Spectral Graph Theory Lecture 22

Sparsification by Random Sampling

Daniel A. Spielman November 26, 2012

22.1 About these notes

These notes are not necessarily an accurate representation of what happened in class. The notes written before class say what I think I should say. The notes written after class way what I wish I said.

22.2 Overview

I am going to prove that every graph on n vertices has an ϵ-approximation with only $O(\epsilon^{-2}n \log n)$ edges (a result of myself and Srivastava [SS11]) Along the way, I will prove the matrix Chernoff bound of Ahlswede and Winter [AW02] and a special case of a concentration result of Rudelson [Rud99].

Fix references!

22.3 Sparsification

We say that a graphs G and H are ϵ-approximations of each other if

$$(1+\epsilon)^{-1}\boldsymbol{L}_H \preccurlyeq \boldsymbol{L}_G \preccurlyeq (1+\epsilon)\boldsymbol{L}_H.$$

Note that this relation is symmetric.

In this lecture, we will show that every graph G has a good approximation by a sparse graph. This is a very strong statement, as graphs that approximate each other have a lot in common. For example,

1. the effective resistance between all pairs of vertices are similar in the two graphs,
2. the eigenvalues of the graphs are similar,
3. the boundaries of all sets are similar, as these are given by $\boldsymbol{\chi}_S^T \boldsymbol{L}_G \boldsymbol{\chi}_S$, and
4. the solutions of linear equations in the two matrices are similar.

We will see how to construct sparse approximations of graphs (called sparsifiers) by random sampling. We will do this by creating a probability distribution on the edges, and then repeatedly using this distribution to choose edges to include in the graph. If we include an edge, we will increase its weight by dividing by the probability that we choose it.

Symbolically, our distribution is specified by letting p_e be the probability that we choose edge e. We require that

$$\sum_e p_e = 1.$$

If we do choose edge e, we multiply its weight by $1/p_e$.

Let $\boldsymbol{L}_e$ be the elementary Laplacian for edge e. If we draw just one edge from this distribution, its expected Laplacian is given by

$$\sum_e p_e(\boldsymbol{L}_e/p_e) = \sum_e \boldsymbol{L}_e = \boldsymbol{L}_G.$$

Let $\boldsymbol{R}$ be a random matrix with this distribution. That is,

$$\Pr\left[\boldsymbol{R} = \boldsymbol{L}_e/p_e\right] = p_e.$$

To create a sparsifier H with q edges, we will draw q edges from this distribution, with replacement, and divide the result by q. That is, we independently sample matrices $\boldsymbol{R}^1, \dots, \boldsymbol{R}^q$, each distributed as $\boldsymbol{R}$, and set

$$\boldsymbol{L}_H \stackrel{\text{def}}{=} \frac{1}{q}\sum_{i=1}^{q} \boldsymbol{R}^i.$$

So, $\boldsymbol{L}_H$ will equal $\boldsymbol{L}_G$ in expectation. To make H be close to G with reasonable probability, we need to sample enough edges and choose the probabilities p_e carefully.

22.4 A Little Transformation

From our study of preconditioning, we know that H is an ϵ-approximation of G if and only if

$$\left\| \boldsymbol{L}_G^{-1/2} \boldsymbol{L}_H \boldsymbol{L}_G^{-1/2} - \boldsymbol{\Pi} \right\| \leq \epsilon',$$

where we need to use an ϵ' that is slightly different from ϵ as we are dealing with slightly different definitions. When ϵ is small, it will be approximately equal to ϵ. Also recall that $\boldsymbol{\Pi}$ is the projection orthogonal to the nullspace, and that

$$\mathbb{E}_{\boldsymbol{R}^1,\dots,\boldsymbol{R}^q}\left[\boldsymbol{L}_G^{-1/2} \boldsymbol{L}_H \boldsymbol{L}_G^{-1/2}\right] = \boldsymbol{L}_G^{-1/2} \boldsymbol{L}_G \boldsymbol{L}_G^{-1/2} = \boldsymbol{\Pi}.$$

We will analyze H under this transformation. So, define

$$\boldsymbol{M}^i = \boldsymbol{L}_G^{-1/2} \boldsymbol{R}^i \boldsymbol{L}_G^{-1/2}.$$

We have

$$\mathbb{E}\left[\boldsymbol{M}^i\right] = \boldsymbol{\Pi}.$$

I will tell you now how we will choose the probabilities p_e. While this choice was initially based on intutition, we will eventually see that it optimizes the application of the concentration inequalities that we will apply. We set

$$p_e \stackrel{\text{def}}{=} \frac{1}{n-1}\left\|\boldsymbol{L}_G^{-1/2}\boldsymbol{L}_e\boldsymbol{L}_G^{-1/2}\right\|.$$

To see that these sum to 1, first let $\boldsymbol{b}_e$ be a vector that is one at one endpoint of e and -1 at the other, so that $\boldsymbol{L}_e = \boldsymbol{b}_e\boldsymbol{b}_e^T$. Then,

$$\begin{aligned}\left\|\boldsymbol{L}_G^{-1/2}\boldsymbol{L}_e\boldsymbol{L}_G^{-1/2}\right\| &= \left\|\boldsymbol{L}_G^{-1/2}\boldsymbol{b}_e\boldsymbol{b}_e^T\boldsymbol{L}_G^{-1/2}\right\| \\ &= \boldsymbol{b}_e^T\boldsymbol{L}_G^{-1/2}\boldsymbol{L}_G^{-1/2}\boldsymbol{b}_e \\ &= \boldsymbol{b}_e^T\boldsymbol{L}_G^{-1}\boldsymbol{b}_e,\end{aligned}$$

which you will recall is the effective resistance in G between the endpoints of edge e.

To see that these probabilities sum to 1, note that

$$\sum_e \boldsymbol{b}_e^T\boldsymbol{L}_G^{-1}\boldsymbol{b}_e = \sum_e \operatorname{Tr}\left(\boldsymbol{L}_G^{-1}\boldsymbol{b}_e\boldsymbol{b}_e^T\right) = \operatorname{Tr}\left(\boldsymbol{L}_G^{-1}\left(\sum_e \boldsymbol{b}_e\boldsymbol{b}_e^T\right)\right) = \operatorname{Tr}\left(\boldsymbol{L}_G^{-1}\boldsymbol{L}_G\right) = \operatorname{Tr}\left(\boldsymbol{\Pi}\right) = n-1.$$

The advantage of these choices is that our distribution is a distribution over the matrices

$$\frac{\boldsymbol{L}_G^{-1/2}\boldsymbol{L}_e\boldsymbol{L}_G^{-1/2}}{p_e},$$

all of which have the same norm, $n-1$. Concentration inequalities work best when all the items being summed have the same magnitude.

22.5 The Matrix Chernoff Bounds

We now prove the main concentration inequality that we will use in this lecture: the matrix Chernoff bound of Ahlswede and Winter [AW02]. The bound applies to a sum of random symmetric matrices. It tells us that such a sum is unlikely to have too large a norm.

Standard proofs of the Chernoff bounds in random variables X work by applying Markov's inequality to e^X. The proof of Ahlswede and Winter does this too, but with matrices. I should begin with a few observations about $e^{\boldsymbol{X}}$ for a symmetric matrix $\boldsymbol{X}$.

First, one can define $e^{\boldsymbol{X}}$ by the power series:

$$e^{\boldsymbol{X}} = \sum_{i\geq 0}\frac{1}{i!}\boldsymbol{X}^i.$$

Second, every eigenvetor of $\boldsymbol{X}$ is an eigenvector of $e^{\boldsymbol{X}}$. So, the eigenvalues of $e^{\boldsymbol{X}}$ are just the exponentials of the eigenvalues of $\boldsymbol{X}$. In particular, this means that $e^{\boldsymbol{X}}$ is positive definite whenever $\boldsymbol{X}$ is symmetric.

Our analysis will rely on a fundamental fact about matrix exponentials that we will unfortunately not have time to prove in this class, the Golden-Thompson inequality:

Theorem 22.5.1. *For symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$\operatorname{Tr}\left(e^{\boldsymbol{A}+\boldsymbol{B}}\right) \leq \operatorname{Tr}\left(e^{\boldsymbol{A}} e^{\boldsymbol{B}}\right).$$

We will also use a fact about traces that I will defer to the last problem set:

Claim 22.5.2. *For positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$\operatorname{Tr}(\boldsymbol{A}\boldsymbol{B}) \leq \|\boldsymbol{A}\| \operatorname{Tr}(\boldsymbol{B}).$$

Theorem 22.5.3. *Let $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^q$ be independent random symmetric n-dimensional matrices. Then, for every $t > 0$ and every $\lambda > 0$,*

$$Pr\left[\left\|\sum_{i=1}^{q} \boldsymbol{X}^i\right\| > t\right] \leq n e^{-\lambda t}\left(\prod_{i=1}^{q}\left\|\mathbb{E}\left[e^{\lambda \boldsymbol{X}_i}\right]\right\| + \prod_{i=1}^{q}\left\|\mathbb{E}\left[e^{-\lambda \boldsymbol{X}_i}\right]\right\|\right).$$

The term λ in the theorem is a parameter that we typically set to optimize the bound after we have chosen t. The utility of this bound clearly depends upon the terms $\mathbb{E}\left[e^{\lambda \boldsymbol{X}_i}\right]$. We get stronger bounds when they are small.

We will apply this theorem to the matrices

$$\boldsymbol{X}^i = \boldsymbol{M}^i - \mathbb{E}\left[\boldsymbol{M}^i\right] = \boldsymbol{M}^i - \boldsymbol{\Pi}.$$

Proof of Theorem 22.5.3. Let $\boldsymbol{S} = \sum_i \boldsymbol{X}^i$. We being with the observation that $\|\boldsymbol{S}\|$ is greater than t if and only the largest eigenvalue of $\boldsymbol{S}$ is greater than t or the smallest eigenvalue is less than $-t$. Exponentiating, we have

$$\lambda_{max}(\boldsymbol{S}) \geq t \quad \text{iff} \quad \lambda_{max}(e^{\boldsymbol{S}}) \geq e^t.$$

We then observe that

$$\lambda_{max}(e^{\boldsymbol{S}}) \geq e^t \quad \Longrightarrow \quad \operatorname{Tr}\left(e^{\boldsymbol{S}}\right) \geq e^t.$$

So,

$$\Pr\left[\lambda_{max}(\boldsymbol{S}) \geq t\right] \leq \Pr\left[\operatorname{Tr}\left(e^{\boldsymbol{S}}\right) \geq e^t\right] \leq e^{-t}\mathbb{E}\left[\operatorname{Tr}\left(e^{\boldsymbol{S}}\right)\right],$$

where the last inequality follows from Markov's inequality. As we will sometimes want a parameter λ in the proof, we use the more general form

$$\Pr\left[\lambda_{max}(\boldsymbol{S}) \geq t\right] \leq e^{-\lambda t}\mathbb{E}\left[\operatorname{Tr}\left(e^{\lambda \boldsymbol{S}}\right)\right].$$

The goal of the rest of the proof is to prove an upper bound on

$$\mathbb{E}\left[\operatorname{Tr}\left(e^{\lambda \boldsymbol{S}}\right)\right].$$

For all k let $\boldsymbol{S}^k = \sum_{i=1}^{k} \boldsymbol{X}^i$. We have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Tr}\left(e^{\lambda \boldsymbol{S}^k}\right)\right] &= \mathbb{E}\left[\mathrm{Tr}\left(e^{\lambda \boldsymbol{X}^k + \lambda \boldsymbol{S}^{k-1}}\right)\right] \\
&\leq \mathbb{E}\left[\mathrm{Tr}\left(e^{\lambda \boldsymbol{X}^k} e^{\lambda \boldsymbol{S}^{k-1}}\right)\right] && \text{(by Golden-Thompson)} \\
&= \mathbb{E}_{\boldsymbol{X}^1,\dots,\boldsymbol{X}^{k-1}}\left[\mathrm{Tr}\left(\mathbb{E}_{\boldsymbol{X}^k}\left[e^{\lambda \boldsymbol{X}^k}\right] e^{\lambda \boldsymbol{S}^{k-1}}\right)\right] && \text{(as trace is linear)} \\
&\leq \mathbb{E}_{\boldsymbol{X}^1,\dots,\boldsymbol{X}^{k-1}}\left[\left\|\mathbb{E}_{\boldsymbol{X}^k}\left[e^{\lambda \boldsymbol{X}^k}\right]\right\| \mathrm{Tr}\left(e^{\lambda \boldsymbol{S}^{k-1}}\right)\right] && \text{(by Claim 22.5.2)} \\
&= \left\|\mathbb{E}_{\boldsymbol{X}^k}\left[e^{\lambda \boldsymbol{X}^k}\right]\right\| \cdot \mathbb{E}_{\boldsymbol{X}^1,\dots,\boldsymbol{X}^{k-1}}\left[\mathrm{Tr}\left(e^{\lambda \boldsymbol{S}^{k-1}}\right)\right] \\
&= \left\|\mathbb{E}_{\boldsymbol{X}^k}\left[e^{\lambda \boldsymbol{X}^k}\right]\right\| \cdot \mathbb{E}\left[\mathrm{Tr}\left(e^{\lambda \boldsymbol{S}^{k-1}}\right)\right].
\end{aligned}$$

By induction on k, we conclude that

$$\mathbb{E}\left[\mathrm{Tr}\left(e^{\lambda \boldsymbol{S}^q}\right)\right] \leq \prod_{i=1}^{q} \left\|\mathbb{E}_{\boldsymbol{X}^i}\left[e^{\lambda \boldsymbol{X}^i}\right]\right\| \mathrm{Tr}\left(\boldsymbol{I}\right) = n \prod_{i=1}^{q} \left\|\mathbb{E}\left[e^{\lambda \boldsymbol{X}^i}\right]\right\|.$$

So,

$$\Pr\left[\lambda_{max}(\boldsymbol{S}) \geq t\right] \leq e^{-\lambda t} n \prod_{i=1}^{q} \left\|\mathbb{E}_{\boldsymbol{X}^i}\left[e^{\lambda \boldsymbol{X}^i}\right]\right\|.$$

We may treat the smallest eigenvalue of $\boldsymbol{S}$ similarly. □

As I mentioned before, we will set

$$\boldsymbol{X}^i = \boldsymbol{M}^i - \mathbb{E}\left[\boldsymbol{M}^i\right] = \boldsymbol{M}^i - \boldsymbol{\Pi},$$

and we will take

$$q \stackrel{\text{def}}{=} 4n \log(2n)/\epsilon^2$$

samples. We need to prove an upper bound on

$$\left\|\mathbb{E}\left[e^{\lambda \boldsymbol{X}}\right]\right\|.$$

To this end, recall that we chose our probability distribution so that $\boldsymbol{M}$ is a random matrix that always satisfies

$$\|\boldsymbol{M}\| = n - 1$$

So,

$$\|\boldsymbol{X}\| = \|\boldsymbol{M} - \boldsymbol{\Pi}\| \leq \|\boldsymbol{M}\| = n - 1,$$

assuming $n \geq 2$. We will also use the fact that

$$\mathbb{E}\left[\boldsymbol{M}\right] = \boldsymbol{0}.$$

Lemma 22.5.4. *Let $\boldsymbol{X}$ be a random symmetric matrix such that $\mathbb{E}\left[\boldsymbol{X}\right] = \boldsymbol{0}$, and let ν be a number so that it is always true that $\|\boldsymbol{X}\| \leq \nu$ and $\lambda_{max}\left(\mathbb{E}\left[\boldsymbol{X}^2\right]\right) \leq \nu^2$. Then, for $\lambda \leq 1/\nu$ we have*

$$\left\|\mathbb{E}\left[e^{\lambda \boldsymbol{X}}\right]\right\| \leq \left\|e^{\lambda^2 \mathbb{E}\left[\boldsymbol{X}^2\right]}\right\|.$$

Proof. We use two inequalities on the exponential that transfer directly to matrices:

$$\begin{aligned} 1 + x &\leq e^x, \text{ and} \\ e^x &\leq 1 + x + x^2, \text{ for } x \in [-1, 1]. \end{aligned}$$

For $\lambda \leq 1/\nu$, $\|\lambda \boldsymbol{X}\| \leq 1$. So,

$$e^{\lambda \boldsymbol{X}} \preccurlyeq \boldsymbol{I} + \lambda \boldsymbol{X} + \lambda^2 \boldsymbol{X}^2.$$

Using $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{0}$, this gives

$$\begin{aligned} \mathbb{E}\left[e^{\lambda \boldsymbol{X}}\right] &\preccurlyeq \boldsymbol{I} + \lambda^2 \mathbb{E}\left[\boldsymbol{X}^2\right]. \\ &\preccurlyeq e^{\lambda^2 \mathbb{E}\left[\boldsymbol{X}^2\right]}, \end{aligned}$$

as $\left\|\mathbb{E}\left[\boldsymbol{X}^2\right]\right\| \leq \nu^2$. So,

$$\left\|\mathbb{E}\left[e^{\lambda \boldsymbol{X}}\right]\right\| \leq \left\|e^{\lambda^2 \mathbb{E}\left[\boldsymbol{X}^2\right]}\right\|.$$

□

It remains to bound the norm of $\mathbb{E}\left[\boldsymbol{X}^2\right]$, and to put the ingredients together.

Lemma 22.5.5. *Let* $\boldsymbol{\Pi}$ *be a projection matrix (symmetric and all eigenvalues in* $\{0, 1\}$*) and let* $\boldsymbol{M}$ *be a random positive semidefinite matrix such that* $\mathbb{E}[\boldsymbol{M}] = \boldsymbol{\Pi}$ *and* $\|\boldsymbol{M}\| \leq \nu$*, always. Let* $\boldsymbol{X} = \boldsymbol{M} - \boldsymbol{\Pi}$*. Then,*

$$\lambda_{max}\left(\mathbb{E}\left[\boldsymbol{X}^2\right]\right) \leq \nu.$$

Proof.

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{X}^2\right] &= \mathbb{E}\left[(\boldsymbol{M} - \mathbb{E}[\boldsymbol{M}])^2\right] \\ &= \mathbb{E}\left[\boldsymbol{M}^2\right] - 2\mathbb{E}[\boldsymbol{M}]\,\mathbb{E}[\boldsymbol{M}] + \mathbb{E}[\boldsymbol{M}]^2 \\ &= \mathbb{E}\left[\boldsymbol{M}^2\right] - \mathbb{E}[\boldsymbol{M}]^2 \\ &\preccurlyeq \mathbb{E}\left[\boldsymbol{M}^2\right]. \end{aligned}$$

To bound the latter term, observe that for a postive semidefinite matrix $\boldsymbol{A}$ we have

$$\boldsymbol{A}^2 \preccurlyeq \|\boldsymbol{A}\|\, \boldsymbol{A}$$

(I could put this on a problem set, but it is too easy). In our case, we always have $\|\boldsymbol{M}\| \leq \nu$. So,

$$\mathbb{E}\left[\boldsymbol{M}^2\right] \preccurlyeq \nu \mathbb{E}[\boldsymbol{M}] = \nu \boldsymbol{\Pi},$$

and

$$\lambda_{max}\left(\mathbb{E}\left[\boldsymbol{M}^2\right]\right) \leq \nu \lambda_{max}\left(\|\boldsymbol{\Pi}\|\right) = \nu.$$

□

22.6 Finishing The Argument

By combining the results of the last section, we obtain a variant of Rudelson's concentration theorem [Rud99].

Theorem 22.6.1. *Let* $\mathbf{\Pi}$ *be a projection matrix (symmetric and all eigenvalues in* $\{0,1\}$*) and let* $\boldsymbol{M}$ *be a random positive semidefinite matrix such that* $\mathbb{E}[\boldsymbol{M}] = \mathbf{\Pi}$ *and* $\|\boldsymbol{M}\| \leq \nu$*, always. Let* $\boldsymbol{M}^1, \ldots, \boldsymbol{M}^q$ *be independent random matrices with the same distribution as* $\boldsymbol{M}$*. Then, for every* $\epsilon > 0$*,*

$$Pr\left[\left\|\frac{1}{q}\sum_i \boldsymbol{M}^i - \mathbf{\Pi}\right\| \geq \epsilon\right] \leq 2ne^{-\epsilon^2 q/4\nu}.$$

Proof. We first multiply through by q to obtain

$$\Pr\left[\left\|\frac{1}{q}\sum_i \boldsymbol{M}^i - \mathbf{\Pi}\right\| \geq \epsilon\right] = \Pr\left[\left\|\sum_i \boldsymbol{M}^i - q\mathbf{\Pi}\right\| \geq \epsilon q\right].$$

By Theorem 22.5.3, for every $\lambda > 0$ this probability is at most

$$ne^{-\lambda t}\left(\prod_{i=1}^{q}\left\|\mathbb{E}\left[e^{\lambda \boldsymbol{X}_i}\right]\right\| + \prod_{i=1}^{q}\left\|\mathbb{E}\left[e^{-\lambda \boldsymbol{X}_i}\right]\right\|\right). \tag{22.1}$$

So, we set $\boldsymbol{X}^i = \boldsymbol{M}^i - \mathbf{\Pi}$ and apply Lemma 22.5.5 to show that $\lambda_{max}(\mathbb{E}\left[(\boldsymbol{X}^i)^2\right]) \leq \nu$. Having established this, we may apply Lemma 22.5.4 to prove that for $\lambda < 1/\nu$,

$$\left\|\mathbb{E}\left[e^{\lambda \boldsymbol{X}_i}\right]\right\| \leq \left\|e^{\lambda^2 \mathbb{E}\left[\boldsymbol{X}^2\right]}\right\| \leq e^{\lambda^2 \nu}.$$

We may similarly prove that

$$\left\|\mathbb{E}\left[e^{-\lambda \boldsymbol{X}_i}\right]\right\| \leq e^{\lambda^2 \nu}.$$

So, we obtain an upper bound on the probabilty in (22.1) of

$$2ne^{-\lambda t}e^{\lambda^2 q\nu} = 2ne^{\lambda^2 q\nu - \lambda t}$$

By differentiating the exponent with respect to t we find that the optimal choice of λ is

$$\lambda = t/2q\nu.$$

Substituting this into the bound on the probability yields

$$2ne^{-t/4q\nu}.$$

Finally substituting $t = \epsilon q$ gives

$$2ne^{-\epsilon^2 q/4\nu}.$$

□

In the case of matrix sparsification, we have $\nu = n - 1 \leq n$. So, if we take

$$q = 5n \ln(2n)/\epsilon^2,$$

we find that the probability that H fails to be an ϵ approximation of G is at most

$$2ne^{-\epsilon^2 q/4n} \leq 2ne^{-5\ln(2n)/4} = (2n)^{-1/4},$$

which is close to 0.

22.7 Open Problem

There is one way that I would like to make this construction cleaner. Instead of choosing edges with replacement, I would rather have each edge e appear in the graph independently with probability p_e. While this is very close to what we are doing, I have been unable to prove that it works. Much of this proof works for this case. We can create a random matrix $\boldsymbol{M}_e$ for each edge e that in expectation equals $\boldsymbol{L}_G^{-1/2} \boldsymbol{L}_e \boldsymbol{L}_G^{-1/2}$. We then have

$$\mathbb{E}\left[\sum_e \boldsymbol{M}_e\right] = \boldsymbol{\Pi}.$$

The problem is that in this situtation Lemma 22.5.5 is just not strong enough.

In the proof presented here, Lemma 22.5.5 is very powerful because the norm of the expected matrix is 1 and all of the eigenvalues of the expected matrix are 1. Howerver, in my alternative proposed proof both the norm of the expected matrix and the sum of its eigenvalues are 2.

References

[AW02] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *Information Theory, IEEE Transactions on*, 48(3):569–579, 2002.

[Rud99] M. Rudelson. Random vectors in the isotropic position,. *Journal of Functional Analysis*, 164(1):60 – 72, 1999.

[SS11] D.A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011.