Lecture 23 — April 3, 2013

Prof. Nick Harvey *Scribe: Samira Samadi*

1 Tropp's theorem applications

Theorem 1.1 (*Tropp's theorem*). Let $X_1, \cdots, X_k$ be independent, $n \times n$ symmetric matrices with $0 \preccurlyeq X_i \preccurlyeq R\, I$. Let $\mu_{min}\, I \preccurlyeq \sum_i \mathbb{E}[X_i] \preccurlyeq \mu_{max}\, I$. Then $\forall \epsilon \in [0,1]$

$$Pr[\lambda_{max}(\sum_i X_i) \geq (1+\epsilon)\mu_{max}] \leq n\, e^{\frac{-\epsilon^2\, \mu_{max}}{3R}} \tag{1.1}$$

$$Pr[\lambda_{min}(\sum_i X_i) \geq (1-\epsilon)\mu_{min}] \leq n\, e^{\frac{-\epsilon^2\, \mu_{min}}{2R}} \tag{1.2}$$

As we remember from Lecture 19

Definition 1.2 (*Graph sparsifiers*). H is sparsifier of G if $(1-\epsilon)L_G \preccurlyeq L_H \preccurlyeq (1+\epsilon)L_G$.

We are going to use Tropp's theorem to prove the following theorem:

Theorem 1.3 (*Spielman, Srivastava*). $\exists H$ with $O(|V| \log \frac{|V|}{\epsilon^2})$ edges.

1.1 Reduction

As usual, we apply the reduction (Lemma 6) described in Lecture 14. This yields vectors $\{\, w_e \,:\, e \in E \,\} \subset \mathbb{R}^n$ where $n = |V| - 1$ and $\sum_e w_e w_e^{\mathsf{T}} = I$ so that the following conditions are equivalent:

(1) subgraph H of G with edge set F and weights $s : F \to \mathbb{R}_+$ is a sparsifier of G

(2) $(1-\epsilon)I \preccurlyeq \sum_{e \in F} s_e\, w_e w_e^{\mathsf{T}} \preccurlyeq (1+\epsilon)I$

So we assume that we have these vectors and try to prove (2).

1.2 Algorithm

- Parameters: $C := \frac{6 \log n}{\epsilon^2}$, $A \leftarrow 0$, $s \leftarrow 0$
- **Pseudocode:**

For $j := 1, \cdots, C$ {

For $e \in E$ {

with probability $\underbrace{w_e^\mathsf{T} w_e}_{=p_e}$:

$$F \leftarrow F \cup \{e\}$$

$$s_e \leftarrow s_e + \frac{1}{C\ \underbrace{w_e^\mathsf{T} w_e}_{p_e}}$$

$$A \leftarrow A + \frac{w_e w_e^\mathsf{T}}{C\ \underbrace{w_e^\mathsf{T} w_e}_{p_e}}$$

}

}

Claim 1.4. With probability at least 0.9, H has $O(n \log n/\epsilon^2)$ edges.

Proof.

$$E[|F|] = \sum_{j=1}^{C} \sum_{e \in E} \underbrace{w_e^\mathsf{T} w_e}_{p_e} = C \cdot \mathrm{tr}[(\sum_e \underbrace{w_e^\mathsf{T} w_e}_{p_e})] = C \cdot \mathrm{tr}[(\sum_e w_e w_e^\mathsf{T})] = C \cdot \mathrm{tr}[(I)] = C \cdot n = \frac{6\, n \log n}{\epsilon^2}$$

By Markov's inequality, H has at most $60n \log n/\epsilon^2$ edges with probability at least 0.9. □

Claim 1.5. With high probability $\lambda_i(A) \in [1-\epsilon, 1+\epsilon]$ for all i. Thus, letting H be the subgraph of G with edges F and weights s, we have that H is a sparsifier of G with high probability.

Proof. We use Tropp's theorem. The change to A in iteration (j, e) is a random matrix which we denote $Z_{j,e}$. Clearly

$$Z_{j,e} \;=\; \left\{ \begin{array}{ll} \dfrac{w_e w_e^\mathsf{T}}{C \cdot w_e^\mathsf{T} w_e} & \text{w.p } w_e^\mathsf{T} w_e \\ 0 & \text{o.w.} \end{array} \right\}.$$

Note that $\mathbb{E}[Z_{j,e}] = \frac{1}{C} w_e w_e^\mathsf{T}$. Also, the matrix $w_e w_e^\mathsf{T} / w_e^\mathsf{T} w_e$ has maximum eigenvalue 1, so $Z_{j,e}$ always has maximum eigenvalue at most $1/C$.

The output of the algorithm is the matrix $A = \sum_{j,e} Z_{j,e}$. Note that

$$\mathbb{E}[A] = \sum_{j,e} \mathbb{E}[Z_{j,e}] = \frac{1}{C} \sum_{j=1}^{C} \sum_{e \in E} w_e w_e^\mathsf{T} = \sum_{e \in E} w_e w_e^\mathsf{T} = I$$

We apply Tropp's theorem to the random matrices $X_1, \ldots, X_k$ which we define by letting

$$\{\ X_i \ :\ i = [k]\ \} \;=\; \{\ Z_{j,e} \ :\ j \in [C], e \in E\ \}.$$

We have $A = \sum_{i=1}^k X_i$, $\sum_{i=1}^k \mathbb{E}[X_i] = I$, and $\lambda_{max}(X_i) = \lambda_{max}(Z_{j,e}) \le \frac{1}{C}$. So we may apply Tropp's theorem with $\mu_{min} = \mu_{max} = 1$ and $R := \frac{1}{C}$. The theorem yields

$$Pr[\lambda_{max}(A) \ge 1+\epsilon] \;\le\; n \cdot e^{\frac{-\epsilon^2 \mu_{max}}{3R}} = n \cdot e^{\frac{-\epsilon^2 C}{3}} = n \cdot e^{-2 \log n} = \frac{1}{n}$$

$$Pr[\lambda_{min}(A) \le 1-\epsilon] \;\le\; n \cdot e^{\frac{-\epsilon^2 \mu_{min}}{2R}} = n \cdot e^{-3 \log n} \le \frac{1}{n}$$

So with probability at least $1 - 2/n$, we have $\lambda_i(A) \in [1-\epsilon, 1+\epsilon]$ for all i. □

1.3 p_e values

Recall that the reduction (Lemma 6 of Lecture 14) has the property that

$$w_e^\mathsf{T} w_e \;=\; (e_u - e_v)^\mathsf{T} L_G^+ (e_u - e_v) \;=\; p_e. \tag{1.3}$$

What is this quantity? We will show that it is the effective resistance between u and v. Think of every edge of G that is present as a 1-ohm resistor and other edges that are not present as an infinite resistor.

Fact 1.6 (Ohm's law). $v = i \cdot r$ (voltage difference equals current times resistance).

Since current has a direction, if we orient edges arbitrarily we can talk about current on each edge, *e.g.*, if the vertices are labeled $\{1, ..., n\}$ we can orient each edge from the smaller vertex to the larger vertex. Let $V \in \mathbb{R}^n$ be a vector giving the voltage level at the vertices.

$$i_{ab} = \text{current on edge } ab = \frac{(V_a - V_b)}{1} = (V_a - V_b) \tag{1.4}$$

We can write this in terms of matrices. Let U be the $E \times V$ matrix that is the the node-arc incidence matrix of this orientation of G. That is, for edge $e = \{u, v\} \in E$, row e of U is $y_e^\mathsf{T} = (e_u - e_v)^\mathsf{T}$.

Let $v \in \mathbb{R}^V$ be a vector giving the voltage at each vertex. Let $i \in \mathbb{R}^E$ be a vector giving the current across each edge. So (1.4) states that

$$i \;=\; U \cdot v \tag{1.5}$$

That is, given voltages v, we get the current on every edge. But the current depends only on the *voltage difference*, so this equation remains true if we increase all voltages by α:

$$i \;=\; U \cdot v \qquad \Longleftrightarrow \qquad i \;=\; U \cdot (v + \alpha\vec{1}) \quad \forall \alpha \in \mathbb{R}. \tag{1.6}$$

Next we state Kirchhoff's current conservation law.

Fact 1.7. Let i_a^{ext} be the external current entering at node a. Then

$$i_a^{\text{ext}} = \sum_{ab \text{ leaves } a} i_{ab} - \sum_{ab \text{ entering } a} i_{ab}.$$

That is, if the amount of current leaving a in the graph exceeds the amount of current entering a in the graph, then there must be an external current source at a supplying the required current.

Letting $i \in \mathbb{R}^E$ and $i^{ext} \in \mathbb{R}^V$, we can also write this in terms of matrices.

$$i^{\text{ext}} \;=\; U^\mathsf{T} \cdot i \;=\; U^\mathsf{T} U (v + \alpha\vec{1}) \;=\; L_G(v + \alpha\vec{1}) \quad \forall \alpha \in \mathbb{R}.$$

That is, given any voltages $v \in \mathbb{R}^V$ and any $\alpha \in \mathbb{R}$, we can find the external current required to support these voltages.

The reverse is true too: given a vector of external currents $i^{\text{ext}} \in \mathbb{R}^V$, we can find voltages $v \in \mathbb{R}^V$ that produce these currents. Recall that $\text{image}(L_G) = \text{span}(\vec{1})^\perp$. Let $v = L_G^+ \cdot i^{\text{ext}}$. Then we have

$$L_G \cdot v \;=\; L_G \cdot L_G^+ \cdot i^{\text{ext}} \;=\; i^{\text{ext}}$$

since $\vec{1}^{\mathsf{T}} i^{\text{ext}} = \sum_{a \in V} i_a^{\text{ext}} = 0$, again by Kirchhoff's current conservation law.

Now connect a 1-amp current source between two nodes a and b. Following Ohm's law ($v = ir$ with $i = 1$), the *effective resistance* between a and b is the voltage difference required to support this single unit of current. So our external current vector is

$$i^{ext} \;:=\; e_a - e_b.$$

The voltages that produce these currents are

$$v \;=\; L_G^+ \cdot i^{ext} \;=\; L_G^+(e_a - e_b).$$

The voltage difference between a and b is

$$v_a - v_b \;=\; (e_a - e_b)^{\mathsf{T}} L_G^+ (e_a - e_b).$$

This is the effective resistance between a and b. For any edge $ab \in E$, it is the same as our sampling probability p_{ab}, as shown in (1.3).