_{\mathscr{H}} \alpha_R(\boldsymbol{\xi}, \boldsymbol{\zeta}) P(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\zeta})\right] \delta_{\boldsymbol{\xi}}(\mathrm{d}\boldsymbol{\eta})$$

where $\alpha_R(\boldsymbol{\xi}, \boldsymbol{\eta}) := \mathbf{1}_{B_R}(\boldsymbol{\eta}) \alpha(\boldsymbol{\xi}, \boldsymbol{\eta})$.

Lemma

If M is self-adjoint and positive on $L^2_\mu(\mathscr{H})$, then so is M_R on $L^2_{\mu_R}(\mathscr{H})$ and

$$\mathrm{gap}_{\mu_R}(\mathrm{M}_R) \geq \mathrm{gap}_\mu(\mathrm{M}) - \sup_{\boldsymbol{\xi} \in B_R} M(\boldsymbol{\xi}, B_R^c).$$

Theorem (Spectral gap of gpCN-Metropolis for restricted measure)

If

$$\mathrm{gap}_\mu(\mathrm{M}_0) > 0,$$

then for any admissible Γ *and any* $\epsilon > 0$ *there exists a number* $R < \infty$ *such that*

$$\|\mu - \mu_R\|_{TV} < \epsilon \quad \textit{and} \quad \mathrm{gap}_{\mu_R}(\mathrm{M}_{\Gamma,R}) > 0.$$

Proof employs previous lemmas and comparison theorem to show

$$\mathrm{gap}_\mu(\mathrm{M}_0) > 0 \quad \Longrightarrow \quad \mathrm{gap}_{\mu_R}(\mathrm{M}_{0,R}) > 0 \quad \Longrightarrow \quad \mathrm{gap}_{\mu_R}(\mathrm{M}_{\Gamma,R}) > 0.$$

Density $\rho_\Gamma(\boldsymbol{\xi}) = \frac{\mathrm{d}P_0(\boldsymbol{\xi})}{\mathrm{d}P_\Gamma(\boldsymbol{\xi})}$ allows for state-dependent proposal covariances:

- Let $\boldsymbol{\xi} \mapsto \Gamma(\boldsymbol{\xi})$ be measurable and define proposal kernel

$$P_{\mathrm{loc}}(\boldsymbol{\xi}) := N(A_{\Gamma(\boldsymbol{\xi})}\boldsymbol{\xi}, s^2 C_{\Gamma(\boldsymbol{\xi})}).$$

- Following an approach as in [Beskos et al., 2008] we get

$$P_{\mathrm{loc}}(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta})\,\mu_0(\mathrm{d}\boldsymbol{\xi}) = \frac{\rho_{\Gamma(\boldsymbol{\eta})}(\boldsymbol{\eta},\boldsymbol{\xi})}{\rho_{\Gamma(\boldsymbol{\xi})}(\boldsymbol{\xi},\boldsymbol{\eta})}\,P_{\mathrm{loc}}(\boldsymbol{\eta}, \mathrm{d}\boldsymbol{\xi})\,\mu_0(\mathrm{d}\boldsymbol{\eta}).$$

- Obtain μ-reversible MH algorithm with proposal P_{loc} and

$$\alpha_{\mathrm{loc}}(\boldsymbol{\xi},\boldsymbol{\eta}) := \min\left\{1, \exp(\Phi(\boldsymbol{\eta}) - \Phi(\boldsymbol{\xi}))\,\frac{\rho_{\Gamma(\boldsymbol{\xi})}(\boldsymbol{\xi},\boldsymbol{\eta})}{\rho_{\Gamma(\boldsymbol{\eta})}(\boldsymbol{\eta},\boldsymbol{\xi})}\right\}$$

- Same approach works for $P'_{\mathrm{loc}}(\boldsymbol{\xi}) := N(\sqrt{1-s^2}\boldsymbol{\xi}, s^2 C_{\Gamma(\boldsymbol{\xi})}).$

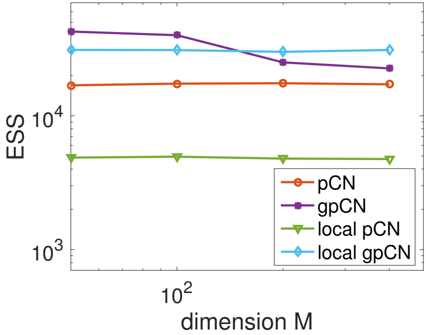
Same problem setting as before, but now also consider "local" proposals

local pCN: $P_5(\boldsymbol{\xi}) = N(\sqrt{1-s^2}\boldsymbol{\xi}, s^2 C_{\Gamma(\boldsymbol{\xi})})$

local gpCN: $P_6(\boldsymbol{\xi}) = N(A_{\Gamma(\boldsymbol{\xi})}\boldsymbol{\xi}, s^2 C_{\Gamma(\boldsymbol{\xi})})$

with $\Gamma(\boldsymbol{\xi}) = \nabla G(\boldsymbol{\xi})^\top \Sigma^{-1} \nabla G(\boldsymbol{\xi})$.

Effective sample size vs. dimension



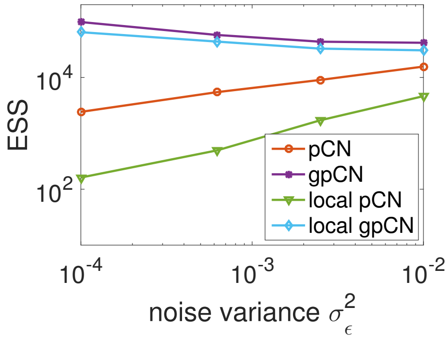
Same problem setting as before, but now also consider “local” proposals

local pCN: $P_5(\boldsymbol{\xi}) = N(\sqrt{1-s^2}\boldsymbol{\xi}, s^2 C_{\Gamma(\boldsymbol{\xi})})$

local gpCN: $P_6(\boldsymbol{\xi}) = N(A_{\Gamma(\boldsymbol{\xi})}\boldsymbol{\xi}, s^2 C_{\Gamma(\boldsymbol{\xi})})$

with $\Gamma(\boldsymbol{\xi}) = \nabla G(\boldsymbol{\xi})^\top \Sigma^{-1} \nabla G(\boldsymbol{\xi})$.

Effective sample size vs. noise variance



- Generalized pCN-Metropolis allows to beneficially employ approximation of posterior covariance for proposal kernel
- gpCN seems to perform independent of dimension and noise variance
- Geometric ergodicity of gpCN-Metropolis proven by general framework for relating spectral gaps of Metropolis algorithms
- gpCN-Metropolis with state-dependent proposal covariance possible

Some open issues:

- Proof of higher efficiency of gpCN-Metropolis
- Proof and theoretical understanding of variance-independence

- Generalized pCN-Metropolis allows to beneficially employ approximation of posterior covariance for proposal kernel
- gpCN seems to perform independent of dimension and noise variance
- Geometric ergodicity of gpCN-Metropolis proven by general framework for relating spectral gaps of Metropolis algorithms
- gpCN-Metropolis with state-dependent proposal covariance possible

Some open issues:

- Proof of higher efficiency of gpCN-Metropolis
- Proof and theoretical understanding of variance-independence

A. Beskos, G. Roberts, A. Stuart, and J. Voss.
MCMC methods for diffusion bridges.
Stoch. Dynam., 8(3):319–350, 2008.

S. Cotter, G. Roberts, A. Stuart, and D. White.
MCMC methods for functions: Modifying old algorithms to make them faster.
Stat. Sci., 28(3):283–464, 2013.

T. Cui, K. Law, and Y. Marzouk.
Dimension-independent likelihood-informed MCMC.
arXiv:1411.3688, 2014.

M. Hairer, A. Stuart, and S. Vollmer.
Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions.
Ann. Appl. Probab., 24(6):2455–2490, 2014.

D. Rudolf, B. Sprungk.
On a generalization of the preconditioned Crank-Nicolson Metropolis algorithm.
arXiv:1504.03461, 2015.

A. Stuart.
Inverse problems: a Bayesian perspective.
Acta Numer., 19:451–559, 2010.