Explicit convergence bounds for preconditioned Crank–Nicolson

Andi Q. Wang

University of Warwick

Joint with: Christophe Andrieu, Anthony Lee, Sam Power.

BayesComp, Levi, Finland

March 2023

Overview

Overview

Quick overview

Focus of the talk: MCMC / sampling.

Quick overview

Focus of the talk: MCMC / sampling.

I will be discussing (new) bounds for the convergence to equilibrium of MCMC algorithms.

Quick overview

Focus of the talk: MCMC / sampling.

I will be discussing (new) bounds for the convergence to equilibrium of MCMC algorithms.

Thus the talk is mostly theoretical, but the purpose of this theory is to guide application!

Quick overview

Focus of the talk: MCMC / sampling.

I will be discussing (new) bounds for the convergence to equilibrium of MCMC algorithms.

Thus the talk is mostly theoretical, but the purpose of this theory is to guide application!

I will present bounds on the spectral gap of preconditioned Crank–Nicolson algorithm.

Quick overview

Focus of the talk: MCMC / sampling.

I will be discussing (new) bounds for the convergence to equilibrium of MCMC algorithms.

Thus the talk is mostly theoretical, but the purpose of this theory is to guide application!

I will present bounds on the spectral gap of preconditioned Crank–Nicolson algorithm.

Along the way I will introduce our refined technique for deriving convergence bounds based on isoperimetry and conductance.

Quick overview

Focus of the talk: MCMC / sampling.

I will be discussing (new) bounds for the convergence to equilibrium of MCMC algorithms.

Thus the talk is mostly theoretical, but the purpose of this theory is to guide application!

I will present bounds on the spectral gap of preconditioned Crank–Nicolson algorithm.

Along the way I will introduce our refined technique for deriving convergence bounds based on isoperimetry and conductance.

See also the keynote talk, Thursday 0920 of Anthony Lee, who will discuss applications of similar techniques to study Random Walk Metropolis.

Quick overview

Focus of the talk: MCMC / sampling.

I will be discussing (new) bounds for the convergence to equilibrium of MCMC algorithms.

Thus the talk is mostly theoretical, but the purpose of this theory is to guide application!

I will present bounds on the spectral gap of preconditioned Crank–Nicolson algorithm.

Along the way I will introduce our refined technique for deriving convergence bounds based on isoperimetry and conductance.

See also the keynote talk, Thursday 0920 of Anthony Lee, who will discuss applications of similar techniques to study Random Walk Metropolis.

See arXiv preprint! `https://arxiv.org/abs/2211.08959`.

Sampling

We wish to evaluate integrals

$$I = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x,$$

where π is a probability density function (our posterior distribution).

Sampling

We wish to evaluate integrals

$$I = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x,$$

where π is a probability density function (our posterior distribution).

Direct integration infeasible in high-dimensions (curse of dimensionality), furthermore only have access to π up to a normalizing constant!

Sampling

We wish to evaluate integrals

$$I = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x,$$

where π is a probability density function (our posterior distribution).

Direct integration infeasible in high-dimensions (curse of dimensionality), furthermore only have access to π up to a normalizing constant!

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$ and consider

$$I_n := \frac{1}{n}\sum_{i=1}^{n} f(X_i) \approx I = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x.$$

Monte Carlo

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$.

Monte Carlo

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$.

Exact sampling hard (e.g. rejection sampling also suffers from a curse of dimensionality)

Monte Carlo

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$.

Exact sampling hard (e.g. rejection sampling also suffers from a curse of dimensionality) so instead: build an ergodic Markov chain X which possesses π as its stationary distribution.

Monte Carlo

So instead, approximate I by sampling $X_1, X_2, \dots, X_n \sim \pi$.

Exact sampling hard (e.g. rejection sampling also suffers from a curse of dimensionality) so instead: build an ergodic Markov chain X which possesses π as its stationary distribution.

We simulate a π-reversible ergodic Markov chain,

$$X_1, X_2, \dots$$

where $X_n \to \pi$ in distribution and considering

$$I_n := \frac{1}{n} \sum_{i=1}^{n} f(X_i) \approx I = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x.$$

Metropolis–Hastings

Algorithm 1 Metropolis–Hastings (MH)

initialise: $X_0 = x_0, i = 0$

while $i < N$ **do**

  $i \leftarrow i + 1$

  simulate $Y_i \sim Q(X_{i-1}, \cdot)$

  $\alpha(X_{i-1}, Y_i) = 1 \wedge \frac{q(Y_i, X_{i-1})\pi(Y_i)}{q(X_{i-1}, Y_i)\pi(X_{i-1})}$

  with probability $\alpha(X_{i-1}, Y_i)$

    $X_i \leftarrow Y_i$

  else

    $X_i \leftarrow X_{i-1}$

return $(X_i)_{i=1,\ldots,n}$

Setting: Bayesian Inverse Problems

Target density

We assume that we are targeting a density on $\mathbb{R}^d$ of the form

$$\pi(\mathrm{d}x) \propto \mathcal{N}(\mathrm{d}x; 0, C) \cdot \exp(-\Psi(x)),$$

where Ψ is assumed convex, L-smooth and minimized at $x = 0$, and $\mathcal{N}$ denotes a Gaussian density, where C is a positive definite covariance matrix.

Setting: Bayesian Inverse Problems

Target density

We assume that we are targeting a density on $\mathbb{R}^d$ of the form

$$\pi(\mathrm{d}x) \propto \mathcal{N}(\mathrm{d}x; 0, C) \cdot \exp(-\Psi(x)),$$

where Ψ is assumed convex, L-smooth and minimized at $x = 0$, and $\mathcal{N}$ denotes a Gaussian density, where C is a positive definite covariance matrix.

Such densities arise naturally in Bayesian Inverse Problems, where C is a finite section of some infinite-dimensional trace-class covariance operator.

Setting: Bayesian Inverse Problems

Target density

We assume that we are targeting a density on $\mathbb{R}^d$ of the form

$$\pi(\mathrm{d}x) \propto \mathcal{N}(\mathrm{d}x; 0, C) \cdot \exp(-\Psi(x)),$$

where Ψ is assumed convex, L-smooth and minimized at $x = 0$, and $\mathcal{N}$ denotes a Gaussian density, where C is a positive definite covariance matrix.

Such densities arise naturally in Bayesian Inverse Problems, where C is a finite section of some infinite-dimensional trace-class covariance operator.

In this case $\nu(\mathrm{d}x) := \mathcal{N}(\mathrm{d}x; 0, C)$ is the prior and Ψ is the (log-)likelihood term corresponding to the observed data.

Preconditioned Crank–Nicolson

$$\pi(\mathrm{d}x) \propto \mathcal{N}(\mathrm{d}x; 0, C) \cdot \exp(-\Psi(x)), \quad \nu(\mathrm{d}x) := \mathcal{N}(\mathrm{d}x; 0, C).$$

Preconditioned Crank–Nicolson

$$\pi(\mathrm{d}x) \propto \mathcal{N}(\mathrm{d}x; 0, C) \cdot \exp(-\Psi(x)), \quad \nu(\mathrm{d}x) := \mathcal{N}(\mathrm{d}x; 0, C).$$

The preconditioned Crank–Nicolson algorithm (pCN) [Beskos et. al. (2008), Stuart (2010), Example 5.3] is a Metropolis–Hastings chain with ν-reversible Gaussian proposal Q: for fixed $\rho \in (0, 1)$,

$$Q(x, A) = \int \mathbf{1}_A(\rho x + \eta z)\, \nu(\mathrm{d}z),$$

where $\rho^2 + \eta^2 = 1$.

Preconditioned Crank–Nicolson

$$\pi(\mathrm{d}x) \propto \mathcal{N}(\mathrm{d}x; 0, C) \cdot \exp(-\Psi(x)), \quad \nu(\mathrm{d}x) := \mathcal{N}(\mathrm{d}x; 0, C).$$

The preconditioned Crank–Nicolson algorithm (pCN) [Beskos et. al. (2008), Stuart (2010), Example 5.3] is a Metropolis–Hastings chain with ν-reversible Gaussian proposal Q: for fixed $\rho \in (0, 1)$,

$$Q(x, A) = \int \mathbf{1}_A(\rho x + \eta z)\, \nu(\mathrm{d}z),$$

where $\rho^2 + \eta^2 = 1$.

pCN remains stable even in infinite dimensions, since the proposal preserves the (prior) measure ν, unlike pure Random Walk Metropolis (see keynote talk of Anthony Lee!).

Tuning of pCN

In general, tuning the 'step-size' η is important for good performance:

Tuning of pCN

In general, tuning the ‘step-size’ η is important for good performance:

η^2 too large $\Rightarrow$ many proposals rejected; wasted computational effort.

Tuning of pCN

In general, tuning the 'step-size' η is important for good performance:

η^2 too large $\Rightarrow$ many proposals rejected; wasted computational effort.

η^2 too small $\Rightarrow$ proposing tiny moves; wasted computational effort.

Tuning of pCN

In general, tuning the 'step-size' η is important for good performance:

η^2 too large $\Rightarrow$ many proposals rejected; wasted computational effort.

η^2 too small $\Rightarrow$ proposing tiny moves; wasted computational effort.

We will be interested to derive non-asymptotic bounds on the resulting spectral gap, which can be applied for a given target and given step-size. (C.f. optimal scaling framework of [Roberts, Gelman, Gilks (1997)].)

Spectral gap

Recall that a reversible π-invariant Markov kernel P defines an operator on $L^2(\pi)$, and its convergence to equilibrium can be bounded by the spectral gap γ (and this is the best rate):

Spectral gap

Recall that a reversible π-invariant Markov kernel P defines an operator on $L^2(\pi)$, and its convergence to equilibrium can be bounded by the spectral gap γ (and this is the best rate):

$$\|P^n f - \pi(f)\|_2 \leq (1-\gamma)^n \|f\|_2.$$

Spectral gap

Recall that a reversible π-invariant Markov kernel P defines an operator on $L^2(\pi)$, and its convergence to equilibrium can be bounded by the spectral gap γ (and this is the best rate):

$$\|P^n f - \pi(f)\|_2 \leq (1-\gamma)^n \|f\|_2.$$

Want an explicit bound; see related work of [Hairer, Stuart, Vollmer (2014)].

Main result

Recall we assumed potential Ψ is convex, L-smooth.

Theorem ([Andrieu, Lee, Power, W. (2022)])

Under our previous assumptions on π, setting $\eta = \varsigma \cdot (L \cdot \mathrm{Tr}(C))^{-1/2}$, we have the following bound on the spectral gap:

$$\gamma \geq 2^{-9} \cdot C_{\mathrm{g}}^2 \cdot \exp(-2\varsigma^2) \cdot \varsigma^2 \cdot (L \cdot \mathrm{Tr}(C))^{-1}.$$

Optimizing over ς, we obtain

$$\gamma \geq 3.62784 \times 10^{-5} \cdot (L \cdot \mathrm{Tr}(C))^{-1}.$$

Main result

Recall we assumed potential Ψ is convex, L-smooth.

Theorem ([Andrieu, Lee, Power, W. (2022)])

Under our previous assumptions on π, setting $\eta = \varsigma \cdot (L \cdot \mathrm{Tr}(C))^{-1/2}$, we have the following bound on the spectral gap:

$$\gamma \geq 2^{-9} \cdot C_{\mathrm{g}}^2 \cdot \exp(-2\varsigma^2) \cdot \varsigma^2 \cdot (L \cdot \mathrm{Tr}(C))^{-1}.$$

Optimizing over ς, we obtain

$$\gamma \geq 3.62784 \times 10^{-5} \cdot (L \cdot \mathrm{Tr}(C))^{-1}.$$

This is an explicit lower bound, which only depends on the dimension through $\mathrm{Tr}(C)$.

Overview

Conductance

Definition: Conductance

The conductance profile of a π-invariant Markov kernel P is

$$\Phi_P(v) := \inf \left\{ \frac{(\pi \otimes P)(A \times A^{\complement})}{\pi(A)} : \pi(A) \leq v \right\}, \quad v \in (0, 1/2].$$

The conductance of P is $\Phi_P^* := \Phi_P(1/2)$.

Conductance

Definition: Conductance

The conductance profile of a π-invariant Markov kernel P is

$$\Phi_P(v) := \inf \left\{ \frac{(\pi \otimes P)(A \times A^{\complement})}{\pi(A)} : \pi(A) \le v \right\}, \quad v \in (0, 1/2].$$

The conductance of P is $\Phi_P^* := \Phi_P(1/2)$.

Theorem (Cheeger inequalities)

For a positive chain, such as pCN, we have the bounds on the spectral gap,

$$\frac{1}{2} \cdot [\Phi_P^*]^2 \le \gamma \le \Phi_P^*.$$

Conductance

Definition: Conductance

The conductance profile of a π-invariant Markov kernel P is

$$\Phi_P(v) := \inf\left\{\frac{(\pi \otimes P)(A \times A^{\complement})}{\pi(A)} : \pi(A) \leq v\right\}, \quad v \in (0, 1/2].$$

The conductance of P is $\Phi_P^* := \Phi_P(1/2)$.

Theorem (Cheeger inequalities)

For a positive chain, such as pCN, we have the bounds on the spectral gap,

$$\frac{1}{2} \cdot [\Phi_P^*]^2 \leq \gamma \leq \Phi_P^*.$$

Thus our goal is to lower bound the conductance.

Isoperimetry

Fix target density π on metric space (E, d).

Definition: isoperimetric profile / minorant, c.f. [Milman (2009)]

Given a measurable set A, define the r-enlargment of A via $A_r := \{x \in \mathrm{E} : \mathrm{d}(x, A) \leq r\}$, and set

$$\pi^+(A) := \liminf_{r \downarrow 0} \frac{\pi(A_r) - \pi(A)}{r}.$$

Then the isoperimetric profile of π is

$$I_\pi(p) := \inf\{\pi^+(A) : A \in \mathcal{E}, \pi(A) = p\}, \quad p \in (0, 1).$$

Isoperimetry

Fix target density π on metric space (E, d).

Definition: isoperimetric profile / minorant, c.f. [Milman (2009)]

Given a measurable set A, define the r-enlargment of A via $A_r := \{x \in \mathsf{E} : \mathsf{d}(x, A) \leq r\}$, and set

$$\pi^+(A) := \liminf_{r \downarrow 0} \frac{\pi(A_r) - \pi(A)}{r}.$$

Then the isoperimetric profile of π is

$$I_\pi(p) := \inf\{\pi^+(A) : A \in \mathcal{E}, \pi(A) = p\}, \quad p \in (0,1).$$

A function $\tilde{I}_\pi : (0,1) \to (0,\infty)$ is a regular isoperimetric minorant of π if $\tilde{I}_\pi$ is continuous, monotone increasing, symmetric about $1/2$ and $\tilde{I}_\pi \leq I_\pi$.

Close coupling

Definition: close coupling

Given $\epsilon, \delta > 0$, we say that a Markov kernel P is $(\mathsf{d}, \delta, \epsilon)$-close coupling if

$$\mathsf{d}(x,y) \le \delta \Rightarrow \|P(x,\cdot) - P(y,\cdot)\|_{\mathrm{TV}} \le 1 - \epsilon, \quad \forall x, y \in \mathsf{E}.$$

Close coupling

Definition: close coupling

Given $\epsilon, \delta > 0$, we say that a Markov kernel P is $(\mathsf{d}, \delta, \epsilon)$-close coupling if

$$\mathsf{d}(x, y) \leq \delta \Rightarrow \|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \leq 1 - \epsilon, \quad \forall x, y \in \mathsf{E}.$$

Lemma: close coupling for Metropolis chains

For Metropolis chains, we have the bound:

$$\|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \leq \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} + 1 - \alpha_0,$$

$$\alpha_0 := \inf_{x \in \mathsf{E}} \alpha(x), \quad \alpha(x) := \int \alpha(x, y) Q(x, \mathrm{d}y).$$

Close coupling

Definition: close coupling

Given $\epsilon, \delta > 0$, we say that a Markov kernel P is $(\mathrm{d}, \delta, \epsilon)$-close coupling if

$$\mathrm{d}(x, y) \leq \delta \Rightarrow \|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \leq 1 - \epsilon, \quad \forall x, y \in \mathrm{E}.$$

Lemma: close coupling for Metropolis chains

For Metropolis chains, we have the bound:

$$\|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \leq \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} + 1 - \alpha_0,$$

$$\alpha_0 := \inf_{x \in \mathrm{E}} \alpha(x), \quad \alpha(x) := \int \alpha(x, y) Q(x, \mathrm{d}y).$$

Thus we can establish P is close coupling provided we can bound α_0!

Close coupling, conductance and isoperimetry

Theorem: Conductance lower bound; c.f. [Dwivedi et. al. (2019)]

Suppose $\tilde{I}_\pi$ is a regular, concave isoperimetric minorant of π. Let P be $(\mathrm{d}, \delta, \epsilon)$-close coupling. Then for any $v \in (0, 1/2]$,

$$\Phi_P(v) \geq \frac{1}{4} \cdot \epsilon \cdot 1 \wedge \left(\frac{\delta}{2} \cdot \frac{\tilde{I}_\pi(v/2)}{v/2} \right).$$

Close coupling, conductance and isoperimetry

Theorem: Conductance lower bound; c.f. [Dwivedi et. al. (2019)]

Suppose $\tilde{I}_\pi$ is a regular, concave isoperimetric minorant of π. Let P be $(\mathrm{d}, \delta, \epsilon)$-close coupling. Then for any $v \in (0, 1/2]$,

$$\Phi_P(v) \geq \frac{1}{4} \cdot \epsilon \cdot 1 \wedge \left(\frac{\delta}{2} \cdot \frac{\tilde{I}_\pi(v/2)}{v/2} \right).$$

Taking $v = 1/2$ immediately gives a lower bound on the conductance Φ_P^*, and hence on the spectral gap.

Close coupling, conductance and isoperimetry

Theorem: Conductance lower bound; c.f. [Dwivedi et. al. (2019)]

Suppose $\tilde{I}_\pi$ is a regular, concave isoperimetric minorant of π. Let P be $(\mathrm{d}, \delta, \epsilon)$-close coupling. Then for any $v \in (0, 1/2]$,

$$\Phi_P(v) \geq \frac{1}{4} \cdot \epsilon \cdot 1 \wedge \left(\frac{\delta}{2} \cdot \frac{\tilde{I}_\pi(v/2)}{v/2} \right).$$

Taking $v = 1/2$ immediately gives a lower bound on the conductance Φ_P^*, and hence on the spectral gap.

This result thus breaks the problem into two pieces:

- For a given target π, establish a regular concave isoperimetric minorant $\tilde{I}_\pi$.
- For the chain P, establish close coupling.

Overview

Isoperimetric minorants for π

There are various ways to establish isoperimetric minorants: for example, they can be derived from functional inequalities, e.g. Poincaré inequalities, log-Sobolev inequalities, c.f. [Bobkov (1999)].

Isoperimetric minorants for π

There are various ways to establish isoperimetric minorants: for example, they can be derived from functional inequalities, e.g. Poincaré inequalities, log-Sobolev inequalities, c.f. [Bobkov (1999)].

The specific case of interest for this talk:

Lemma

Under our assumptions on π, we have minorant

$$I_\pi(p) \geq \varphi(\Phi^{-1}(p)) =: \tilde{I}_\pi(p),$$

with respect to metric $d = |\cdot|_C^{-1}$, where φ, Φ are the standard Gaussian p.d.f. and c.d.f., and furthermore

$$\tilde{I}_\pi(1/4) = C_{\mathrm{g}},$$

where $C_{\mathrm{g}} \geq 0.317776$.

Controlling acceptance probabilities

We saw that to establish close coupling, needed to get a handle on $\alpha_0 := \inf_{x \in \mathsf{E}} \alpha(x)$.

Controlling acceptance probabilities

We saw that to establish close coupling, needed to get a handle on $\alpha_0 := \inf_{x \in E} \alpha(x)$.

Through a direct calculation, we obtain:

Lemma

Let $\eta = \varsigma \cdot (L \cdot \mathrm{Tr}(C))^{-1/2}$*, some* $\varsigma > 0$*. Then*

$$\alpha_0 \geq \frac{1}{2} \cdot \exp\left(-\frac{\varsigma^2}{2}\right).$$

Main result

Putting together all of these pieces, we obtain the main result.

Theorem

We obtain the lower bound on the spectral gap of pCN, for $\eta = \varsigma \cdot (L \cdot \mathrm{Tr}(C))^{-1/2}$

$$\gamma \geq 2^{-9} \cdot C_{\mathrm{g}}^2 \cdot \varsigma^2 \cdot \exp(-2\varsigma^2) \cdot (L \cdot \mathrm{Tr}(C))^{-1}.$$

Main result

Putting together all of these pieces, we obtain the main result.

Theorem

We obtain the lower bound on the spectral gap of pCN, for $\eta = \varsigma \cdot (L \cdot \mathrm{Tr}(C))^{-1/2}$

$$\gamma \geq 2^{-9} \cdot C_{\mathrm{g}}^2 \cdot \varsigma^2 \cdot \exp(-2\varsigma^2) \cdot (L \cdot \mathrm{Tr}(C))^{-1}.$$

In this convex, L-smooth case, we have a nice isoperimetric minorant; but can be applied in other cases too.

Using the full conductance profile can get much more intricate analysis of the mixing times.

Overview

Concluding remarks

I have presented explicit lower bound on the spectral gap of the pCN algorithm, focussing on the convex, L-smooth case.

Concluding remarks

I have presented explicit lower bound on the spectral gap of the pCN algorithm, focussing on the convex, L-smooth case.

However the general framework developed is applicable much more broadly!

Concluding remarks

I have presented explicit lower bound on the spectral gap of the pCN algorithm, focussing on the convex, L-smooth case.

However the general framework developed is applicable much more broadly!

Furthermore the full conductance profile can give much more detailed mixing time bounds (not presented today; see paper).

Concluding remarks

I have presented explicit lower bound on the spectral gap of the pCN algorithm, focussing on the convex, L-smooth case.

However the general framework developed is applicable much more broadly!

Furthermore the full conductance profile can give much more detailed mixing time bounds (not presented today; see paper).

Our paper actually discusses Random Walk Metropolis as the main example – see keynote talk Thursday 0920 of Anthony Lee!

Concluding remarks

I have presented explicit lower bound on the spectral gap of the pCN algorithm, focussing on the convex, L-smooth case.

However the general framework developed is applicable much more broadly!

Furthermore the full conductance profile can give much more detailed mixing time bounds (not presented today; see paper).

Our paper actually discusses Random Walk Metropolis as the main example – see keynote talk Thursday 0920 of Anthony Lee!

Natural next steps would be to consider more advanced algorithms such as MALA, HMC, etc...

Thanks for listening! I

- Andrieu, C., Lee, A., Power, S., Wang, A. Q. (2022). Poincaré inequalities for Markov chains: a meeting with Cheeger, Lyapunov and Metropolis. *Technical report*. https://doi.org/10.48550/arxiv.2208.05239.
- **Andrieu, C., Lee, A., Power, S., Wang, A. Q. (2022)**. Explicit convergence bounds for Metropolis Markov chains: isoperimetry, spectral gaps and profiles. https://doi.org/10.48550/arxiv.2211.08959.
- Belloni, A., Chernozhukov, V. (2009). On the computational complexity of MCMC-based estimators in large samples. *Ann. Statist.*, 37(4), 20112055.
- Beskos, A., Roberts, G., Stuart, A., Voss, J. (2008). MCMC methods for diffusion bridges. *Stoch. Dyn.*, 8(3), 319350.
- Bobkov, S. G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.*, 27(4), 19031921.
- Chen, Y., Dwivedi, R., Wainwright, M. J., Yu, B. (2019). Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients. *J. Mach. Learn. Res.*, 21.
- Doucet, A., Pitt, M. K., Deligiannidis, G., Kohn, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika*, 102(2), 295313.
- Dwivedi, R., Chen, Y., Wainwright, M. J., Yu, B. (2019). Log-concave sampling: Metropolis-Hastings algorithms are fast. *J. Mach. Learn. Res.*, 20, 142.
- Goel, S., Montenegro, R., Tetali, P. (2006). Mixing time bounds via the spectral profile. *Elec. J. Probab.*, 11(2000), 126.

Thanks for listening! II

Hairer, M., Stuart, A. M., Vollmer, S. J. (2014). Spectral gaps for a metropolis-hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6), 24552490.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21(6), 10871092.

Milman, E. (2009). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.*, 177(1), 143.

Roberts, G. O., Gelman, A., Gilks, W. R. (1997). Weak Convergence and Optimal Scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1), 110120.

Stuart, A. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, 19, 451559.

Convergence framework

Recall a reversible chain P is positive if for any $f \in \mathrm{L}^2(\pi)$,

$$\langle Pf, f \rangle \geq 0.$$

Convergence framework

Recall a reversible chain P is positive if for any $f \in \mathrm{L}^2(\pi)$,

$$\langle Pf, f \rangle \geq 0.$$

Lemma ([Doucet et. al. (2015)])

pCN with Gaussian proposals is a positive chain.

Close coupling for pCN

Prevously: provided we can choose δ such that $|x - y| \leq \delta \Rightarrow \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq \alpha_0/2$, we obtain that P is close coupling with $\epsilon \geq \alpha_0/2$.

Close coupling for pCN

Prevously: provided we can choose δ such that $|x-y| \le \delta \Rightarrow \|Q(x,\cdot) - Q(y,\cdot)\|_{\mathrm{TV}} \le \alpha_0/2$, we obtain that P is close coupling with $\epsilon \ge \alpha_0/2$.

Since we have Gaussian $\mathcal{N}(0, \sigma^2 \mathbf{I}_d)$ proposals, we can use Pinsker's inequality to obtain

Lemma

For $v > 0$,

$$|x-y|_{C^{-1}} \le v \cdot \frac{\eta}{\rho} \Rightarrow \|Q(x,\cdot) - Q(y,\cdot)\|_{\mathrm{TV}} \le v/2.$$

Close coupling for pCN

Prevously: provided we can choose δ such that $|x-y| \le \delta \Rightarrow \|Q(x,\cdot)-Q(y,\cdot)\|_{\mathrm{TV}} \le \alpha_0/2$, we obtain that P is close coupling with $\epsilon \ge \alpha_0/2$.

Since we have Gaussian $\mathcal{N}(0, \sigma^2 \mathbf{I}_d)$ proposals, we can use Pinsker's inequality to obtain

Lemma

For $v > 0$,

$$|x-y|_{C^{-1}} \le v \cdot \frac{\eta}{\rho} \Rightarrow \|Q(x,\cdot)-Q(y,\cdot)\|_{\mathrm{TV}} \le v/2.$$

Thus by taking $v = \alpha_0$, i.e. $\delta = \alpha_0 \cdot \eta/\rho$, we have that P is close coupling with $\epsilon = \alpha_0/2$.

Close coupling for pCN

Prevously: provided we can choose δ such that $|x - y| \leq \delta \Rightarrow \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq \alpha_0/2$, we obtain that P is close coupling with $\epsilon \geq \alpha_0/2$.

Since we have Gaussian $\mathcal{N}(0, \sigma^2 \mathbf{I}_d)$ proposals, we can use Pinsker's inequality to obtain

Lemma

For $v > 0$,

$$|x - y|_{C^{-1}} \leq v \cdot \frac{\eta}{\rho} \Rightarrow \|Q(x, \cdot) - Q(y, \cdot)\|_{\mathrm{TV}} \leq v/2.$$

Thus by taking $v = \alpha_0$, i.e. $\delta = \alpha_0 \cdot \eta/\rho$, we have that P is close coupling with $\epsilon = \alpha_0/2$.

So all that remains is to get a handle on α_0.

Convergence of MCMC

What is the criteria for an MCMC chain to be ‘good’?

Convergence of MCMC

What is the criteria for an MCMC chain to be 'good'?

Classically, MCMC is good if it converges fast to equilibrium and mixes well.

Convergence of MCMC

What is the criteria for an MCMC chain to be 'good'?

Classically, MCMC is good if it converges fast to equilibrium and mixes well.

One measure of the former is to look at rates of convergence:

Convergence of MCMC

What is the criteria for an MCMC chain to be 'good'?

Classically, MCMC is good if it converges fast to equilibrium and mixes well.

One measure of the former is to look at rates of convergence:

Theorem ([?, ?])

RWM converges to equilibrium exponentially fast if and only if π has an exponential moment (e.g. $\pi(x) \propto \exp(-\|x-\mu\|^{\alpha})$, $\alpha \geq 1$.). Otherwise, the chain converges at a subgeometric (e.g. polynomial) rate.*

L^2 convergence and Dirichlet forms

We work on $\mathrm{L}^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_2^2 < \infty\}, \quad \langle f, g\rangle := \int fg \, \mathrm{d}\pi$,
$\mathrm{L}^2_0(\pi) := \{f \in \mathrm{L}^2(\pi) : \pi(f) = 0\}$.

L^2 convergence and Dirichlet forms

We work on $\mathrm{L}^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_2^2 < \infty\}$, $\quad \langle f, g \rangle := \int fg \, \mathrm{d}\pi$,
$\mathrm{L}_0^2(\pi) := \{f \in \mathrm{L}^2(\pi) : \pi(f) = 0\}$.

For a π-invariant Markov transition kernel P with $\mathrm{L}^2(\pi)$-adjoint P^*, define the Dirichlet form $\mathcal{E}(P^*P, f)$, for $f \in \mathrm{L}_0^2(\pi)$:

$$\mathcal{E}(P^*P, f) := \langle (I - P^*P)f, f \rangle = \|f\|^2 - \|Pf\|^2.$$

L^2 convergence and Dirichlet forms

We work on $\mathrm{L}^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_2^2 < \infty\}, \quad \langle f, g\rangle := \int fg \, \mathrm{d}\pi$,
$\mathrm{L}_0^2(\pi) := \{f \in \mathrm{L}^2(\pi) : \pi(f) = 0\}$.

For a π-invariant Markov transition kernel P with $\mathrm{L}^2(\pi)$-adjoint P^*, define the Dirichlet form $\mathcal{E}(P^*P, f)$, for $f \in \mathrm{L}_0^2(\pi)$:

$$\mathcal{E}(P^*P, f) := \langle (I - P^*P)f, f\rangle = \|f\|^2 - \|Pf\|^2.$$

This acts like a discrete derivative, and we will seek to lower bound it.

L^2 convergence and Dirichlet forms

We work on $\mathrm{L}^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_2^2 < \infty\}$, $\quad \langle f, g \rangle := \int fg \,\mathrm{d}\pi$,
$\mathrm{L}_0^2(\pi) := \{f \in \mathrm{L}^2(\pi) : \pi(f) = 0\}$.

For a π-invariant Markov transition kernel P with $\mathrm{L}^2(\pi)$-adjoint P^*, define the Dirichlet form $\mathcal{E}(P^*P, f)$, for $f \in \mathrm{L}_0^2(\pi)$:

$$\mathcal{E}(P^*P, f) := \langle (I - P^*P)f, f \rangle = \|f\|^2 - \|Pf\|^2.$$

This acts like a discrete derivative, and we will seek to lower bound it.

Furthermore if P is reversible and positive (so its spectrum $\sigma(P) \subset [0, 1]$), we have that

$$\mathcal{E}(P^*P, f) = \mathcal{E}(P^2, f) \geq \mathcal{E}(P, f).$$

L^2 convergence and Dirichlet forms

We work on $\mathrm{L}^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_2^2 < \infty\}$, $\quad \langle f, g \rangle := \int fg \, \mathrm{d}\pi$,
$\mathrm{L}_0^2(\pi) := \{f \in \mathrm{L}^2(\pi) : \pi(f) = 0\}$.

For a π-invariant Markov transition kernel P with $\mathrm{L}^2(\pi)$-adjoint P^*, define the Dirichlet form $\mathcal{E}(P^*P, f)$, for $f \in \mathrm{L}_0^2(\pi)$:

$$\mathcal{E}(P^*P, f) := \langle (I - P^*P)f, f \rangle = \|f\|^2 - \|Pf\|^2.$$

This acts like a discrete derivative, and we will seek to lower bound it.

Furthermore if P is reversible and positive (so its spectrum $\sigma(P) \subset [0, 1]$), we have that

$$\mathcal{E}(P^*P, f) = \mathcal{E}(P^2, f) \geq \mathcal{E}(P, f).$$

So it will be sufficient to lower bound $\mathcal{E}(P, f)$.

Conductance and spectral profiles

Lemma ([Goel et. al. (2006)])

For nonconstant nonnegative $g \in \mathrm{L}_0^2(\pi)$, *we have the lower bound*

$$\mathcal{E}(P, g) \geq \mathrm{Var}_\pi(g) \cdot \frac{1}{2} \cdot \Lambda_P \left(\frac{4[\pi(g)]^2}{\mathrm{Var}_\pi(g)} \right),$$

where Λ_P *is the spectral profile of* P.

Lemma

For π*-reversible* P, *we have the further lower bound*

$$\Lambda_P(v) \geq \begin{cases} \frac{1}{2}\Phi_P(v)^2 & 0 < v \leq 1/2, \\ \frac{1}{2}[\Phi_P^*]^2 & v > 1/2. \end{cases}$$