1 Overview

In this lecture we will do the non-commutative Bernstein Inequality and Graph Sparsification problem.

2 Bernstein Inequality

Let $X_1, X_2, \ldots, X_n$ be $n$ independent, not necessarily identically distributed random variables. Further, $|X_i| \leq K \forall i \in [n]$ and $\mathbb{E} \left[ \sum_{i=1}^{n} X_i^2 \right] \leq \sigma^2$.

We wish to find the tail bounds for $|\sum_{i=1}^{n} X_i|$, i.e., $\mathbb{P} \left[ |\sum_{i=1}^{n} X_i| \geq t \right] \leq ?$

Note that $X_i$s are sub-gaussian($K$). This in turn implies that $\sum_{i=1}^{n} X_i$ is sub-gaussian($K\sqrt{n}$). Thus,

$$\mathbb{P} \left[ \left| \sum_{i=1}^{n} X_i \right| \geq t \right] \leq e^{-\frac{t^2}{2K^2n}}.$$

This in turn implies $|\sum_{i=1}^{n} X_i| \simeq K\sqrt{n}$. However, note that the bound is weak when $\sigma \ll K\sqrt{n}$.

Note that $X_i$s are also sub-gamma random variables. $\mathbb{E}[X_i^2] \leq \sigma_i^2 K^2$, $|X_i| \leq K$ implies that $X_i$ is sub-gamma($2\sqrt{2}\sigma, K, 4K$). Let us assume $\sigma_i$ is such that it subsumes $K$ in the argument. Thus, $X_i \in$ sub-gamma($2\sqrt{2}\sigma, 4K$), and $\sum_{i=1}^{n} X_i \in$ sub-gamma($2\sqrt{2}\sigma, 4K$). Using bounds for sub-gamma random variables, we can now write

$$\mathbb{P} \left[ \left| \sum_{i=1}^{n} X_i \right| \geq t \right] \leq 2e^{-\min \left\{ \frac{t^2}{16\sigma^2}, \frac{4Kt}{2\sqrt{2}\sigma} \right\}}.$$

But the mean may not be 0. We use $\mathbb{E}[\sum_{i=1}^{n} X_i] \leq \mathbb{E}[\sum_{i=1}^{n} X_i^2]^{\frac{1}{2}} = \sigma$ to write

$$\mathbb{P} \left[ \left| \sum_{i=1}^{n} X_i \right| \geq t \right] \leq 2e^{-\min \left\{ \frac{(t-\sigma)^2}{16\sigma^2}, \frac{(t-\sigma)}{4K} \right\}} \leq 2e^{-\frac{(t-\sigma)^2}{16\sigma^2}} \leq 2e^{-\frac{t^2}{16\sigma^2}}.$$

1
where $C > 0$ is some constant. Also, note that $(a)$ is meaningful only if $(t - \sigma) \geq 4\sigma\sqrt{\ln(2)}$.

**Notation:** $\|A\| = \sup_{\|x\| \leq 1} \|Ax\|_2$.

**Theorem 1** (Non-commutative Bernstein inequality). *Extension of Bernstein-type inequalities to matrices.*

Let $X_1, \ldots, X_m$ be independent symmetric matrices with zero mean, i.e., $E[X_i] = 0$ $\forall i \in [m]$. Also, $\|X_i\| \leq K \forall i \in [m]$, and $\sum_{i=1}^{n} E[X_i^2] \leq \sigma^2$. Then, $\exists C < 0$, such that

$$
P \left[ \left\| \sum_{i=1}^{n} X_i \right\| \geq t \right] \leq 2n \cdot e^{C \min \left\{ \frac{t^2}{4}, \frac{t}{K} \right\}}$$

We omit the proof of this theorem.

**Theorem 2** (R-V theorem). Let $X_1, \ldots, X_m$ be independent, and identically distributed vectors in $\mathbb{R}^n$ such that $\|X_i\|_2 \leq K$ ($K \geq 1$), and $\|E[X_iX_i^\top]\| \leq 1 \forall i \in [m]$. Then,

$$
E \left[ \left\| \frac{1}{m} \sum_{i=1}^{m} X_iX_i^\top - E[XX^\top] \right\| \right] \leq K \sqrt{\frac{\log n}{m}}
$$

**Proof.** Let $Y_i = X_iX_i^\top - E[X_iX_i^\top]$. We want to apply the non-commutative Bernstein theorem to $\sum_{i=1}^{m} Y_i$.

**Upper bound for $Y$:**

$$
\|Y_i\| \leq \|X_iX_i^\top\| + \|E[X_iX_i^\top]\| \leq 2K^2
$$

**Upper bound for $\left\| \sum_{i=1}^{m} E[Y_i^2] \right\|$**

$$
\left\| \sum_{i=1}^{m} E[Y_i^2] \right\| \leq m\|E[Y_i^2]\| = m\left(\|E[(XX^\top)^2 - E[XX^\top]^2]\| \leq m\left(\|E[\|X\|_2^2 \cdot XX^\top]\| + \|E[XX^\top]\|^2\right) \leq 2mK^2
$$

We can now apply the non-commutative Bernstein inequality.

$$
P \left[ \left\| \sum_{i=1}^{m} E[Y_i]\right\| \geq mt \right] \leq 2n \cdot e^{-C \min \left\{ \frac{mt^2}{2K^2}, \frac{mt}{K^2} \right\}}
$$

Hence, when $t \geq \frac{K^2}{m} \log \left( \frac{2}{\delta} \right)$, and $t \geq K \sqrt{\frac{\log(m)}{m}}$
\[
\mathbb{P}\left[\left\| \sum_{i=1}^{m} \mathbb{E}[Y_i] \right\| \geq C_2 K \sqrt{\log \left( \frac{n}{\epsilon} \right) / m} \right\] \leq \delta
\]

More on the subject can be found here [2].

3 Graph Sparsifier

Graph Sparsification problem is the following: Given a dense graph \( G = (V, E_G, W_G) \), find a sparse graph \( H = (V, E_H, W_H) \), which approximately preserves some properties of \( G \). The vertex set will remain the same, but the edge set and their weights can be different. We will henceforth denote \(|V|\) by \( n \).

3.1 Cut-Sparsifier

In the first lecture we studied a randomized algorithm to compute the min-cut in a graph. Here we study a related problem of finding a cut-sparsifier, namely, a sparse graph \( H \), that approximately preserves all the cuts in \( G \).

For a given graph \( G = (V, E, W) \), a cut \( S \subseteq V \) has size:

\[
C_G(S) = \sum_{(u,v) \in E} W(u,v) \cdot I_{\{u \in S, v \notin S\}}
\]

**Definition 3 (Cut-sparsifier).** \( H \) is a cut-sparsifier for \( G \) if:

\[
\forall S \subseteq V, C_H(S) = (1 \pm \epsilon)C_G(S)
\]

3.2 Spectral Sparsifier

The Spectral Sparsifier is a generalized form of cut-sparsification [1]. Let us define

\[
L_G = \sum_{(u,v) \in E_G} A_{u,v}
\]

so that,

\[
L_G(u,v) = \begin{cases} 
-W(u,v) & u \neq v \\
\sum_t W(u,t) & u = v 
\end{cases}
\]

\( L_G \) is called the Laplacian Matrix of the graph. Let \( P_G(x) = x^T L_G x \)

**Definition 4 (Spectral Sparsifier).** A spectral sparsifier is a graph that spectrally approximates the graph Laplacian, i.e. for all vectors \( x \), we should have

\[
P_H(x) = (1 \pm \epsilon)P_G(x)
\]
\[
(1 - \epsilon)x^T L_G x \leq x^T L_H x \leq (1 + \epsilon)x^T L_G x \quad \forall \ x \in \mathbb{R}^n \\
\iff (1 - \epsilon)L_G \preceq L_H \preceq (1 + \epsilon)L_G
\]

**Notation:** \( \preceq \) is the generalized matrix inequality on symmetric matrices: two symmetric matrices \( A \) and \( B \) satisfy \( A \preceq B \) iff \( (B - A) \) is positive semidefinite.

**Theorem 5.** *Spectral Sparsifier \( \implies \) Cut-sparsifier*

*Proof.* Will be done in next class. \( \square \)

**References**
