Fast Mixing for Discrete Point Processes

Patrick Rebeschini[1] **and Amin Karbasi**[2]

[1]patrick.rebeschini@yale.edu [2]amin.karbasi@yale.edu
Yale Institute for Network Science, Yale University

Yale Institute for Network Science

1. Role of submodularity in probability?

- **Combinatorial optimization:** Submodularity extensively studied.

Let V be a finite set and $f : S \in 2^V \longrightarrow f(S) \in \mathbb{R}$ be a set function.

$$\Delta_i f(S) := f(S \cup \{i\}) - f(S) \qquad \textbf{(gradient of } f\textbf{)}$$

Function f is **submodular** if for each $i, j \notin S$, $i \neq j$

$$\Delta_i \Delta_j f(S) \equiv \Delta_j f(S \cup \{i\}) - \Delta_j f(S) \le 0 \qquad \textbf{(Hessian of } f\textbf{)}$$

- **Probability:** Submodularity recently investigated to compute $\mathbb{P}(\mathbf{S} \ni i)$ in

$$\mathbb{P}(\mathbf{S} = S) := \frac{e^{-\beta f(S)}}{Z}, \qquad \beta > 0.$$

ISSUE: (Djolonga and Krause, 2014) **bounds exp. bad in** $|V| := \mathrm{card}\, V$.

Q: Can we get dimension-free bounds?

Q: Is submodularity right notion?

2. Fast mixing MCMC: control on Hessian

GOAL: Investigate **fundamental** property of f to get fast mixing MCMC.

Consider local-update (Glauber dynamics type) Markov chains:
(systematic-scan, Metropolis-Hasting algorithm also considered in the paper)

Algorithm 1. Random-scan Gibbs sampler

Sample $\mathbf{S}_0 \in 2^V$ from a given distribution (e.g., uniform); Set $S \leftarrow \mathbf{S}_0$;
for $s = 1, \ldots, t$ **do**
  for $|V|$ *times* **do**
    Sample $i \in V$ uniformly. Draw $C \in \{0,1\}$ with $\mathbb{P}(C=0) = \frac{1}{1+\exp \Delta_i f(S \setminus \{i\})}$;
    If $C = 0$ then set $S \leftarrow S \setminus \{i\}$, else set $S \leftarrow S \cup \{i\}$;
  $\mathbf{S}_s \leftarrow S$;
Output: Markov chain $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_t$.

MAIN RESULT: For a **generic** set function f, if

$$\beta \|M\|_\infty \equiv \beta \max_{i \in V} \sum_{j \in V} M_{ij} \le \gamma < 1 \qquad \text{where} \qquad M_{ij} \propto \max_{S \in 2^V : S \not\ni i,j} |\Delta_i \Delta_j f(S)|$$

then $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_t$ is fast mixing (mixing time $\tau(\varepsilon) \le \left\lceil \frac{\log(|V|\varepsilon^{-1})}{1-\gamma} \right\rceil$) and

$$\left\| \frac{1}{N} \sum_{k=1}^{N} \mathbf{1}(\mathbf{S}_t^{[k]} \ni i) - \mathbb{P}(\mathbf{S} \ni i) \right\|_2 \le \gamma^t + \frac{1}{\sqrt{N}},$$

where $\mathbf{S}^{[1]}, \ldots, \mathbf{S}^{[N]}$ are N independent copies of the Markov chain.

- Proof relies on theory of Dobrushin uniqueness for Gibbs measures.
- **Key result:** Bound does not depend on dimension $|V|$.
- **Key property: Dimension-free uniform control on Hessian.**
- **Submodularity not enough**: Phase transition as a function of β for convergence rate of Glauber dynamics for Ising model.

NOTE: No previous literature on Hessian of set functions.

3. Hessian and decay of correlation

Hessian captures decay of correlations in probability.

Examples in metric space (d is metric):

- Exponential decay of correlations: $\max_{S \in 2^V : S \not\ni i,j} |\Delta_i \Delta_j f(S)| \le \alpha e^{-\alpha' d(i,j)}$.
- Finite-range correlations: $\max_{S \in 2^V : S \not\ni i,j} |\Delta_i \Delta_j f(S)| \le \begin{cases} c & \text{if } d(i,j) \le r, \\ 0 & \text{if } d(i,j) > r. \end{cases}$

4. Hessian and *curvature*

- Many results in submodular optimization for *monotone* functions (i.e., $\Delta_i f(S) \ge 0$ for each i, S) rely on notion of *curvature* (based on **gradient**):

$$c := 1 - \min_{i \in V} \frac{\min_{S \in 2^V : S \not\ni i} \Delta_i f(S)}{f(\{i\})} = 1 - \min_{i \in V} \frac{\Delta_i f(V \setminus \{i\})}{f(\{i\})} \in [0,1].$$

We have $c = 0$ if and only if function is *modular*, i.e., $f(S) = \sum_{i \in S} w_i$.
Curvature is convenient as easy to compute (minimum is over $|V|$ terms).

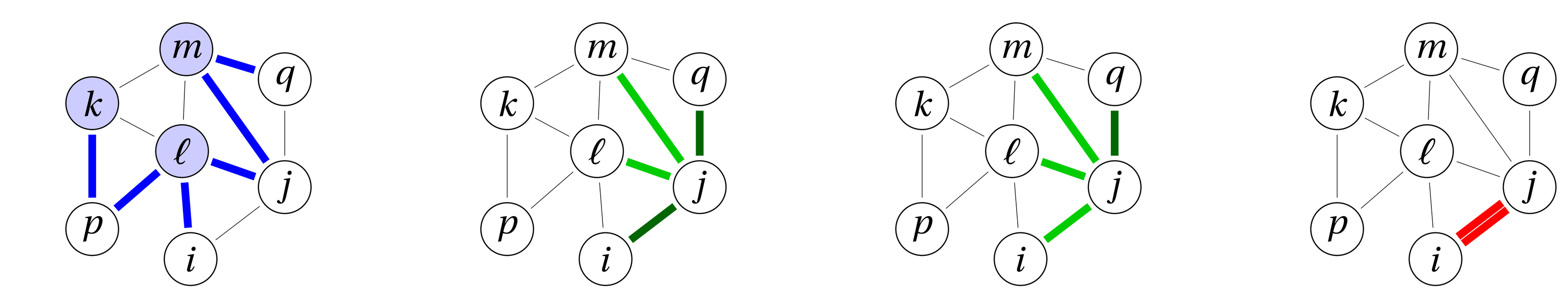
- **Hessian** is a **more natural concept to characterize "curvature"**.
- **Hessian** also **captures locality**.

$$\begin{pmatrix} -1 & -c & -c \\ -c & \ddots & -c \\ -c & -c & -1 \end{pmatrix} \le \frac{\Delta_i \Delta_j f(S)}{f(\{i\}) \wedge f(\{j\})} \le \begin{pmatrix} c-1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & c-1 \end{pmatrix}$$

- In general $\max_{S \in 2^V : S \not\ni i,j} |\Delta_i \Delta_j f(S)|$ is not easy to compute (maximum is over $2^{|V|-2}$ terms). However:
  - In many canonical applications (cut function, coverage function, etc.) Hessian is **sparse** and can be **easily computed or uniformly bounded**.
  - In other applications (e.g., determinantal point processes) more assumptions are needed to uniformly control Hessian.

5. Cut function

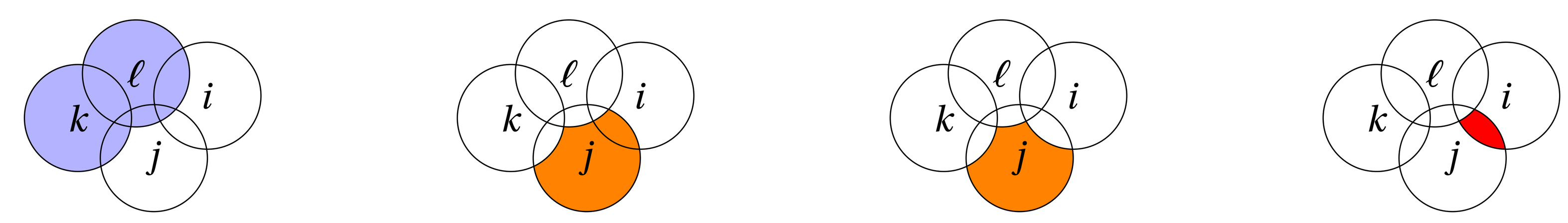
- (V, E) complete graph. $L_{ij} = L_{ji} \ge 0$ weight associated to edge $\{i,j\} \in E$.
- $f(S) = f(V \setminus S) := \sum_{k \in S} \sum_{\ell \in V \setminus S} L_{k\ell}$ and $f(\varnothing) = f(V) := 0$.
- $\boxed{\Delta_i \Delta_j f(S) = -2L_{ij}}$ for any $S \in 2^V$.



$f(S)$ with $S = \{k, \ell, m\}$ $\quad$ $\Delta_j f(S) = L_{ji} + L_{jq} - L_{j\ell} - L_{jm}$ $\quad$ $\Delta_j f(S \cup \{i\})$ $\quad$ $\Delta_i \Delta_j f(S) = -2L_{ij}$

6. Coverage function

- V set of points in $\mathbb{R}^2$. B_i ball centered at $i \in V$. $f(S) := \mathrm{Vol}(\bigcup_{i \in S} B_i)$.
- $\boxed{|\Delta_i \Delta_j f(S)| = \mathrm{Vol}\left(B_i \cap B_j \setminus \bigcup_{k \in S} B_k\right) \le \mathrm{Vol}(B_i \cap B_j)}$



$f(S)$ with $S = \{k, \ell\}$ $\quad$ $\Delta_j f(S)$ $\quad$ $\Delta_j f(S \cup \{i\})$ $\quad$ $|\Delta_i \Delta_j f(S)|$

7. Determinantal point process

- $V = \{1, \ldots, n\}$. $L \in \mathbb{R}^{n \times n}$ pos. definite. $(X_v)_{v \in V}$ Gaussian r.v.'s covariance L.
- $f(S) := \log\det L_S$ where $L_S := (L_{ij})_{i,j \in S}$ and $f(\varnothing) := 0$.
- $\boxed{\Delta_i \Delta_j f(S) = -2I(X_i; X_j | X_S)}$
- **CAVEAT:** Conditional mutual information not monotone in S ($\neq$ entropy).

8. Back to optimization!

NEXT STEP: Use Hessian in **combinatorial optimization**.

- Dimension-free uniform control on Hessian can be exploited to **get fastest convergence rates for ordinary greedy-type algorithms**.
- Work to be posted soon on **arXiv**.