Explicit convergence bounds for Metropolis Markov chains

Isoperimetry, Spectral Gaps, and Complexity

Sam Power

University of Bristol

18 January, 2023

Links & Acknowledgements

- Main paper today: arXiv 2211.08959;
- Related: arXiv 2208.05239
- All joint work with
  - Christophe Andrieu (Bristol)
  - Anthony Lee (Bristol)
  - Andi Q. Wang (Bristol $\rightsquigarrow$ Warwick)

- Funded by Bayes4Health EPSRC Grant

My Interests

- Computational aspects of statistical inference
- Stochastic algorithms for questions from statistics, machine learning, ...
- Theoretical properties of algorithms (efficiency, complexity, comparisons, ...)
- Motivated by the task of understanding structured probability distributions in high-dimensional spaces
  - posterior inference in Bayesian statistics
  - latent variable models, hidden Markov models
  - generative modeling
  - non-convex optimisation
  - ...

My Goals

- Understanding structured probability distributions in high-dimensional spaces:
  - Developing and adapting mathematical tools for analysis of practical algorithms
  - Application of this understanding towards guiding practice
- Today:
  - A specific class of methods for this task,
  - A specific algorithm within this class, and
  - A mathematical analysis of this algorithm.

Markov Chain Monte Carlo (MCMC)

- Task: Generate approximate samples from a probability distribution π to which we have *limited access*.
- MCMC: An iterative approach to this task.
  - Simulate a time-homogeneous Markov chain $(X_n)_{n \geqslant 0}$ such that

$$\mathrm{Law}(X_n) \to \pi \text{ as } n \to \infty.$$

    (and hopefully, quickly)
- Current status:
  - Mature algorithmic field, many 'correct' solutions are known and practical.
  - Quantitative convergence theory is *challenging*; *important*.
    - 'Is (this algorithm) $\{$ performant, reliable, preferable, ... $\}$?'
    - 'Given π, which algorithm do I choose?'

Random Walk Metropolis

- Today: Study the *Random Walk Metropolis* (RWM) algorithm
  - Only requires access to density of π, up to a multiplicative constant (typical).
  - Widely-used, simple, 'representative'

1. At x,
   1.1 Propose $x' \sim \mathcal{N}\left(x, \sigma^2 \cdot I_d\right)$.
   1.2 Evaluate $r\left(x, x'\right) = \frac{\pi(x')}{\pi(x)}$.
   1.3 With probability $\min\left\{1, r\left(x, x'\right)\right\}$, move to x'; otherwise, remain at x.

- Leaves π invariant, ergodic under mild conditions, exponentially so under tail conditions

Quantitative Convergence of RWM

- Despite ubiquity, sharp complexity analysis of RWM has long been open
- We obtain a convincing complexity analysis with
  - sharp dependence on the dimension of the problem
  - conjecturally sharp dependence on the conditioning of the problem
- Our proof techniques are remarkably robust, and largely new to this area
- Gives a relatively complete resolution to the question of RWM's mixing

Convergence of Markov Chains

- For nice f, define $Pf(x) = \mathbb{E}[f(X_1) \mid X_0 = x]$, where $X_1 \sim P(X_0 \to \cdot)$.
  - $\rightsquigarrow P^n f(x) = \mathbb{E}[f(X_n) \mid X_0 = x]$
- If the Markov chain converges in law to π, then

$$\forall x, \quad \lim_{n\to\infty} P^n f(x) = \mathbb{E}_\pi [f(X)] .$$

- 'Convergence in L^2': for $f \in \mathrm{L}^2(\pi)$ with $\mathbb{E}_\pi [f(X)] = 0$, have

$$\|P^n f\|_2^2 := \int \pi(x) \left(P^n f(x)\right)^2 \, \mathrm{d}x \to 0.$$

- Nice to work with, implies other common notions of convergence

L^2 Convergence of Markov Chains

- If chain is exponentially ergodic and reversible, then there hold *uniform* bounds of the form

$$f \in L^2(\pi) \implies \|P^n f\|_2 \leqslant (1 - \gamma_P)^n \cdot \|f\|_2,$$

  where $\gamma_P > 0$ is the 'spectral gap' of the chain.
- Estimates on γ_P give control of mixing time, variance of MCMC estimators, etc.
  - $\rightsquigarrow$ *practically relevant*.
- First goal: characterise γ_P for RWM.
- Further goals: more detailed questions about convergence to equilibrium.

Conductance Methods for Markov Chains

- **Many** tools exist for studying convergence of Markov chains.
- *Conductance* analysis is well-suited to study of chains making ‘local’ moves.
  - Facilitated by recent progress in *isoperimetry* of probability measures.
- Consider for $A \subseteq \mathbb{R}^d$

$$\pi(A) := \int_{x \in A} \pi(x)\,\mathrm{d}x$$

$$\pi \otimes P(A \times A^{\complement}) := \int_{x \in A, y \in A^{\complement}} \pi(x) P(x, y)\,\mathrm{d}x\mathrm{d}y.$$

- If $\pi \otimes P(A \times A^{\complement}) \geqslant c \cdot \pi(A)$, then $\mathrm{P}(X_1 \notin A \mid X_0 \in A) \geqslant c$,
  - so if $c \gg 0$, then the set A is easy for P to escape.
- If every set A is easy for P to escape, then P cannot get stuck ...
  - ... and hence must converge quickly.

Cheeger's Inequality for Markov Chains

- Define the 'conductance' of P as

$$\Phi_P := \inf\left\{\frac{\pi\otimes P(A\times A^{\complement})}{\pi(A)} : \pi(A)\leqslant\frac{1}{2}\right\}.$$

  Then for (positive, reversible) P, it holds that $\gamma_P \geqslant \frac{1}{2}\Phi_P^2$.
- For $P = P^{\mathrm{RWM}}$, we will lower bound Φ_P, and hence γ_P.
- Remark: For $\pi(A)\approx 0$, things can be much better; gives sharper description of convergence when far from equilibrium.

Bounding Conductance for 'Local' Markov Chains

- The following is true for general Markov chains on metric spaces:
  - Suppose that
    1. ('close coupling') For some $(\delta, \tau) \in \mathbb{R}_+ \times (0, 1)$, it holds that
    $$\mathsf{d}(x, y) \leqslant \delta \implies \mathrm{TV}\left(P_x, P_y\right) \leqslant 1 - \tau.$$
    2. ('good isoperimetry') For some $\Phi_\pi > 0$, the target measure π satisfies
    $$\pi^+(A) \geqslant \Phi_\pi \cdot \pi(A),$$
    where π^+ is the 'Minkowski content' ($\approx$ boundary mass) of A.
    - Interpretation: under 'natural, local dynamics' on π, all sets are easy to escape.
  - Then, it holds that
  $$\Phi_P \gtrsim \tau \cdot \delta \cdot \Phi_\pi.$$
- P is 'nice' at small scales $+$ π is 'nice' at large scales $\rightsquigarrow$ good mixing!

Conductance for RWM

- We argue as follows:
  1. To guarantee that ‘close coupling’ holds, it suffices to control

$$\alpha_0 := \text{inf}\left\{\mathrm{P}\left(\text{ accept proposed move } \mid \text{current state } = x\right) : x \in \mathrm{R}^d\right\}$$
$$= \text{‘worst-case acceptance rate out of a state’}$$

  2. For ‘well-concentrated’ targets, Φ_π can be controlled explicitly in cases of interest.
- Failure of these conditions corresponds to known failure modes for RWM.

Application to Log-Concave Targets

- Write $U = -\log \pi$, assume that $0 < m \leqslant U'' \leqslant L$ (in matrix sense).
- Consider RWM with $\sigma \asymp (L \cdot d)^{-1/2}$. Then
  1. P satisfies 'close coupling' with $(\delta, \tau) \asymp \left((L \cdot d)^{-1/2}, 1 \right)$.
  2. π has 'good isoperimetry', with $\Phi_\pi \gtrsim m^{1/2}$.
- It thus follows that

$$
\begin{aligned}
\Phi_P &\gtrsim \left(\frac{m}{L \cdot d} \right)^{1/2} \\
\gamma_P &\gtrsim \frac{m}{L \cdot d}.
\end{aligned}
$$

- Interpretation: 'difficulty' of sampling from π scales as $\approx \kappa \cdot d$, where $\kappa = \frac{L}{m}$.

Other Results

- γ_P implies an estimate for the asymptotic variance
- γ_P implies an estimate for the relaxation time
- More detailed analysis gives a good estimate of the mixing time
- Can handle U with lower regularity (e.g. U' Hölder).
- Can handle π with tails 'from exponential to Gaussian'.
- Can treat related 'Metropolis-type' algorithms (pCN).

Future Work

- Analysis of RWM on Heavy-Tailed π (where $\Phi_\pi = 0$)
- Analysis of other practical samplers ($\{$ Langevin, Hamiltonian $\}$ Monte Carlo)
- Development of new samplers inspired by insights suggested by our proof techniques

Recap

- Markov chain analysis for computational statistics
- Sharp analysis of Random Walk Metropolis algorithm
- New theoretical approach centered on isoperimetry
- Proof tools should generalise well to other ‘local’ algorithms