

## Sparsification by Effective Resistance Sampling

*Daniel A. Spielman*

November 2, 2015

**Disclaimer**

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. I sometimes edit the notes after class to make them way what I wish I had said.

There may be small mistakes, so I recommend that you check any mathematically precise statement before using it in your own work.

These notes were last revised on November 3, 2015.

**17.1 Overview**

I am going to prove that every graph on  $n$  vertices has an  $\epsilon$ -approximation with only  $O(\epsilon^{-2}n \log n)$  edges (a result of myself and Srivastava [SS11]). We will prove this using a matrix Chernoff bound due to Tropp [Tro12].

We originally proved this theorem using a concentration bound of Rudelson [Rud99]. This required an argument that used sampling with replacement. When I taught this result in 2012, I asked if one could avoid sampling with replacement. Nick Harvey pointed out to me the argument that avoids replacement that I am presenting today.

**17.2 Sparsification**

For this lecture, I define a graph  $H$  to be an  $\epsilon$ -approximation of a graph  $G$  if

$$(1 - \epsilon)\mathbf{L}_G \preceq \mathbf{L}_H \preceq (1 + \epsilon)\mathbf{L}_G.$$

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  $\chi_S^T \mathbf{L}_G \chi_S$ , and

4. the solutions of linear equations in the two matrices are similar.

We will prove this by using a very simple random construction. For each edge  $(a, b) \in G$ , let  $R_{\text{eff}}(a, b)$  denote the effective resistance between  $a$  and  $b$  in  $G$ . Recall that this equals

$$(\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \mathbf{L}_G^+ (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b),$$

and that it is smaller when there are many short paths between  $a$  and  $b$ . Define

$$q_{a,b} = w_{a,b} R_{\text{eff}}(a, b) \quad \text{and} \quad p_{a,b} = \min(1, C(\log n) \epsilon^{-2} q_{a,b}),$$

where  $C$  is some absolute constant we will set later, and  $w_{a,b}$  is the weight of edge  $(a, b)$ . Our algorithm is simple: we include edge  $(a, b)$  in  $H$  with probability  $p_{a,b}$ . If we do include edge  $(a, b)$ , we give it weight  $w_{a,b}/p_{a,b}$ . We will show that the resulting graph  $H$  has  $O(n \log n/\epsilon^2)$  edges and is an  $\epsilon$  approximation of  $G$  with high probability.

The reason we employ this sort of sampling—blowing up the weight of an edge by dividing by the probability that we choose it—is that it preserves the matrix in expectation. Let  $\mathbf{L}_{a,b}$  denote the elementary Laplacian on edge  $(a, b)$  with weight 1, so that

$$\mathbf{L}_G = \sum_{(a,b) \in E} w_{a,b} \mathbf{L}_{a,b}.$$

We then have that

$$\mathbb{E} \mathbf{L}_H = \sum_{(a,b) \in E} p_{a,b} (w_{a,b}/p_{a,b}) \mathbf{L}_{a,b} = \mathbf{L}_G.$$

## 17.3 Matrix Chernoff Bounds

The main tool that we will use in our analysis is a theorem about the concentration of random matrices. These may be viewed as matrix analogs of the Chernoff bound that we saw in Lecture 5. These are a surprisingly recent development, with the first ones appearing in the work of Rudelson and Vershynin [Rud99, RV07] and Ahlswede and Winter [AW02]. The best present source for these bounds is Tropp [Tro12], in which the following result appears as Corollary 5.2.

**Theorem 17.3.1.** *Let  $\mathbf{X}_1, \dots, \mathbf{X}_m$  be independent random  $n$ -dimensional symmetric positive semidefinite matrices so that  $\|\mathbf{X}_i\| \leq R$  almost surely. Let  $\mathbf{X} = \sum_i \mathbf{X}_i$  and let  $\mu_{\min}$  and  $\mu_{\max}$  be the minimum and maximum eigenvalues of*

$$\mathbb{E}[\mathbf{X}] = \sum_i \mathbb{E}[\mathbf{X}_i].$$

Then,

$$\Pr \left[ \lambda_{\min} \left( \sum_i \mathbf{X}_i \right) \leq (1 - \epsilon) \mu_{\min} \right] \leq n \left( \frac{e^{-\epsilon}}{(1 - \epsilon)^{1 - \epsilon}} \right)^{\mu_{\min}/R}, \quad \text{for } 0 < \epsilon < 1, \text{ and}$$

$$\Pr \left[ \lambda_{\max} \left( \sum_i \mathbf{X}_i \right) \geq (1 + \epsilon) \mu_{\max} \right] \leq n \left( \frac{e^{\epsilon}}{(1 + \epsilon)^{1 + \epsilon}} \right)^{\mu_{\max}/R}, \quad \text{for } 0 < \epsilon.$$

It is important to note that the matrices  $\mathbf{X}_1, \dots, \mathbf{X}_m$  can have different distributions. Also note that as the norms of these matrices get bigger, the bounds above become weaker. As the expressions above are not particularly easy to work with, we often use the following approximations.

$$\begin{aligned} \left( \frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}} \right) &\leq e^{-\epsilon^2/2}, && \text{for } 0 < \epsilon < 1, \text{ and} \\ \left( \frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}} \right) &\leq e^{-\epsilon^2/3}, && \text{for } 0 < \epsilon < 1. \end{aligned}$$

Chernoff (and Hoeffding and Bernstein) bounds rarely come in exactly the form you want. Sometimes you can massage them into the needed form. Sometimes you need to prove your own. For this reason, you may some day want to spend a lot of time reading how these are proved.

I am going to want a bound on the probability that the smallest eigenvalue is small in terms of  $\mu_{max}$  rather than in terms of  $\mu_{min}$ . Fortunately, the bound I want is easy to obtain by substitution:

$$\begin{aligned} \Pr \left[ \lambda_{min} \left( \sum_i \mathbf{X}_i \right) \leq \mu_{min} - \epsilon \mu_{max} \right] &\leq n \exp \left( -\frac{1}{2} \left( \frac{\epsilon \mu_{max}}{\mu_{min}} \right)^2 \mu_{min} / R \right) \\ &= n \exp \left( -\frac{1}{2} \epsilon^2 \left( \frac{\mu_{max}}{\mu_{min}} \right) \mu_{max} / R \right) \\ &\leq n \exp \left( -\frac{1}{2} \epsilon^2 \mu_{max} / R \right). \end{aligned}$$

## 17.4 The key transformation

To apply the matrix Chernoff bound, and to explain why we sample by effective resistances, we make a simplifying transformation.

For positive definite matrices  $\mathbf{A}$  and  $\mathbf{B}$ , we have

$$\mathbf{A} \preceq (1 + \epsilon) \mathbf{B} \iff \mathbf{B}^{-1/2} \mathbf{A} \mathbf{B}^{-1/2} \preceq (1 + \epsilon) \mathbf{I}.$$

The same thing holds for singular semidefinite matrices that have the same nullspace:

$$\mathbf{L}_H \preceq (1 + \epsilon) \mathbf{L}_G \iff \mathbf{L}_G^{+2} \mathbf{L}_H \mathbf{L}_G^{+2} \preceq (1 + \epsilon) \mathbf{L}_G^{+2} \mathbf{L}_G \mathbf{L}_G^{+2},$$

where  $\mathbf{L}_G^{+2}$  is the square root of the pseudo-inverse of  $\mathbf{L}_G$ . Let

$$\mathbf{\Pi} = \mathbf{L}_G^{+2} \mathbf{L}_G \mathbf{L}_G^{+2},$$

which is the projection onto the range of  $\mathbf{L}_G$ .

As multiplication by a fixed matrix is a linear operation and expectation commutes with linear operations,

$$\mathbb{E} \mathbf{L}_G^{+2} \mathbf{L}_H \mathbf{L}_G^{+2} = \mathbf{L}_G^{+2} (\mathbb{E} \mathbf{L}_H) \mathbf{L}_G^{+2} = \mathbb{E} \mathbf{L}_G^{+2} \mathbf{L}_G \mathbf{L}_G^{+2} = \mathbf{\Pi}.$$

So, we really just need to show that this random matrix is probably close to its expectation,  $\mathbf{\Pi}$ . It would probably help to pretend that  $\mathbf{\Pi}$  is in fact the identity, as it will make it easier to understand the analysis. In fact, you don't have to pretend: you could project all the vectors and matrices onto the span of  $\mathbf{\Pi}$  and carry out the analysis there.

## 17.5 Why effective resistances

We will now see why we set the edge sampling probabilities in terms of effective resistances. We have

$$\begin{aligned}
 \sum_{(a,b) \in E} q_{a,b} &= \sum_{(a,b) \in E} w_{a,b} \text{R}_{\text{eff}}(a,b) \\
 &= \sum_{(a,b) \in E} w_{a,b} (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \mathbf{L}_G^+ (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b) \\
 &= \sum_{(a,b) \in E} w_{a,b} \text{Tr} \left( \mathbf{L}_G^+ (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b) (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \right) \\
 &= \text{Tr} \left( \sum_{(a,b) \in E} \mathbf{L}_G^+ w_{a,b} (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b) (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \right) \\
 &= \text{Tr} \left( \mathbf{L}_G^+ \sum_{(a,b) \in E} w_{a,b} \mathbf{L}_{a,b} \right) \\
 &= \text{Tr} \left( \mathbf{L}_G^+ \mathbf{L}_G \right) \\
 &= \text{Tr} \left( \mathbf{\Pi} \right) \\
 &= n - 1.
 \end{aligned}$$

There is a combinatorial reason that this is true:  $q_{a,b}$  is the probability that edge  $(a,b)$  appears in a random spanning tree of  $G$  when we sample spanning trees with probability proportional to the product of their edge weights. As every spanning tree has  $n - 1$  edges, the sum of these probabilities must be  $n - 1$ .

We can use this to bound the expected number of edges in  $H$ :

$$\sum_{(a,b) \in E} p_{a,b} = \sum_{(a,b) \in E} \min \left( 1, C(\log n) q_{a,b} / \epsilon^2 \right) \leq \sum_{(a,b) \in E} C(\log n) q_{a,b} / \epsilon^2 \leq Cn \log n / \epsilon^2$$

One can use a Chernoff bound (on real variables rather than matrices) to prove that it is exponentially unlikely that the number of edges in  $H$  is more than any small multiple of this.

The other advantage of sampling by effective resistances is that it guarantees that all the random matrices have small norm. Let

$$\mathbf{X}_{a,b} = \begin{cases} (w_{a,b}/p_{a,b}) \mathbf{L}_G^{+2} \mathbf{L}_{(a,b)} \mathbf{L}_G^{+2} & \text{with probability } p_{a,b} \\ 0 & \text{otherwise,} \end{cases}$$

so that

$$\mathbf{L}_G^{+/2} \mathbf{L}_H \mathbf{L}_G^{+/2} = \sum_{(a,b) \in E} \mathbf{X}_{a,b}.$$

We have

$$\begin{aligned} \|\mathbf{X}_{a,b}\| &= (w_{a,b}/p_{a,b}) \left\| \mathbf{L}_G^{+/2} \mathbf{L}_{(a,b)} \mathbf{L}_G^{+/2} \right\| \\ &= (w_{a,b}/p_{a,b}) \left\| \mathbf{L}_G^{+/2} (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b) (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \mathbf{L}_G^{+/2} \right\| \\ &= (w_{a,b}/p_{a,b}) \text{Tr} \left( \mathbf{L}_G^{+/2} (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b) (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \mathbf{L}_G^{+/2} \right) \\ &= (w_{a,b}/p_{a,b}) \text{Tr} \left( (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \mathbf{L}_G^{+/2} \mathbf{L}_G^{+/2} (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b) \right) \\ &= (w_{a,b}/p_{a,b}) \text{Tr} \left( (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \mathbf{L}_G^+ (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b) \right) \\ &= \frac{w_{a,b} \mathbf{R}_{\text{eff}}(a,b)}{p_{a,b}}. \end{aligned}$$

In the above chain of equalities, the equality of the norm and the trace follows from the fact that the matrix in question has rank 1. But, we really only needed the fact that the norm of a positive semidefinite matrix is at most its trace.

So, when  $p_{a,b} < 1$ , we have

$$\|\mathbf{X}_{a,b}\| \leq \frac{1}{C(\log n)\epsilon^{-2}}.$$

The point is that all of these norms are the same, and do not depend on  $(a,b)$ . We choose the probabilities  $p_{a,b}$  for exactly this reason.

## 17.6 The analysis

We have that

$$\sum_{(a,b) \in E} \mathbb{E} \mathbf{X}_{a,b} = \mathbf{\Pi}.$$

It remains to show that it is unlikely to deviate from this by too much.

We first consider the case in which  $p_{(a,b)} < 1$  for all edges  $(a,b)$ . If this were the case, then we could apply Theorem 17.3.1 with the bound

$$\|\mathbf{X}_{a,b}\| \leq \frac{1}{C(\log n)\epsilon^{-2}} \stackrel{\text{def}}{=} R.$$

Theorem 17.3.1 tells us that

$$\Pr \left[ \sum_{a,b} \mathbf{X}_{a,b} \geq (1 + \epsilon) \mathbf{\Pi} \right] \leq n \exp(-\epsilon^2 C \epsilon^{-2} \log n / 3) = n \exp(-(C/3) \log n) = n^{-(C/3)+1},$$

which is small for  $C > 3$ .

For the lower bound, we need to remember that we can just work orthogonal to the all-1s vector, and so treat the smallest eigenvalue of  $\mathbf{\Pi}$  as 1. We then find that

$$\Pr \left[ \sum_{a,b} \mathbf{X}_{a,b} \leq (1 - \epsilon)\mathbf{\Pi} \right] \leq n \exp(-\epsilon^2 C \epsilon^{-2} \log n / 2) = n \exp(-(C/2) \log n) = n^{-(C/2)+1},$$

which is small for  $C > 2$ . So, we should choose something like  $C = 4$ .

We finally return to deal with the fact that there might be some edges for which  $p_{a,b} = 1$  and so definitely appear in  $H$ . As these aren't really random, our intuition is that they shouldn't affect the bound. I will make this formal in a way that will look like cheating. The only reason these edges pose a difficulty is that the corresponding  $\mathbf{X}_{a,b}$  can have large norm. But, they are certainly independent.

For each  $(a,b)$  for which  $p_{a,b} = 1$ , we split  $\mathbf{X}_{a,b}$  into many independent random variables. For example, we could replace it with  $K$  copies of the random variable  $\mathbf{X}_{a,b}/K$  for some large  $K$ . This does not change the expectation of their sum, or the distribution of their sum as they all appear in  $H$  with probability 1. Each of these now has small norm, and so we can apply Theorem 17.3.1 as desired.

What is going on here is that we are really using another slight variation of Theorem 17.3.1 that we can derive from the original.

## 17.7 Open Problem

If I have time in class, I will sketch a way to quickly approximate the effective resistances of every edge in the graph. The basic idea, which can be found in [SS11] and which is carried out better in [LKP12], is that we can compute the effective resistance of an edge  $(a,b)$  from the solution to a logarithmic number of systems of random linear equations in  $\mathbf{L}_G$ . That is, after solving a logarithmic number of systems of linear equations in  $\mathbf{L}_G$ , we have information from which we can estimate all of the effective resistances.

In order to sparsify graphs, we do not actually need estimates of effective resistances that are always accurate. We just need a way to identify many edges of low effective resistance, without listing any that have high effective resistance. I believe that better algorithms for doing this remain to be found. Current fast algorithms that make progress in this direction and that exploit such estimates may be found in [LKP12, Kou14, CLM<sup>+</sup>14, LPS15].

## References

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

- [CLM<sup>+</sup>14] Michael B Cohen, Yin Tat Lee, Cameron Musco, Christopher Musco, Richard Peng, and Aaron Sidford. Uniform sampling for matrix approximation. *arXiv preprint arXiv:1408.5099*, 2014.
- [Kou14] Ioannis Koutis. Simple parallel and distributed algorithms for spectral graph sparsification. In *Proceedings of the 26th ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '14, pages 61–66, New York, NY, USA, 2014. ACM.
- [LKP12] Alex Levin, Ioannis Koutis, and Richard Peng. Improved spectral sparsification and numerical algorithms for sdd matrices. In *Proceedings of the 29th Symposium on Theoretical Aspects of Computer Science (STACS)*, 2012. to appear.
- [LPS15] Yin Tat Lee, Richard Peng, and Daniel A. Spielman. Sparsified cholesky solvers for SDD linear systems. *CoRR*, abs/1506.08204, 2015.
- [Rud99] M. Rudelson. Random vectors in the isotropic position,. *Journal of Functional Analysis*, 164(1):60 – 72, 1999.
- [RV07] Mark Rudelson and Roman Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *J. ACM*, 54(4):21, 2007.
- [SS11] D.A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011.
- [Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.