Fast Mixing for Discrete Point Processes

Patrick Rebeschini
(joint work with Amin Karbasi)

YALE INSTITUTE FOR
NETWORK SCIENCE

COLT 2015
28th Annual Conference on Learning Theory

July 6, 2015

Role of submodularity in probability?

- **Combinatorial optimization:** **Submodular functions** 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 $i, j \notin S$, $i \neq j$

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

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

  $$\mathbb{P}(\mathbf{S} = S) := \frac{e^{-\beta f(S)}}{Z}, \quad \beta > 0, \quad S \in 2^V.$$

  ISSUE: (Djolonga, Krause, 2014) have **bounds exp. bad in** $\operatorname{card} V$.

Q: Can we get **dimension-free bounds**? Is submodularity right notion?

Fast mixing MCMC: control on Hessian

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

Main result

For a **generic** set function f, if

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

then Gibbs sampler $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_t$ is fast mixing (mixing time $\sim O(\log(\operatorname{card} V))$)

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

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

- **Key result: Bound does not depend on dimension** $\operatorname{card} V$.
- **Fundamental property: Dimension-free uniform control on Hessian**.

NOTE: No previous literature on Hessian of set functions.

Hessian captures "curvature" much better than *curvature*!

- Many results on submodularity rely on *curvature* c (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., $f(S) = \sum_{i \in S} w_i$.

- **Hessian** more natural concept to **characterize "curvature"** and **locality**:

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

- In many canonical applications (facility location, cut-function, etc.) Hessian is **sparse** and can be **easily computed or uniformly bounded**.

WORK IN PROGRESS: Use Hessian in combinatorial optimization!