CS 496 Lecture 20: Matrix concentration and graph sparsification

Sidhanth Mohanty

November 20, 2025

1 Spectral sparsification of graphs

Let G be an undirected graph on n vertices, and let L_G denote the Laplacian of G. For every $x \in \mathbb{R}^V$ we have the quadratic form

$$x^\top L_G x = \sum_{e=\{u,v\}\in E(G)} (x_u - x_v)^2.$$

Definition 1.1 (Spectral sparsifier)**.** For $\varepsilon \in (0,1)$, a weighted subgraph $H = (V, E_H, w)$ is a $(1 \pm \varepsilon)$-spectral sparsifier of G if

$$(1-\varepsilon) L_G \preceq L_H \preceq (1+\varepsilon) L_G\,.$$

In this lecture, we will construct such an H with only $O\big(n \log n / \varepsilon^2\big)$ edges, based on the work of Spielman and Srivastava [SS08].

1.1 Importance sampling for PSD matrices and effective resistances

Before specializing to graphs, consider a finite collection of positive semidefinite (PSD) matrices $M_1, \ldots, M_m$ and their sum

$$M = \sum_{e=1}^{m} M_e.$$

The basic importance sampling trick is:

- Choose probabilities $p_1, \ldots, p_m > 0$.
- Sample each index e independently: let $\boldsymbol{z}_e \sim \text{Bernoulli}(p_e)$.
- Form the random sum
$$\widehat{\boldsymbol{M}} = \sum_{e=1}^{m} \frac{\boldsymbol{z}_e}{p_e} M_e.$$

Then

$$\mathbf{E}\,\widehat{\boldsymbol{M}} = M\,.$$

In particular, $\widehat{\boldsymbol{M}}$ is an unbiased estimator of the sum.

We now specialize the M_e to be the normalized edge Laplacians of a graph, and henceforth specialize these linear operators to the space orthogonal to the span of the indicator vectors of connected components.

For each edge e, define

$$M_e = L_G^{-1/2} L_e L_G^{-1/2}.$$

Each M_e is PSD and

$$\sum_{e \in E} M_e = \mathbb{I}\,.$$

Define the *effective resistance* of an edge e as

$$R_e := \|M_e\|\,.$$

We may verify that $R_e \leqslant 1$ since $0 \preceq M_e \preceq \mathbb{I}$.

We will use the following key bound.

Observation 1.2. $\sum_e R_e \leqslant n$.

Proof. Since each M_e is rank-1, we have $R_e = \operatorname{tr}(M_e)$ and $\sum_e R_e = \operatorname{tr}(\sum_e M_e) = \operatorname{tr}(\mathbb{I}) \leqslant n$. □

We now construct a random sparsifier $\boldsymbol{H}$ by sampling edges of G. Fix parameters $\varepsilon \in (0,1)$ and a sufficiently large constant $C > 0$. For each edge $e \in E$ define the sampling probability

$$p_e = \min\left\{1, \frac{C \log n}{\varepsilon^2} R_e\right\}.$$

We then form a random weighted subgraph $\boldsymbol{H}$ according to the importance sampling scheme described by the above family of p_e. By Observation 1.2, the number of edges in $\boldsymbol{H}$ is bounded by $O(n \log n/\varepsilon^2)$ with high probability.

To prove that $\boldsymbol{H}$ is a spectral sparsifier of G, we use the matrix Chernoff bound (see, e.g., [Tro15]).

Theorem 1.3 (Matrix Chernoff)**.** *Let X_0 be a deterministic positive semidefinite matrix, and let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_N$ be independent random self-adjoint matrices in $\mathbb{R}^{d \times d}$ such that*

$$0 \preceq \boldsymbol{X}_k \preceq R \cdot \mathbb{I} \qquad \text{for } k = 1, \dots, N,$$

for some scalar $R > 0$. Define

$$\boldsymbol{Y} = X_0 + \sum_{k=1}^{N} \boldsymbol{X}_k, \qquad \mu_{\min} = \lambda_{\min}(\mathbf{E}\,\boldsymbol{Y}), \qquad \mu_{\max} = \lambda_{\max}(\mathbf{E}\,\boldsymbol{Y}).$$

Then for $0 \leqslant \varepsilon \leqslant 1$,

$$\mathbb{P}\left[\lambda_{\min}(\boldsymbol{Y}) \leqslant (1-\varepsilon)\mu_{\min}\right] \leqslant d \cdot \exp\left(-\frac{\varepsilon^2}{2}\frac{\mu_{\min}}{R}\right),$$

$$\mathbb{P}\left[\lambda_{\max}(\boldsymbol{Y}) \geqslant (1+\varepsilon)\mu_{\max}\right] \leqslant d \cdot \exp\left(-\frac{\varepsilon^2}{3}\frac{\mu_{\max}}{R}\right).$$

We now analyze $L_{\boldsymbol{H}}$ using matrix Chernoff. Consider the normalized random matrix:

$$\boldsymbol{Z} = L_G^{-1/2} L_{\boldsymbol{H}} L_G^{-1/2}$$

By design, $\mathbf{E}\,\boldsymbol{Z} = \mathbb{I}$.

We now calculate a uniform bound satisfied by the summands $\boldsymbol{X}_e$ corresponding to edge e satisfies

$$0 \preceq \boldsymbol{X}_e \preceq \frac{1}{p_e} M_e \,,$$

so

$$\lambda_{\max}(\boldsymbol{X}_e) \leqslant \frac{1}{p_e} \lambda_{\max}(M_e) \leqslant \frac{1}{p_e} R_e.$$

By the definition of p_e, whenever $p_e < 1$ we have $p_e = \frac{C \log n}{\varepsilon^2} R_e$, and hence

$$\lambda_{\max}(\boldsymbol{X}_e) \leqslant \frac{\varepsilon^2}{C \log n}.$$

If $p_e = 1$, then $X_e = M_e$ deterministically, and can be accounted for by the X_0 term.

Thus, we can apply the matrix Chernoff bound with $R = \frac{\varepsilon^2}{C \log n}$, which then tells us that with high probability $(1-\varepsilon)\mathbb{I} \preceq M \preceq (1+\varepsilon)\mathbb{I}$.

Rearranging this inequality yields $(1-\varepsilon)L_G \preceq L_{\boldsymbol{H}} \preceq (1+\varepsilon)L_G$.

References

[SS08] Daniel A Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 563–568, 2008. 1

[Tro15] Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015. 2