Spectral Telescope: Convergence Rate Bounds for Random-Scan Gibbs Samplers Based on a Hierarchical Structure

Qian Qin Guanyang Wang

Unviersity of Minnesota and Rutgers University

Contents

- Introduction to Markov chain Monte Carlo and Gibbs samplers
- Convergence rates of MCMC algorithms
- Recent developments in computer science concerning convergence analyses of Gibbs samplers
- A hierarchical structure of Gibbs samplers and the spectral telescope
- Illustration

Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a class of algorithms for simulating complex distributions.

- Target Distribution: Π.
- The algorithm simulates a Markov chain $X(1), X(2), \ldots$ such that $\mathcal{L}(X(t)) \to \Pi$ in some sense as $t \to \infty$.
- It is important to understand the convergence rate of a given algorithm.

Random-Scan Gibbs Sampler

The random-scan Gibbs sampler is a commonly- used and studied class of MCMC algorithms.

- The target distribution $\Pi = \mathcal{L}(X_1, \ldots, X_n)$.
- In each step of a standard Gibbs sampler, when the current state is $(x_1, \ldots, x_n)$, select 1 coordinate $j \in \{1, \ldots, n\}$, and update the value of x_j using $\mathcal{L}(X_j \mid X_{-\{j\}} = x_{-\{j\}})$.

Random-Scan Gibbs Sampler (Glauber Dynamics)

Example:

$$\pi(x_1, \ldots, x_n) \propto \begin{cases} 1 & \sum_{i=1}^{n} x_i < 1, \\ 0 & \text{elsewhere} \end{cases},$$

where $x_i \in (0, 1)$.

$$\pi(x_j \mid x_{-\{j\}}) \propto \begin{cases} 1 & x_j < 1 - \sum_{i \neq j} x_i, \\ 0 & \text{elsewhere}, \end{cases}$$

where $x_j \in (0, 1)$.

Random-Scan Gibbs Sampler

Example:

Let (V, E) be a graph. Each vertex is associated with a binary random variable whose range is $\{-1, 1\}$.

$$\pi(x_1, \ldots, x_{|V|}) \propto \exp\left(-c \sum_{\{i,j\} \in E} x_i x_j\right),$$

where $x_i \in \{-1, 1\}$.

$$\pi(x_j \mid x_{-\{j\}}) \propto \exp\left(-c \sum_{i:\ \{i,j\} \in E} x_i x_j\right).$$

Convergence Rates of Reversible Chains

Random-scan Gibbs samplers simulate Markov chains reversible with respect to their target distributions. Under mild conditions, this implies that the associated Markov chain

$$X(t) = (X_1(t), \ldots, X_n(t)), \quad t \geq 1,$$

satisfies $\mathcal{L}(X(t)) \to \Pi$ in some sense.

Rate of convergence?

Convergence Rates of Reversible Chains

- Measure the difference between $\mathcal{L}(X(t))$ and Π through the L^2 distance:

$$\|\mathcal{L}(X(t)) - \Pi\| = \sup_f |E[f(X(t))] - E_\Pi[f(X)]|,$$

  where $\mathrm{var}_\Pi[f(X)] = 1$.

- The chain is said to be geometrically convergent if, there is a constant $\rho < 1$, a function C, such that, for a broad class of initialization $\mathcal{L}(X(0))$,

$$\|\mathcal{L}(X(t)) - \Pi\| < C[\mathcal{L}(X(0))]\rho^t \quad \forall t.$$

- The smallest $\rho \in [0, 1]$ for which the above holds is the chain's **convergence rate**.
- Smaller rate = faster convergence.

Spectral Gap

- For reversible chains, it is common to study the spectral gap:

  spectral gap $= 1 -$ convergence rate $\in [0, 1]$.

- Larger gap $=$ faster convergence.

- Bounding the spectral gap from below tells us how fast a chain converges.

Spectral Gap Bounds for Gibbs Samplers

- Recently, in the theoretical computer science literature, a new technique for bounding the spectral gap of Glauber dynamics emerged.
- The technique: **spectral independence**.
- Initially introduced in Anari et al. (2021) to establish a polynomial mixing time of the Gibbs sampler for hardcore models.
- It has since received a tremendous amount of attention in computer science as it provides a powerful tool for proving fast, and sometimes optimal, mixing time bounds for Gibbs samplers for several important discrete models.

Spectral Independence

- Target probability mass function: $\pi(x_1, \ldots, x_n)$, where $x_i \in \{1, \ldots, q\}$.
  This is the joint distribution of the random variables $X_1, \ldots, X_n$.
- To construct bounds using the spectral independence technique, one needs information on the joint distribution of (X_i, X_j), $i \neq j$, given any subset of other components, e.g.,
  Conditional distribution of (X_i, X_j) given all other components $X_{-\{i,j\}}$;
  Marginal distribution of (X_i, X_j);
  Conditional distribution of (X_i, X_j) given a subset of other components.

Spectral Independence

- The information about the joint distribution of (X_i, X_j) given some subset of other components is summarized in an **influence matrix**.
- Roughly, the ijth component of an influence matrix characterizes how much the value of X_i affect the conditional distribution of X_j given X_i and some subset of other components.
- The spectral gap can be bounded using the features (e.g., spectral radii) of influence matrices.

Spectral Independence

- Attractive alternative to Dobrushin's uniqueness condition.
- It has been applied to Glauber dynamics (Gibbs samplers) for spin systems and coloring systems.
- Previously, it was developed only for discrete models, where the target distribution lives on a finite state space.

A Hierarchical Structure

Gibbs sampler targeting $\mathcal{L}(X_1, X_2, X_3, X_4)$, current state (x_1, x_2, x_3, x_4):

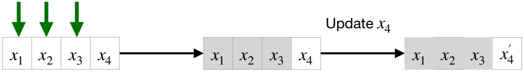
1. Randomly select $j \in \{1, 2, 3, 4\}$.
2. Update x_j based on $\mathcal{L}(X_j \mid X_{-\{j\}} = x_{-\{j\}})$.

Hierarchical representation:

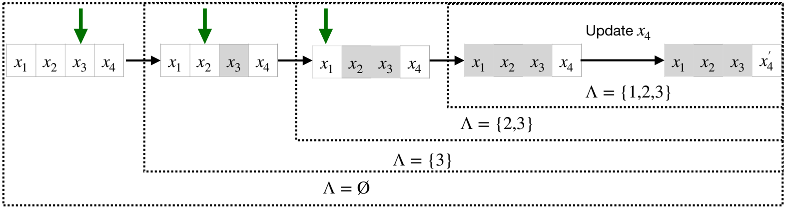
1. Randomly select $i_1 \in \{1, 2, 3, 4\}$.
2. Randomly select $i_2 \in -\{i_1\} := \{1, 2, 3, 4\} \setminus \{i_1\}$.
3. Randomly select $i_3 \in -\{i_1, i_2\}$.
4. Let $j = i_4 \in -\{i_1, i_2, i_3\}$ be the remaining index. Update x_{i_4} based on $\mathcal{L}(X_{i_4} \mid X_{-\{i_4\}} = x_{-\{i_4\}})$.

A Hierarchical Structure

Standard Gibbs sampler



Recursive Gibbs sampler



A Hierarchical Structure

A Gibbs sampler targeting a joint distribution with m components: Randomly selects $m-1$ components to fix, and updates the remaining component.

Gibbs sampler targeting $\mathcal{L}(X_1, X_2, X_3, X_4)$, current state (x_1, x_2, x_3, x_4):

1. Randomly select $i_1 \in \{1, 2, 3, 4\}$.
2. Randomly select $i_2 \in -\{i_1\} := \{1, 2, 3, 4\} \setminus \{i_1\}$.
3. Randomly select $i_3 \in -\{i_1, i_2\}$.
4. Let $i_4 \in -\{i_1, i_2, i_3\}$ be the remaining index. Update x_{i_4} based on $\mathcal{L}(X_{i_4} \mid X_{-\{i_4\}} = x_{-\{i_4\}})$.

Steps 2-4 make up one iteration of a Gibbs sampler targeting $\mathcal{L}(X_{-\{i_1\}} \mid X_{i_1} = x_{i_1})$.

A Hierarchical Structure

A Gibbs sampler targeting a joint distribution with m components: Randomly selects $m-1$ components to fix, and updates the remaining component.

Gibbs sampler targeting $\mathcal{L}(X_1, X_2, X_3, X_4)$, current state (x_1, x_2, x_3, x_4):

1. Randomly select $i_1 \in \{1, 2, 3, 4\}$. Say, $i_1 = 3$.
2. Randomly select $i_2 \in -\{i_1\} = \{1, 2, 4\}$.
3. Randomly select $i_3 \in -\{3, i_2\} = \{1, 2, 4\} \setminus \{i_2\}$.
4. Let $i_4 \in -\{3, i_2, i_3\} = \{1, 2, 4\} \setminus \{i_2, i_3\}$ be the remaining index. Update x_{i_4} based on $\mathcal{L}(X_{i_4} \mid X_{-\{i_4\}} = x_{-\{i_4\}})$.

Steps 2-4 make up one iteration of a Gibbs sampler targeting $\mathcal{L}(X_1, X_2, X_4 \mid X_3 = x_3)$.

A Hierarchical Structure

Gibbs sampler targeting $\mathcal{L}(X_1, X_2, X_3, X_4)$, current state (x_1, x_2, x_3, x_4):

1. Randomly select $i_1 \in \{1, 2, 3, 4\}$.
2. Call one iteration of the Gibbs sampler targeting $\mathcal{L}(X_{-\{i_1\}} \mid X_{i_1} = x_{i_1})$, current state $x_{-\{i_1\}}$.

A Hierarchical Structure

Gibbs sampler targeting $\mathcal{L}(X_{-\{i_1\}} \mid X_{i_1} = x_{i_1})$, current state $x_{-\{i_1\}}$:

1. Randomly select $i_2 \in -\{i_1\}$.
2. Call one iteration of the Gibbs sampler targeting $\mathcal{L}(X_{-\{i_1,i_2\}} \mid X_{\{i_1,i_2\}} = x_{\{i_1,i_2\}})$, current state $x_{-\{i_1,i_2\}}$.

The structure connects Gibbs samplers targeting higher dimensional distributions to those targeting lower dimensional ones.

A Hierarchical Structure

- Full target $\mathcal{L}(X_1, \ldots, X_n)$: n components.

  Gibbs$(n, 1)$: selects 1 component from n components to update.

- Any conditional distribution with m components, $m \leq n$.

  Gibbs$(m, 1)$: selects 1 component from m components to update.

- The hierarchical structure connects Gibbs$(m, 1)$ and Gibbs$(m-1, 1)$ samplers for $m \in \{2, \ldots, n\}$.

Spectral Telescope

Using the hierarchical structure, we can find a connection between the spectral gaps of Gibbs$(m, 1)$ and Gibbs$(m-1, 1)$ samplers.

- Let Gap$(m, 1)$ and Gap$(m-1, 1)$ be respectively, the smallest spectral gaps of Gibbs$(m, 1)$ and Gibbs$(m-1, 1)$ samplers.
- In particular, Gap$(n, 1)$ is the spectral gap of the Gibbs sampler targeting the full joint distribution.
- Main result: Gap$(m, 1) \geq$ Gap$(m, m-1)$Gap$(m-1, 1)$.

Spectral Telescope

- $\mathrm{Gap}(m, 1) \geq \mathrm{Gap}(m, m-1)\mathrm{Gap}(m-1, 1)$.
- $\mathrm{Gap}(m, m-1)$ gives the smallest spectral gap of $\mathrm{Gibbs}(m, m-1)$ samplers, which updates $m-1$ components out of m components in each iteration.
- $\mathrm{Gibbs}(n, n-1)$ targeting $\mathcal{L}(X_1, \ldots, X_n)$:
    1. Randomly select $n-1$ indices $j_1, \ldots, j_{n-1} \in \{1, \ldots, n\}$.
    2. Update $X_{j_1}, \ldots, X_{j_{n-1}}$ based on $\mathcal{L}(X_{j_1}, \ldots, X_{j_{n-1}} \mid X_i = x_i)$, where i is the remaining index.
- $\mathrm{Gibbs}(m, m-1)$ targets an m-component conditional distribution, and updates $m-1$ components at a time.

Spectral Telescope

- $\text{Gap}(m, 1) \geq \text{Gap}(m, m-1)\text{Gap}(m-1, 1)$.
- Spectral telescope (Qin and Wang, 2022):

$$\text{Gap}(n, 1) \geq \prod_{m=2}^{n} \text{Gap}(m, m-1).$$

Similar properties were found for related systems (Carlen et al., 2003).

Gap($m, m-1$)

This quantity is related to several interesting features of the target distribution, including the influence matrices utilized by the spectral independence technique.

- Consider an m-component conditional distribution ω, e.g., the conditional distribution of $(X_1, \ldots, X_m)$ given $(X_{m+1}, \ldots, X_n) = (x_{m+1}, \ldots, x_n)$.
- A Gibbs($m, m-1$) sampler targeting this distribution selects $m-1$ components from $(x_1, \ldots, x_m)$ to update.
- Its spectral gap is related to...

Gap$(m, m-1)$

- ... The spectral gap $\geq 1 - S(\omega)$ where $S(\omega)$ is a correlation coefficient for the target distribution ω that describes its **dependence structure**.
- For $Y_1, \ldots, Y_m \sim \omega$,

$$S(\omega) = \sup_{f_1,\ldots,f_m} \frac{\operatorname{var}\left[\sum_{i=1}^m f_i(Y_i)\right]}{m\sum_{i=1}^m \operatorname{var}[f_i(Y_i)]}.$$

- If $Y_1, \ldots, Y_m$ are weakly dependent on each other, $S(\omega)$ is small, and the spectral gap is large (fast convergence).
- Upper bounds on $S(\omega)$ give a lower bound on Gap$(m, m-1)$.

$\mathrm{Gap}(m, m-1) \sim 1 - S(\omega)$

- $S(\omega)$ is further related to the convergence rate of a certain random walk.
- The state of the random walk has the form (j, y), where $j \in \{1, \ldots, m\}$, and $y \in \mathrm{Range}(Y_j)$. $(Y_1, \ldots, Y_m \sim \omega.)$
- If the current state is (j, y), this random walk proceeds as follows:
  1. Randomly select $j' \in \{1, \ldots, m\}$.
  2. Draw y' based on $\mathcal{L}(Y_{j'} \mid Y_j = y)$.
  3. Update (j, y) to (j', y').
- The spectral gap of this chain $G(\omega)$ is precisely $1 - S(\omega)$.
- Lower bounds on $G(\omega)$ give a lower bound on $\mathrm{Gap}(m, m-1)$.
- This extends results in Alev and Lau (2020) who studied random walks on pure simplical complexes, a discrete structure frequently studied in computer science.

$\mathrm{Gap}(m, m-1) \sim 1 - S(\omega) \sim G(\omega)$

- $G(\omega)$ is in turn related to the spectral radius of an influential matrix.
- The influence matrix is an $m \times m$ matrix whose (i, j)th element describes how much $\mathcal{L}(Y_j \mid Y_i = y)$ changes as y varies. $(Y_1, \ldots, Y_m \sim \omega.)$
- Smaller spectral radius implies larger $G(\omega)$.
- Upper bounds on the spectral radii of influence matrices give a lower bound on $\mathrm{Gap}(m, m-1)$ — Spectral independence on general state spaces.

Summary

- A hierarchical structure of Gibbs samplers yields a hierarchical structure of their spectral gaps.
- It draws a connection between a Gibbs sampler targeting a joint distribution and Gibbs samplers targeting conditional distributions.
- We can related the spectral gap of a Gibbs sampler to
  1. dependence structure of the target distribution
  2. convergence rates of some random walk chains
  3. some influence matrices (spectral independence)
- These relations could be use to bound the spectral gap, i.e., show how fast a Gibbs chain converges.

Illustration

$$\pi(x_1, \ldots, x_n) \propto \begin{cases} 1 & \sum_{i=1}^{n} x_i < 1, \\ 0 & \text{elsewhere} \end{cases},$$

where $x_i \in (0, 1)$.

$$\pi(x_j \mid x_{-\{j\}}) \propto \begin{cases} 1 & x_j < 1 - \sum_{i \neq j} x_i, \\ 0 & \text{elsewhere}, \end{cases}$$

where $x_j \in (0, 1)$.

Illustration

By studying the dependence structure of the target distribution and its conditionals using orthogonal polynomials, we are able to show that

$$\mathrm{Gap}(n,1) \geq \frac{5}{36} \prod_{m=4}^{n} \frac{m-3}{m-2} = \frac{5}{36(n-2)}.$$

When $n \to \infty$, this bound gives the correct order $1/n$.

Reference

ALEV, V. L. and LAU, L. C. (2020). Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*.

ANARI, N., LIU, K. and GHARAN, S. O. (2021). Spectral independence in high-dimensional expanders and applications to the hardcore model. *SIAM Journal on Computing* FOCS20–1.

CARLEN, E. A., CARVALHO, M. C. and LOSS, M. (2003). Determination of the spectral gap for Kac's master equation and related stochastic evolution. *Acta mathematica* **191** 1–54.

QIN, Q. and WANG, G. (2022). Spectral telescope: Convergence rate bounds for random-scan Gibbs samplers based on a hierarchical structure. arXiv preprint.