1. Role of submodularity in probability?

- **Combining submodularity:** Submodularity extensively studied. Let $V$ be a finite set and $f : S \in 2^V \rightarrow \mathbb{R}$ be a set function.

$$\Delta_f(S) := f(S \cup i) - f(S) \quad \text{(gradient of } f)$$

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

$$\Delta \Delta_f(S) \leq \Delta f(S \cup i) - \Delta f(S) \leq 0 \quad \text{(Hessian of } f)$$

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

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

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

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 $S_0 \in 2^V$ from a given distribution (e.g., uniform); Set $S = S_0$; for $s = 1, \ldots, t$

for $i \in V$ times do

- Sample $i \in V$ uniformly. Draw $C \in \{0, 1\}$ with $P(C=0) = \frac{1}{1 + \exp \Delta f(i)}$.
- If $C = 0$ then set $S \leftarrow S \setminus \{i\}$, else set $S \leftarrow S \cup \{i\}$.

Output: Markov chain $S_0, S_1, \ldots, S_t$.

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

$$\beta \|M\|_\infty = \beta \max_{i \notin S} \sum_j M_{ij} \leq \gamma < 1 \quad \text{where} \quad M_{ij} \propto \max_{S \in 2^V \setminus \{S\}} |\Delta \Delta f(S)|$$

then $S_0, S_1, \ldots, S_t$ is fast mixing (mixing time $\tau(x) \leq \left\lceil \frac{\log(1 - \gamma)}{\gamma} \right\rceil$) and

$$\frac{1}{N} \sum_1^N 1(S[S] \ni i) - P(S \ni i) \leq \gamma + \frac{1}{\sqrt{N}}$$

where $S[1], \ldots, S[N]$ 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 $\beta$ 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 correlaction: $\max_{S \in 2^V \setminus \{S\}} |\Delta \Delta f(S)| \leq \alpha e^{-d(i,j)}$.
- Finite-range correlations: $\max_{S \in 2^V \setminus \{S\}} |\Delta \Delta f(S)| \leq \left\{ \begin{array}{ll} c & \text{if } d(i,j) \leq r, \\ 0 & \text{if } d(i,j) > r. \end{array} \right.$

4. Hessian and curvature

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

$$c := 1 - \min_{i \in V} \min_{S \in 2^V \setminus \{S\}} \frac{\Delta f(S \cup i) - \Delta f(S)}{f(i)} = 1 - \min_{i \in V} \frac{\Delta f(V \setminus i)}{f(i)} \in [0, 1].$$

We have $c = 0$ if and only if function is modular, i.e., $f(S) = \sum_{i \in S} w_i$.

Curvature is easy to compute (minimum is over $|V|$ terms).

- **Hessian** is a more natural concept to characterize “curvature”.

- Hessian also captures locality.

  - In general $\max_{S \in 2^V \setminus \{S\}} |\Delta \Delta f(S)|$ is not easy to compute (maximum is over $2^{2^V - 2}$ 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.
- $f(S) = f(V \setminus S) := \sum_{i \in S} w_i$ and $f(\emptyset) = f(V) = 0$.

- $\Delta \Delta f(S) = -2I_{\not\ni}$ for any $S \in 2^V$.

6. Coverage function

- $V$ set of points in $\mathbb{R}^d$, $B_i$ ball centered at $i \in V$. $f(S) := B_i \cup B_j \subseteq V$. $\beta \Delta \Delta f(S) = 0$.

7. Determinantal point processes

- $V = \{1, \ldots, n\}$. $L \in \mathbb{R}^{n \times n}$ pos. definite. $(X, \xi_i)$ Gaussian r.v.'s covariance $L$.

- $f(S) := \log \det L_S$ where $L_S := (L_{ij})_{ij \in S}$ and $f(\emptyset) := 0$.

- $\Delta \Delta f(S) = -2L_{X_i X_j}$

- **CAVEAT:** Conditional mutual information not monotone in $S(\neq \emptyset)$.

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.