Metropolis–Hastings Algorithms For Bayesian Inference In Hilbert Spaces

Oliver Ernst, Daniel Rudolf, Claudia Schillings, Björn Sprungk

Institute of Mathematics, University of Mannheim

3rd GAMM AG UQ Workshop, TU Dortmund
March 13th, 2018

Outline

1. Motivation for Bayesian Inference in Hilbert Spaces
2. Metropolis–Hastings Algorithms in Hilbert Spaces
3. Analysis of Metropolis–Hastings Algorithms

Uncertainty Quantification in Groundwater Flow

Groundwater flow modelling:

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

$$-\nabla \cdot (\mathrm{e}^{u(\mathbf{x})} \nabla p(\mathbf{x})) = 0 \qquad \text{in } D$$

  with uncertain $u \in C(D)$

- Noisy observations of u and p at locations $\mathbf{x}_j \in D$, $j = 1, \ldots, J$

- Functional f of flux $-\mathrm{e}^{u(\mathbf{x})} \nabla p(\mathbf{x})$, e.g., exit time of pollutants

Uncertainty Quantification in Groundwater Flow

Groundwater flow modelling:

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

$$-\nabla \cdot (\mathrm{e}^{u(\mathbf{x})} \nabla p(\mathbf{x})) = 0 \qquad \text{in } D$$

  with uncertain $u \in C(D)$

- Noisy observations of u and p at locations $\mathbf{x}_j \in D$, $j = 1, \ldots, J$
- Functional f of flux $-\mathrm{e}^{u(\mathbf{x})} \nabla p(\mathbf{x})$, e.g., exit time of pollutants

UQ approach: (underlying probability space $(\Omega, \mathscr{A}, \mathbb{P})$)

- Model uncertain u by (Gaussian) random field $u(\cdot, \omega) \in C(D)$ a.s.
- Employ observational data to fit stochastic model for u
- Compute expectations or probabilities for resulting random $f(\omega)$

Stochastic Model for u

- Continuous random field yields random variable $U\colon \Omega \to L^2(D)$ with

$$U(\omega) = \sum_{m\geq 1} \xi_m(\omega)\, \phi_m, \qquad \{\phi_m\}_{m\in\mathbb{N}} \text{ ONS in } L^2(D) \tag{KLE}$$

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

- Convenient: fit Gaussian prior μ_0 for u resp. $\boldsymbol{\xi}$ given data $u(\mathbf{x}_j)$ (geostatistics):

$$\boldsymbol{\xi} \sim \mu_0 = N(m_0, C_0) \qquad \text{on } \ell^2 =: \mathscr{H}$$

- Incorporate indirect data $p(\mathbf{x}_j)$ by conditioning prior μ_0 on it (Bayes)

- Sample from resulting posterior measure μ to compute statistics of $f(\boldsymbol{\xi})$

Bayesian Inference

- Let $G\colon \mathscr{H} \to \mathbb{R}^J$ denote a forward map, here:

$$\boldsymbol{\xi} \xrightarrow{KLE} u \xrightarrow{PDE} p \xrightarrow{Observation} (p(\mathbf{x}_j))_{j=1}^J$$

- Let $\mathbf{d} \in \mathbb{R}^J$ be a realization of noisy observable

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

Theorem (e.g., [Stuart, 2010])

If G is measurable and $\boldsymbol{\xi} \perp\!\!\!\perp \varepsilon$, then the conditional or posterior measure μ of $\boldsymbol{\xi} \sim \mu_0$ given that $G(\boldsymbol{\xi}) + \varepsilon = \mathbf{d}$ is

$$\mu(\mathrm{d}\boldsymbol{\xi}) \propto \exp\left(-\Phi(\boldsymbol{\xi})\right) \mu_0(\mathrm{d}\boldsymbol{\xi}),$$

where $\quad \Phi(\boldsymbol{\xi}) := \frac{1}{2}|\mathbf{d} - G(\boldsymbol{\xi})|^2_{\Sigma^{-1}} \quad$ *is negative log-likelihood.*

Markov Chain Monte Carlo (MCMC)

- **Basic idea:** construct Markov chain $(\xi_k)_{k\in\mathbb{N}}$ with $\xi_k \xrightarrow{\mathscr{D}} \mu$ as $k \to \infty$...

Markov Chain Monte Carlo (MCMC)

- **Basic idea:** construct Markov chain $(\xi_k)_{k\in\mathbb{N}}$ with $\xi_k \xrightarrow{\mathscr{D}} \mu$ as $k \to \infty$...
- ... by constructing a transition kernel

$$K(x, A) := \mathbb{P}(\xi_{k+1} \in A \,|\, \xi_k = x), \qquad x \in \mathscr{H},\ A \in \mathscr{B}(\mathscr{H}),$$

which is μ-invariant, i.e.,

$$\mu K = \int K(x, \cdot)\mu(\mathrm{d}x) = \mu$$

- Then, under suitable conditions, there holds for $f \in L^1_\mu(\mathbb{R})$

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

Efficiency of MCMC

- Autocorrelation of Markov chain effects efficiency: given $\xi_1 \sim \mu$

$$\lim_{n\to\infty} n\,\mathbb{E}\left[\left|S_n(f) - \mathbb{E}_\mu[f]\right|^2\right] = \mathrm{Var}_\mu(f) + 2\sum_{j=1}^{\infty} \mathrm{Cov}(f(\xi_1), f(\xi_{1+j}))$$

- Effective sample size (ESS): number of independent samples which yield same mean squared error

Efficiency of MCMC

- Autocorrelation of Markov chain effects efficiency: given $\xi_1 \sim \mu$

$$\lim_{n\to\infty} n\,\mathbb{E}\left[\left|S_n(f) - \mathbb{E}_\mu[f]\right|^2\right] = \mathrm{Var}_\mu(f) + 2\sum_{j=1}^{\infty} \mathrm{Cov}(f(\xi_1), f(\xi_{1+j}))$$

- Effective sample size (ESS): number of independent samples which yield same mean squared error

- Many common MCMC algorithms show decreasing efficiency for
  - (a) increasing dimension of state space, i.e., $\mathscr{H} = \mathbb{R}^M$ and $M \to \infty$
  - (b) decreasing noise $\mathrm{Var}(\varepsilon) \to 0$, i.e., posterior μ more concentrated

- We will address and resolve both issues in the following

The Metropolis–Hastings (MH) Algorithm

Metropolis–Hastings algorithm [Metropolis et al., 1953] [Hastings, 1970]

Given current state $\xi_k = x$,

1. draw new state y according to proposal kernel $P(x, \cdot)$: $\quad Y_k \sim P(x)$
2. accept proposed y with acceptance probability $\alpha(x, y)$, i.e., set

$$\xi_{k+1} = \begin{cases} y, & \text{with probability } \alpha(x, y), \\ x, & \text{with probability } 1 - \alpha(x, y). \end{cases}$$

The Metropolis–Hastings (MH) Algorithm

Metropolis–Hastings algorithm [Metropolis et al., 1953] [Hastings, 1970]

Given current state $\xi_k = x$,

1. draw new state y according to proposal kernel $P(x, \cdot)$: $\quad Y_k \sim P(x)$
2. accept proposed y with acceptance probability $\alpha(x, y)$, i.e., set

$$\xi_{k+1} = \begin{cases} y, & \text{with probability } \alpha(x, y), \\ x, & \text{with probability } 1 - \alpha(x, y). \end{cases}$$

[Tierney, 1998]: μ-invariance ensured if

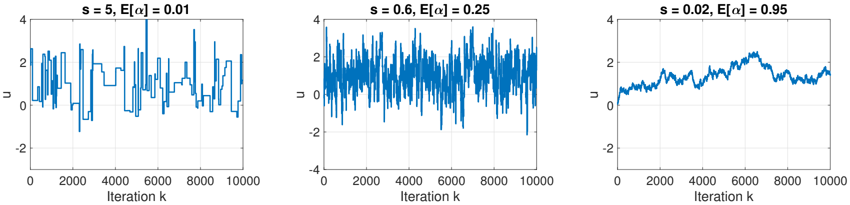
$$\alpha(x, y) = \min\left\{1, \frac{\mathrm{d}\nu^\top}{\mathrm{d}\nu}(x, y)\right\},$$

where $\nu(\mathrm{d}x, \mathrm{d}y) := P(x, \mathrm{d}y)\, \mu(\mathrm{d}x)$ and $\nu^\top(\mathrm{d}x, \mathrm{d}y) := \nu(\mathrm{d}y, \mathrm{d}x)$.

Gaussian Random Walk-MH

Gaussian Random Walk-MH: proposal $P(x) = N(x, s^2 C_0)$

- $s > 0$ tunable stepsize parameter:



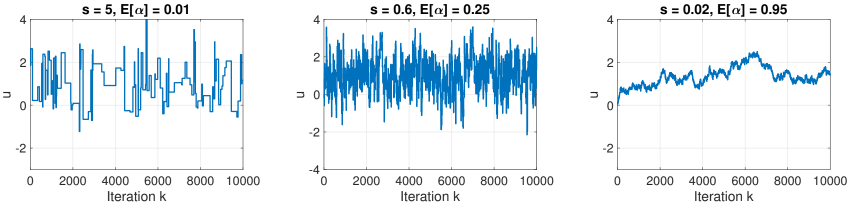
- If $\mathscr{H} = \mathbb{R}^M$ and $\pi\colon \mathbb{R}^M \to (0,\infty)$ Lebesgue density of μ, then:

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

Gaussian Random Walk-MH

Gaussian Random Walk-MH: proposal $P(x) = N(x, s^2 C_0)$

- $s > 0$ tunable stepsize parameter:



- If $\mathscr{H} = \mathbb{R}^M$ and $\pi \colon \mathbb{R}^M \to (0, \infty)$ Lebesgue density of μ, then:

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

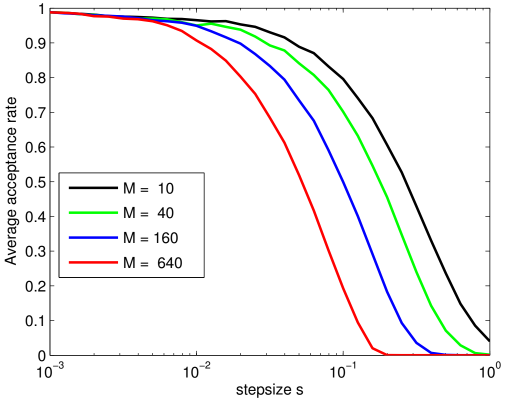
- However, for fixed s there holds $\mathbb{E}\left[\alpha(\xi_k, Y_k)\right] \xrightarrow{M\to\infty} 0$

Numerical Example

Problem: Bayesian inference in 2D groundwater flow model

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

Random walk-proposal $P(x) = N(x, s^2 C_0)$

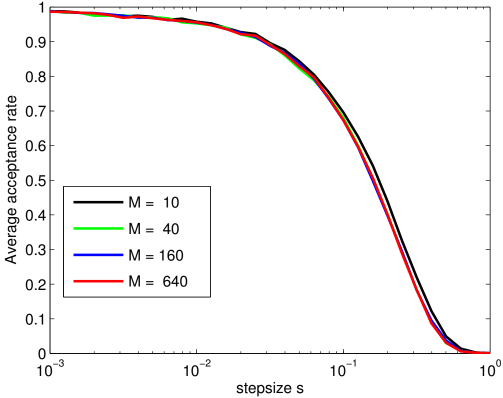


Numerical Example

Problem: Bayesian inference in 2D groundwater flow model

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

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



MH Algorithms in Hilbert Spaces

- Due to Bayes' formula, $\alpha(x,y)$ is well-defined **iff** $\frac{\mathrm{d}\nu_0^\top}{\mathrm{d}\nu_0}$ exists for

$$\nu_0(\mathrm{d}x,\mathrm{d}y) := P(x,\mathrm{d}y)\,\mu_0(\mathrm{d}x), \qquad \nu_0^\top(\mathrm{d}x,\mathrm{d}y) := \nu_0(\mathrm{d}y,\mathrm{d}x)$$

- [Cotter et al., 2013]: RW proposal $P(x) = N(x, s^2 C_0)$ yields $\nu_0 \not\sim \nu_0^\top$ in infinite dimensions

- [Beskos et al., 2008]: pCN proposal

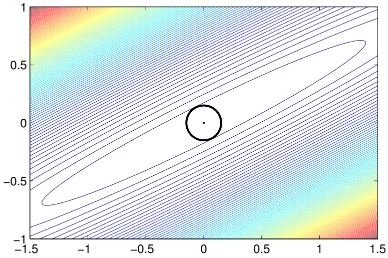
$$P(x) = N(\sqrt{1-s^2}x, s^2 C_0)$$

yields $\nu_0 = \nu_0^\top$ and, thus, $\quad \alpha(x,y) = \min\left\{1, \mathrm{e}^{\Phi(x)-\Phi(y)}\right\}$

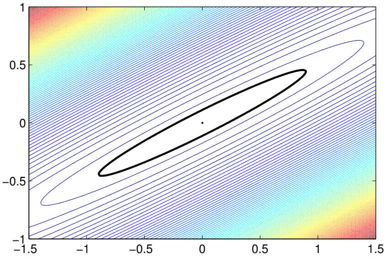
Motivation For Improvement

Observation: [Tierney, 1994], [Roberts & Rosenthal, 2001], . . .
Higher efficiency when proposal P uses posterior covariance matrix

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



$$P(x) = N(x, s^2 I)$$

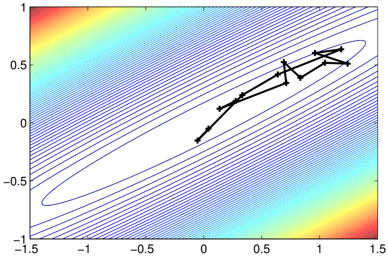


$$P(x) = N(x, s^2 C)$$

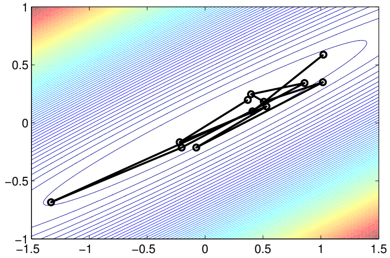
Motivation For Improvement

Observation: [Tierney, 1994], [Roberts & Rosenthal, 2001], ...
Higher efficiency when proposal P uses posterior covariance matrix

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



$P(x) = N(x, s^2 I)$

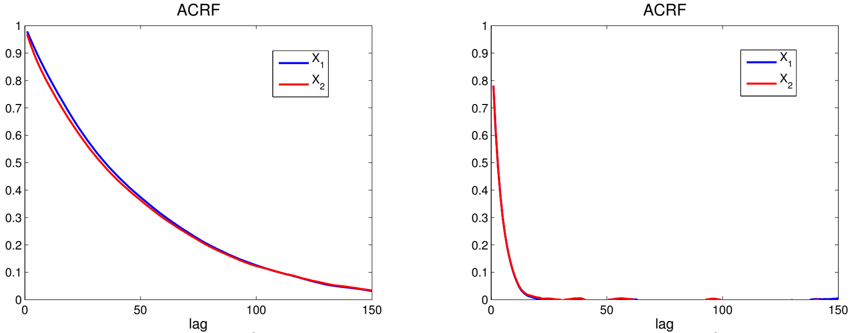


$P(x) = N(x, s^2 C)$

Motivation For Improvement

Observation: [Tierney, 1994], [Roberts & Rosenthal, 2001], . . .
Higher efficiency when proposal P uses posterior covariance matrix

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



$P(x) = N(x, s^2 I)$ $P(x) = N(x, s^2 C)$

How to approximate posterior covariance in advance

- If forward map G were linear, then

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

How to approximate posterior covariance in advance

- 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 $x_0 \in \mathscr{H}$

$$G(x) \approx \widetilde{G}(x) := G(x_0) + Jx, \qquad J = \nabla G(x_0)$$

  yields approximation to posterior covariance

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

How to approximate posterior covariance in advance

- 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 $x_0 \in \mathscr{H}$

$$G(x) \approx \widetilde{G}(x) := G(x_0) + Jx, \qquad J = \nabla G(x_0)$$

yields approximation to posterior covariance

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

- Possible choice for x_0:

$$x_{\mathrm{MAP}} = \operatorname*{argmin}_x |\mathbf{d} - G(x)|^2 + \|C_0^{-1/2} x\|^2$$

Generalized pCN-Proposal

- Class of proposal covariances:

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

- Generalized pCN-proposal:

$$P_\Gamma(x) = N(A_\Gamma x, s^2 C_\Gamma),$$

where enforcing $\nu_0 = \nu_0^\top$ yields

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

(cf. operator weighted proposals [Law, 2013] and [Cui et al., 2016])

Generalized pCN-Proposal

- Class of proposal covariances:

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

- Generalized pCN-proposal:

$$P_\Gamma(x) = N(A_\Gamma x, s^2 C_\Gamma),$$

where enforcing $\nu_0 = \nu_0^\top$ yields

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

(cf. operator weighted proposals [Law, 2013] and [Cui et al., 2016])

Lemma ([Rudolf, S., 2016])

There holds $A_\Gamma \in \mathscr{L}(\mathscr{H})$. The MH algorithm using the gpCN-proposal P_Γ is well-defined in Hilbert spaces and yields the μ-invariant gpCN-MH kernel K_Γ.

Numerical Experiment

- **Setting**
  - 1D model: $\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{e}^{u(t)}\ \frac{\mathrm{d}p}{\mathrm{d}t}(t)\right) = 0, \quad p(0) = 0,\ p(1) = 2$
  - Prior: $u(t, \boldsymbol{\xi}) \approx \sum_{m=1}^{M} \frac{\xi_m}{m\pi}\sqrt{2}\ \sin(m\pi t), \ \ \boldsymbol{\xi} \sim N(0, I)$
  - Observations: $y = \left[\, p(0.2j)\,\right]_{j=1}^{4} + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\epsilon^2 I)$
  - Quantity of interest: $f(\boldsymbol{\xi}) = \int_0^1 \mathrm{e}^{u(t,\boldsymbol{\xi})}\,\mathrm{d}t$
- **Proposals for MH-MCMC**
- **Results**

Numerical Experiment

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

Gaussian random walk: $P_1(x) = N(x, s^2 C_0)$

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

Gauss-Newton RW: $P_3(x) = N(x, s^2 C_\Gamma)$

gpCN: $P_4(x) = N(A_\Gamma x, s^2 C_\Gamma)$

$$\Gamma = \sigma_\epsilon^{-2} J^\top J, \quad J = \nabla G(x_{\mathrm{MAP}})$$

where $x_{\mathrm{MAP}} = \operatorname{argmin}_x \sigma_\epsilon^{-2} |\mathbf{d} - G(x)|^2 + \|C_0^{-1/2} x\|^2$

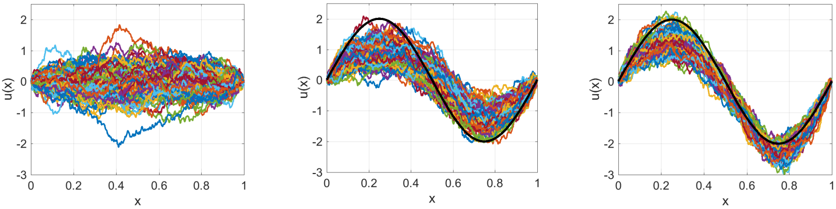
- **Results**

Numerical Experiment

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

100 prior and posterior realizations

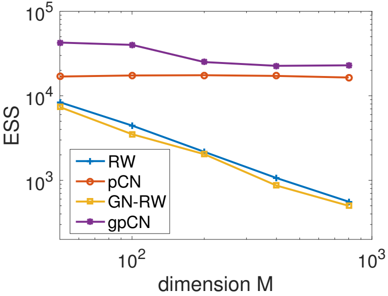
$\boldsymbol{\xi} \sim \mu_0$ | $\boldsymbol{\xi} \sim \mu,\ \sigma_\epsilon = 0.1$ | $\boldsymbol{\xi} \sim \mu,\ \sigma_\epsilon = 0.01$



Numerical Experiment

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

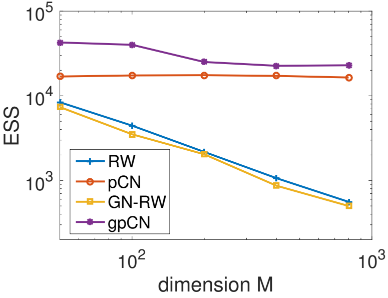
Effective sample size vs. dimension



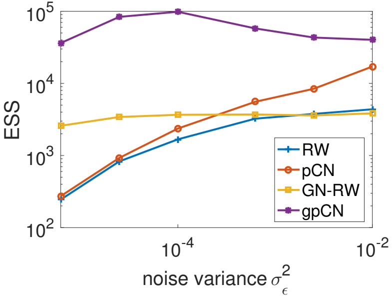
Numerical Experiment

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

Effective sample size vs. dimension



Effective sample size vs. noise variance



Geometric Ergodicity And Spectral Gaps

- MH kernel K is L^2_μ-geometrically ergodic if for an $r > 0$

$$\|\mu - \upsilon K^n\|_{\mathrm{TV}} \leq C_\upsilon \, \mathrm{e}^{-r\, n} \qquad \forall \upsilon : \frac{\mathrm{d}\upsilon}{\mathrm{d}\mu} \in L^2_\mu(\mathscr{H})$$

Geometric Ergodicity And Spectral Gaps

- MH kernel K is L^2_μ-geometrically ergodic if for an $r > 0$

$$\|\mu - \upsilon K^n\|_{\mathrm{TV}} \leq C_\upsilon \,\mathrm{e}^{-r\,n} \qquad \forall \upsilon : \frac{\mathrm{d}\upsilon}{\mathrm{d}\mu} \in L^2_\mu(\mathscr{H})$$

- Markov operator $\mathrm{K} : L^2_\mu(\mathscr{H}) \to L^2_\mu(\mathscr{H})$ associated with MH kernel K:

$$\mathrm{K}f(x) := \int_{\mathscr{H}} f(y)\, K(x, \mathrm{d}y),$$

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

- [Roberts & Rosenthal, 1997]: K is L^2_μ-geometrically ergodic iff $\mathrm{gap}_\mu(\mathrm{K}) > 0$ and, moreover, there holds

$$\lim_{n\to\infty} n\,\mathbb{E}\left[\left|S_n(f) - \mathbb{E}_\mu[f]\right|^2\right] \leq \frac{2\|f\|^2_{L^2_\mu}}{\mathrm{gap}_\mu(\mathrm{K})}, \qquad f \in L^2_\mu(\mathscr{H})$$

Proving Geometric Ergodicity of gpCN-MH Kernel

- For pCN-MH kernel K_0 an L^2_μ-spectral gap was proven in [Hairer et al., 2014] under certain conditions on Φ
- **Our Strategy:** a comparative approach by relating $\mathrm{gap}_\mu(\mathrm{K}_\Gamma)$ to $\mathrm{gap}_\mu(\mathrm{K}_0)$:

Theorem (Comparison of spectral gaps [Rudolf, S., 2016])

If

1. *the associated Markov operators K_0 and K_Γ are positive,*
2. *there exists the Radon-Nikodym derivative* $\rho_\Gamma(x,y) := \frac{\mathrm{d}P_0(x)}{\mathrm{d}P_\Gamma(x)}(y)$
3. *and for a $\beta > 1$ there holds*

$$\sup_{\mu(A)\in(0,\frac{1}{2}]} \frac{\int_A \int_{A^c} \rho_\Gamma^\beta(x,y)\, P_\Gamma(x,\mathrm{d}y)\, \mu(\mathrm{d}x)}{\mu(A)} < \infty,$$

then

$$\mathrm{gap}_\mu(\mathrm{K}_0)^{2\beta} \leq c_\beta\, \mathrm{gap}_\mu(\mathrm{K}_\Gamma)^{\beta-1}.$$

Convergence Result

- Assumptions 1 and 2 fulfilled for pCN- and gpCN-proposal
- To ensure assumption 3, we have to consider restriction of μ:

$$\mu_R(\mathrm{d}x) \propto \mathbf{1}_{B_R}(x)\,\mu(\mathrm{d}x), \qquad B_R := \{x \in \mathscr{H} : \|x\| < R\}$$

and restricted gpCN-MH kernel $K_{\Gamma,R}$ with $\alpha_R(x,y) := \mathbf{1}_{B_R}(y)\alpha(x,y)$

Convergence Result

- Assumptions 1 and 2 fulfilled for pCN- and gpCN-proposal
- To ensure assumption 3, we have to consider restriction of μ:

$$\mu_R(\mathrm{d}x) \propto \mathbf{1}_{B_R}(x)\,\mu(\mathrm{d}x), \qquad B_R := \{x \in \mathscr{H} : \|x\| < R\}$$

and restricted gpCN-MH kernel $K_{\Gamma,R}$ with $\alpha_R(x,y) := \mathbf{1}_{B_R}(y)\alpha(x,y)$

Theorem (Spectral gap of restricted gpCN-MH [Rudolf, S., 2016])

If

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

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

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

Variance-Independent Performance Of MH Algorithms

- Scaled observational noise $\varepsilon \sim N(0, \sigma^2\, \Sigma)$ yields family of posteriors

$$\mu_\sigma(\mathrm{d}x) \propto \exp\left(-\sigma^{-2}\, \Phi(x)\right)\, \mu_0(\mathrm{d}x), \qquad \sigma > 0$$

- Given μ_σ-invariant MH kernels K_σ, we can investigate if

$$\lim_{\sigma \to 0} \mathrm{gap}_{\mu_\sigma}(\mathrm{K}_\sigma) = \beta > 0$$

Variance-Independent Performance Of MH Algorithms

- Scaled observational noise $\varepsilon \sim N(0, \sigma^2\, \Sigma)$ yields family of posteriors

$$\mu_\sigma(\mathrm{d}x) \propto \exp\left(-\sigma^{-2}\, \Phi(x)\right)\, \mu_0(\mathrm{d}x), \qquad \sigma > 0$$

- Given μ_σ-invariant MH kernels K_σ, we can investigate if

$$\lim_{\sigma\to 0}\, \mathrm{gap}_{\mu_\sigma}(\mathrm{K}_\sigma) = \beta > 0$$

- Hard to analyze, thus, we examine limits for $\sigma \to 0$ of

Expected acceptance rate: $\mathbb{E}\left[\alpha_\sigma(\boldsymbol{\xi}_k, Y_k)\right],$

Expected squared jump size: $\mathbb{E}\left[\left|\boldsymbol{\xi}_k - \boldsymbol{\xi}_{k+1}\right|^2\right],$

where $\left(\boldsymbol{\xi}_k\right)_{k\in\mathbb{N}}$ Markov chain generated by K_σ starting at $\boldsymbol{\xi}_1 \sim \mu_\sigma$

A Result For Gaussian Posteriors

Theorem (Variance independence for Gaussian posterior [S., 2017])

Let $\mu_0 = N(0, C_0)$ *on* $\mathbb{R}^M$ *and* $G \colon \mathbb{R}^M \to \mathbb{R}^d$ *be linear with* $d < M$*, i.e.,*

$$\mu_\sigma = N(m_\sigma, C_\sigma), \qquad C_\sigma = (C_0^{-1} + \sigma^{-2} G^\top \Sigma^{-1} G)^{-1}.$$

Then for Markov chains $(\xi_k)_{k\in\mathbb{N}}$ *generated by MH algorithm with*

A Result For Gaussian Posteriors

Theorem (Variance independence for Gaussian posterior [S., 2017])

Let $\mu_0 = N(0, C_0)$ on $\mathbb{R}^M$ and $G\colon \mathbb{R}^M \to \mathbb{R}^d$ be linear with $d < M$, i.e.,

$$\mu_\sigma = N(m_\sigma, C_\sigma), \qquad C_\sigma = (C_0^{-1} + \sigma^{-2} G^\top \Sigma^{-1} G)^{-1}.$$

Then for Markov chains $(\xi_k)_{k\in\mathbb{N}}$ generated by MH algorithm with

- *RW-proposal* $P_\sigma(x) = N(x, s^2 C_\sigma)$,
- *gpCN-proposal* $P_{\Gamma_\sigma}(x) = N(A_{\Gamma_\sigma} x, s^2 C_\sigma)$

there holds

$$\lim_{\sigma\to 0} \mathbb{E}\left[\alpha_\sigma(\xi_k, Y_k)\right] = \beta > 0, \qquad \lim_{\sigma\to 0} \mathbb{E}\left[|\xi_{k+1} - \xi_k|^2\right] = \tilde\beta > 0,$$

with $\beta = \beta(M, s)$ for RW and $\beta = \beta(d, s)$ for gpCN.

Gaussian Approximation of Non-Gaussian Posterior

- Bernstein-von Mises Theorem states approximate Gaussianity of posteriors on $\mathbb{R}^M$ in the large data limit

- Common Gaussian approximation of posterior

$$\mu_\sigma(\mathrm{d}x) \propto \exp\left(-\sigma^{-2}\,\Phi(x)\right)\,\mu_0(\mathrm{d}x), \qquad \sigma \in \mathbb{N},$$

with $\Phi \in C^2(\mathbb{R}^M)$, is Laplace approximation $\tilde{\mu}_\sigma = N(x_\sigma, \tilde{C}_\sigma)$, where

$$x_\sigma := \operatorname*{argmin}_{x \in \mathbb{R}^M} \Phi(x) + \sigma^2 \|x\|^2_{C_0^{-1}}, \quad \tilde{C}_\sigma^{-1} := C_0^{-1} + \sigma^{-2} \nabla^2 \Phi(x_\sigma)$$

Theorem

If unique minimizer $x_\star := \operatorname{argmin}_{x \in \mathbb{R}^M} \Phi(x)$ *exists with* $\nabla^2 \Phi(x_\star) > 0$, $x_\sigma \to x_\star$ *as* $\sigma \to 0$, *and* $\Phi \in C^3(\mathbb{R}^M)$, *then*

$$\|\mu_\sigma - \tilde{\mu}_\sigma\|_{\mathrm{TV}} \in \mathcal{O}(\sigma).$$

Variance Robustness For Non-Gaussian Posteriors

- Theorem extendable to Hellinger norm, arbitrary priors, sequence of Φ_σ,...
- ...and underdetermined case $\Phi\colon \mathbb{R}^M \to \mathbb{R}^d$, $d < M$, if Φ acts only on linear subspace $\mathscr{M}$, $\Phi(x+m) = \Phi(x)$ for $m \in \mathscr{M}^\perp$, with $\dim(\mathscr{M}) \le d$
- **Claim:** Whenever there holds

$$\lim_{\sigma\to 0} \|\mu_\sigma - \tilde{\mu}_\sigma\|_{\mathrm{TV}} = 0,$$

the MH algorithm based on RW or gpCN proposal

$$P_\sigma(x) = N(x, s^2\tilde{C}_\sigma), \qquad P_{\Gamma_\sigma}(x) = N(A_{\Gamma_\sigma}x, s^2\tilde{C}_\sigma)$$

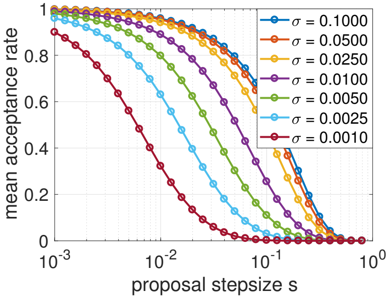
yields

$$\lim_{\sigma\to 0} \mathbb{E}\left[\alpha_\sigma(\boldsymbol{\xi}_k, Y_k)\right] = \beta > 0.$$

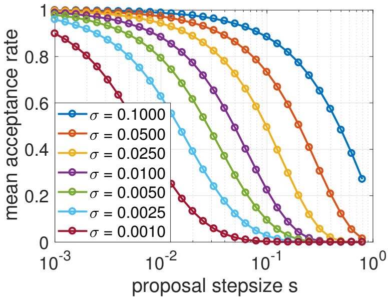
Numerical Experiment

- Linear forward map G (convolution operator) applied to unknown function
- Gaussian prior and noise $\varepsilon \sim N(0, \sigma^2\, I_4)$ yield Gaussian posterior

$$P(x) = N(x, s^2 C_0)$$



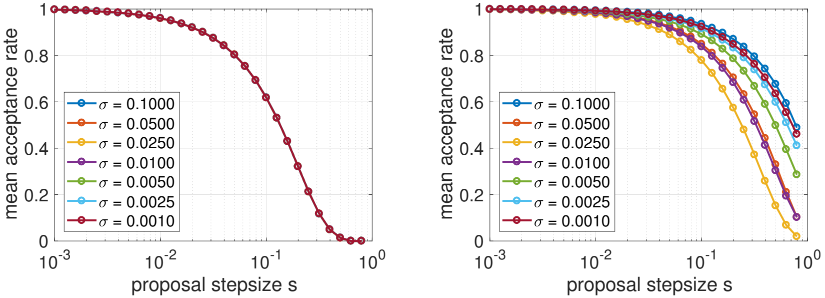
$$P(x) = N(\sqrt{1-s^2}x, s^2 C_0)$$



Numerical Experiment

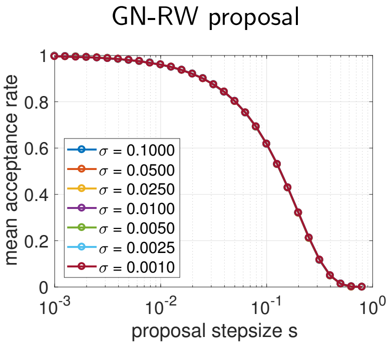
- Linear forward map G (convolution operator) applied to unknown function
- Gaussian prior and noise $\varepsilon \sim N(0, \sigma^2\ I_4)$ yield Gaussian posterior

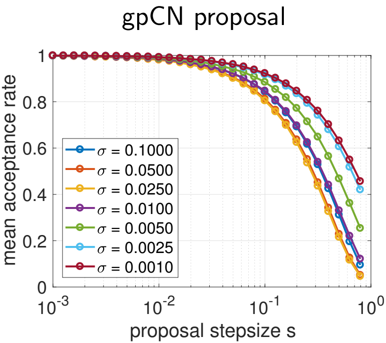
$P(x) = N(x, s^2 C_\sigma)$ $\qquad$ $P(x) = N(A_{\Gamma_\sigma} x, s^2 C_\sigma)$



Numerical Experiment cont'd

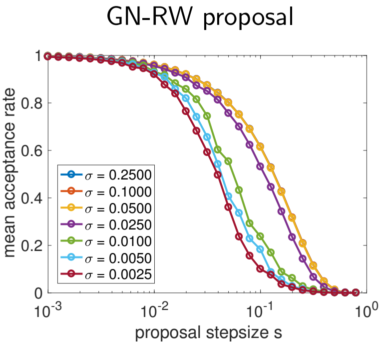
- Nonlinear forward G (exp $\circ$ convolution operator), $\dim(\mathscr{M}) = d$
- Gaussian prior and noise $\varepsilon \sim N(0, \sigma^2\ I_4)$, but **non**-Gaussian posterior
- Use covariance $\tilde{C}_\sigma$ of Laplace approximation for proposal

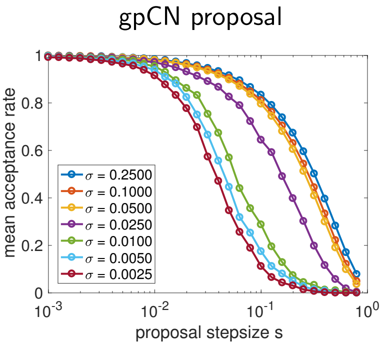




Numerical Experiment cont'd

- Nonlinear forward G (convolution operator $\circ$ exp), $\dim(\mathscr{M}) = M$
- Gaussian prior and noise $\varepsilon \sim N(0, \sigma^2\, I_4)$, but **non**-Gaussian posterior
- Use covariance $\tilde{C}_\sigma$ of Laplace approximation for proposal

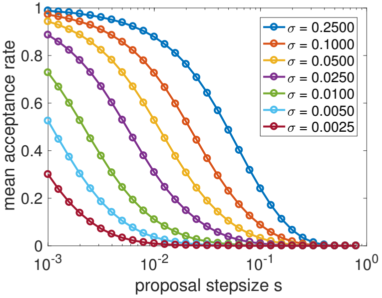




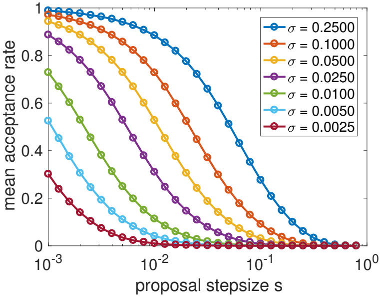
Numerical Experiment cont'd

- Nonlinear forward G (convolution operator $\circ \exp$), $\dim(\mathscr{M}) = M$
- Gaussian prior and noise $\varepsilon \sim N(0, \sigma^2 \, I_4)$, but **non**-Gaussian posterior
- Use covariance $\tilde{C}_\sigma$ of Laplace approximation for proposal

RW proposal



pCN proposal



Conclusions

- Bayesian inference for functions calls MH algorithms well-defined in infinite-dimensional spaces
- Existing MH algorithms with dimension-independent efficiency
- Introduced modification by incorporating approximate information about posterior covariance...
- ... which seems to perform dimension-independent & variance-robust
- Proved L^2_μ-geometric ergodicity of gpCN-MH algorithm via spectral gaps
- First steps to analyze variance-robustness of MH algorithms

References

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

T. Cui, K. Law, and Y. Marzouk. Dimension-independent likelihood-informed MCMC.
J. Comp. Physics, 304:109–137, 2016.

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.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines.
J. Chem. Phys., 21(6):1087–1092, 1953.

G. Roberts and J. Rosenthal. Geometric ergodicity and hybrid Markov chains.
Electron. Comm. Probab., 2(2):13–25, 1997.

D. Rudolf, B. Sprungk. On a generalization of the preconditioned Crank-Nicolson Metropolis algorithm.
Found. Comput. Math., 2016.

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

L. Tierney. Markov chains for exploring posterior distributions.
Ann. Stat., 22(4):1701–1762, 1994.