1 Overview

In this lecture we give a proof of the Rudelson-Vershynin (RV) theorem [RV05], and then begin graph sparsification.

2 Inequalities

We derive the Bernstein inequality for scalar random variables, extend this result to symmetric matrices, and then prove the RV theorem.

2.1 Bernstein Inequality

Let \(X_1, X_2, \ldots, X_n\) be independent random variables such that

\[
\mathbb{E} \left[ \sum_i X_i \right] = 0 \quad \text{max } \mathbb{E} |X_i| \leq K \quad \sum_i \mathbb{E} [X_i^2] \leq \sigma^2.
\]

Then \(\sum_i X_i = \text{subgamma}(2\sigma^2, 2K)\), which gives the bound

\[
\Pr \left[ \left| \sum_i X_i \right| \geq t \right] \leq 2 \exp \left( -\frac{1}{4} \min \left( \frac{t^2}{\sigma^2}, \frac{t}{K} \right) \right).
\]

2.2 Matrix Bernstein Inequality

Let \(X_1, X_2, \ldots, X_N\) be independent, symmetric matrices in \(\mathbb{R}^{n \times n}\) such that

\[
\mathbb{E} [X_i] = 0 \quad \text{max } \|X_i\| \leq K \quad \left\| \sum_i \mathbb{E} [X_i^2] \right\| \leq \sigma^2
\]

where we define \(\|\|\) as the spectral norm for matrices, the \(l^2\) norm for vectors and the absolute value norm for scalars. Analogizing from the scalar case,

\[
\Pr \left[ \left\| \sum_i X_i \right\| \geq t \right] \leq 2n \exp \left( -C \min \left( \frac{t^2}{\sigma^2}, \frac{t}{K} \right) \right)
\]

for some constant \(C\). For a proof see [Tro15].
2.3 Rudelson-Vershynin Inequality

**Theorem 1.** Let there be $m$ independent vectors $X_i \in \mathbb{R}^n$ such that

$$\max_i \|X_i\| \leq K \quad \|E[X_i X_i^T]\| \leq 1.$$  

Then

$$E \left[ \left\| \frac{1}{m} \sum_i X_i X_i^T - \frac{1}{m} \sum_i E[X_i X_i^T] \right\| \right] \lesssim K \sqrt{\frac{\log(n)}{m}},$$

for $K \sqrt{\frac{\log(n)}{m}} \leq 1$.

**Proof.** Let $Y_i = X_i X_i^T - E[X_i X_i^T]$. We would like to bound $\|\sum_i Y_i\|$ similar to equation (1), which can be done by bounding $\max_i \|Y_i\|$ and $\|\sum_i E[Y_i^2]\|$. We have that

$$\max_i \|Y_i\| \leq \max_i \|X_i X_i^T\| + \|E[X_i X_i^T]\|$$

$$\leq \max_i \|X_i\|^2 + \|E[X_i X_i^T]\|$$

$$\leq K^2 + 1$$

$$\leq 2K^2$$

and that

$$\left\| \sum_i E[Y_i^2] \right\| \leq \sum_i \|E[(X_i X_i^T - E[X_i X_i^T])^2]\|$$

$$\leq \sum_i \|E[X_i X_i^T X_i X_i^T] - E[X_i X_i^T] E[X_i X_i^T]\|$$

$$\leq \sum_i K^2 \|E[X_i X_i^T]\| + \|E[X_i X_i^T]\|^2$$

$$\leq m(K^2 + 1)$$

$$\leq 2mK^2.$$  

Then using the Matrix Bernstein inequality,

$$\Pr \left[ \left\| \sum_i Y_i \right\| \geq t \right] \leq 2n \exp \left( -C \min \left( \frac{t^2}{2mK^2}, \frac{t}{2K^2} \right) \right).$$  

for some constant $C$. For our application it’s the case that $\frac{t^2}{2mK^2} \leq \frac{t}{2K^2}$, so

$$E \left[ \left\| \frac{1}{m} \sum_i Y_i \right\| \right] \lesssim K \sqrt{\frac{\log(n)}{m}}.$$  

For the original proof see theorem 3.1 in [RV05].
3 Graph Sparsification

We approximate a weighted graph \( G = (V, E, \omega) \) by another graph \( H = (V, \tilde{E}, \tilde{\omega}) \) where \( |\tilde{E}| = O(K^2 \log(|V|)/\varepsilon^2) \). We do this by proving through the RV theorem that the Laplacian of \( G \) is approximately similar to the Laplacian of \( H \), a random matrix that results from sampling edges from \( E \) dependent on effective resistances and edge weights. The original material for this section can be seen in [SS08].

3.1 Laplacian

For a weighted graph \( G = (V, E, \omega) \), we define its Laplacian as

\[
L_G = \begin{cases} 
-\omega(u,v) & \text{if } u \neq v \\
\sum_{z \in \{x | (u,x) \in E\}} w(u,z) & \text{if } u = v.
\end{cases}
\]

This is equivalently defined as

\[
L_G = \sum_{e \in E} w_e y_e y_e^T
\]

where \( y_e \in \mathbb{Z}^{|V|} \) is the all 0s vector besides \( y_e(u) = 1 \) and \( y_e(v) = -1 \).

The Laplacian is useful because we can use it to approximate \( G \). For instance, consider \( G \) a representation of a circuit and define the total power needed to run the circuit for particular node-voltages as

\[
P_G(x) = \sum_{e=(u,v) \in E} (x_u - x_v)^2 w_e = x^T L_G x,
\]

where \( x \in \mathbb{R}^{|V|} \) is the node-voltages. For our approximation we would like to prove for all \( x \) that

\[
(1 - \varepsilon)P_G(x) \leq P_H(x) \leq (1 + \varepsilon)P_G(x)
\]

since this implies that

\[
(1 - \varepsilon)x^T L_G x \leq x^T L_H x \leq (1 + \varepsilon)x^T L_G x.
\]

3.2 Effective Resistance

In order to determine how likely each edge should be included in an approximation, we analogize edge weights, \( w_e \), to conductance, and measure the effective resistances, \( R_e \), between two nodes that have an edge. We define \( p_e \), the probability of sampling an edge, as

\[
p_e = w_e R_e.
\]

For an intuitive sense why sampling dependent on \( w_e \) and \( R_e \) works, note that the effective resistance between two adjacent nodes is the same as the probability that a random spanning tree contains those nodes’ connector edge.
3.3 Spectral Sparsification

Remember that our goal is to approximate $G$ using a graph $H$ that has only $O(K^2 \log(|V|)/\varepsilon^2)$ edges. We can do this by continually sampling an edge from $G$, modifying this sampled edge’s weight and adding this modified edge to $\tilde{H}$, where $\tilde{H}$ starts as an all-0s matrix and tends to $H$ after enough samples. Let $Z$ be a random variable such that

$$Z = y_e \sqrt{\frac{w_e}{p_e}}$$

with probability $p_e$.

Then the expectation of $ZZ^T$ is

$$E \left[ ZZ^T \right] = \sum_{e \in E} p_e \frac{w_e}{p_e} y_e y_e^T = L_G.$$

In the next lecture we will show that using the RV theorem and setting $L_H = \frac{1}{m} \sum_{i=1}^{m} Z_i Z_i^T$ gives the bound

$$E[\|L_H - L_G\|] \lesssim K \sqrt{\frac{\log n}{m}}$$

when $m = O(K^2 \log(|V|)/\varepsilon^2)$.

References

