Randomized Algorithms 2021A – Lecture 12
Graph Laplacians and Spectral Sparsification*

Robert Krauthgamer

1 Graph Laplacians

High-level motivation: We saw dimension reduction for ℓ_2 (the JL-lemma). What is the analogue for graphs (and combinatorial objects in general)? The idea is to find a *sparse* graph G' that is "similar" to G, either (1) in the sense of cuts in the graph, or (2) viewing a graph as a real matrix (i.e., a linear operator).

Graph Laplacians: Let $G = (V, E, w)$ be an undirected graph with edge weights $w_e \geq 0$, where $w_{ij} = 0$ effectively means that $ij \notin E$. As usual, it is equivalent to think of the unweighted case, and treat an edge weight as representing parallel edges.

Notation: Assume $V = \{1, \ldots, n\}$ and let $e_i \in \mathbb{R}^n$ be the i-th standard basis (column) vector. For an edge $uv \in E$, define

$$z_{uv} := e_u - e_v \in \mathbb{R}^n$$
$$Z_{uv} := z_{uv} z_{uv}^\intercal \in \mathbb{R}^{n\times n}.$$

Remark: $z_{uv} = -z_{vu}$ but $Z_{uv} = Z_{vu}$.

Definition: The *Laplacian matrix* of G is the matrix

$$L_G := \sum_{uv\in E} w_{uv} Z_{uv} \in \mathbb{R}^{n\times n}. \tag{1}$$

Fact 1: The matrix $L = L_G$ is symmetric, non-diagonals entries are $L_{ij} = -w_{ij}$, and its diagonal entries are $L_{ii} = d_i$, where $d_i = \sum_{j:ij\in E} w_{ij}$ is the degree of vertex i.

It is useful to put the degrees $\{d_i\}$ in a diagonal matrix $D = \mathrm{diag}(\vec{d})$. If G is unweighted, then $L = D - A$ where A is the adjacency matrix.

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. The exercises are for self-practice and need not be handed in. In the interest of brevity, most references and credits were omitted.

2 Basics of Symmetric Matrices

The Spectral Theorem: Every symmetric matrix $M \in \mathbb{R}^{n\times n}$ can be written as

$$M = U\Lambda U^\intercal,$$

where Λ is a diagonal matrix and U is an orthogonal matrix (i.e., $UU^\intercal = I$). This is called the *spectral decomposition* of M. Denoting the i-th column of U by $u_i \in \mathbb{R}^n$, we get that $\{u_1, \ldots, u_n\}$ is an orthonormal basis consisting of the eigenvectors of M, each associated with the eigenvalue $\lambda_i = \Lambda_{ii}$, and we can rewrite the above as

$$M = \sum_{i=1}^{n} \lambda_i u_i u_i^\intercal.$$

PSD matrices: A symmetric matrix $M \in \mathbb{R}^{n\times n}$ is called *positive semidefinite (PSD)* if it can be written as $M = BB^\intercal$. This is equivalent to requiring that all eigenvalues of M are non-negative, and also equivalent to requiring that

$$\forall x \in \mathbb{R}^n, \qquad x^\intercal M x \geq 0.$$

Exer: Show that every Symmetric Diagonally Dominant (SDD) matrix M (defined as $M_{ii} \geq \sum_{j\neq i}|M_{ij}|$ for all i) is PSD.

Fact 2: For every graph G, the Laplacian matrix L_G is PSD. Moreover, the number of nonzero eigenvalues of L_G (which is $\text{rank}(L_G)$ by basic linear algebra), is exactly n minus the number of connected components in G. Thus, G is connected if and only if L_G has $n-1$ nonzero eigenvalues.

Proof: For every $x \in \mathbb{R}^n$,

$$x^\intercal L_G x = \sum_{uv\in E} w_{uv}(x^\intercal Z_{uv} x) = \sum_{uv\in E} w_{uv}(z_{uv}^\intercal x)^2 = \sum_{uv\in E} w_{uv}(x_u - x_v)^2 \geq 0.$$

We leave the second part as an exercise, and just observe that for $x = \vec{1}$, the above expression is 0, and thus we always have an eigenvalue $\lambda = 0$, i.e., $\text{rank}(L_G) \leq n-1$.

3 Spectral Sparsifiers

Definition: A $(1\pm\varepsilon)$-*spectral sparsifier* of a graph $G = (V, E, w)$ is a graph $G' = (V, E', w')$ (on the same vertex set) such that

$$\forall x \in \mathbb{R}^n, \qquad x^\intercal L_{G'} x \in (1\pm\varepsilon)\; x^\intercal L_G x. \tag{2}$$

Theorem 3 [Spielman-Srivastava, 2008]: For every $\varepsilon \in (0, 1/2)$, every n-vertex graph $G = (V, E, w)$ has a $(1\pm\varepsilon)$-spectral sparsifier G' with $|E'| = O(\varepsilon^{-2} n \log n)$ edges. Moreover, G' is a reweighted subgraph of G, and it can be computed in randomized polynomial time (given G and ε as input).

Remarks:

(1) This theorem improves [Spielman-Teng, 2004] and [Benczur-Karger, 1996]. It was later improved by removing the $\log n$ factor in sparsity, which is the optimal bound [Batson-Spielman-Srivastava].

(2) We will focus on the existence of G'; a randomized polynomial-time algorithms is quite straightforward, and with more effort the running time can be further improved to near-linear.

(3) We assume WLOG that G is connected.

Proposition 4: Suppose G' is a $(1 \pm \varepsilon)$-spectral sparsifier of G, and denote the weight of a cut $(S, \bar{S})$ by $w(S, \bar{S}) := \sum_{uv \in E: u \in S, v \in \bar{S}} w_{uv}$ (and similarly for G'). Then

$$\forall S \subset V, \qquad w'(S, \bar{S}) \in (1 \pm \varepsilon)\ w(S, \bar{S}).$$

(Such a graph G' is usually a called a *cut sparsifier*.)

Proof: Was seen in class by considering 0-1 vectors x.

Exer: Suppose G' is a $(1 \pm \varepsilon)$-spectral sparsifier of G, and denote the eigenvalues of L_G by $\lambda_1 \geq \cdots \geq \lambda_n$, and those of L'_G by $\lambda'_1 \geq \cdots \geq \lambda'_n$. Show that

$$\forall i \in [n], \qquad \lambda'_i \in (1 \pm \varepsilon)\lambda_i.$$

Hint: use the Courant-Fischer (min-max) characterization of eigenvalues.

4 Construction of Spectral Sparsifiers

We prove Theorem 3 using the following algorithm.

Algorithm SS:

1. Init $w' = \vec{0}$ and $k := 6\varepsilon^{-2} n \ln n$

2. Viewing G as an electrical network where each edge $e \in E$ has resistance $r_e = 1/w_e$, compute for every edge $e \in E$ its effective resistance $\mathrm{R_{eff}}(e)$

3. For $i = 1, \ldots, k$

4. Pick an edge e at random with probability $p_e := \frac{w_e\, \mathrm{R_{eff}}(e)}{n-1}$

5. Increase w'_e by $\frac{1}{k}\frac{1}{p_e} w_e = \frac{n-1}{k \cdot \mathrm{R_{eff}}(e)}$

6. Output the graph defined by w', i.e., the Laplacian $L_{G'} = \sum_{e \in E} w'_e Z_e$, similarly to (1).

Observe that G' is sparse, because $E' = \{e \in E : w'_e > 0\}$ has size $|E'| \leq k$.

The next lemma shows that this algorithm (step 4) is well-defined. It requires expressing effective resistances explicitly using the Laplacian.

Lemma 5: The edge probabilities p_e sum up to 1.

Expressing effective resistances via Laplacians: Consider the electrical network corresponding to G, i.e., each edge $e \in E$ is resistor with resistance $r_e = 1/w_e$. If we fix the potentials according to some vector $\phi \in \mathbb{R}^n$, then some electrical flow (current) f will go through the resistors, and some will flow in/out of the vertices. Denote by a vector $x \in \mathbb{R}^n$ the flow injected to the vertices (opposite of the *excess flow* at each vertex). Then for every $u \in V$ (recall $d_u := \sum_{v\in N(u)} w_{uv}$),

$$\begin{aligned} x_u &= \sum_{v\in N(u)} f_{uv} && \text{(KCL)} \\ &= \sum_{v\in N(u)} \frac{\phi_u - \phi_v}{r_{uv}} && \text{(Ohm)} \\ &= d_v \cdot \phi_v - \sum_{v\in N(u)} w_{vu}\phi_u. \end{aligned}$$

In matrix notation, this is just

$$x = L_G \phi.$$

It also works in the opposite direction, i.e., if we inject flow $x \in \mathbb{R}^n$ to the vertices, then the vertex potentials will be fixed to $\phi = L_G^{-1} x$ (formally, this should be the pseudo-inverse because L_G is singular, see more below, but we will generally gloss over this issue).

Recall that the effective resistance $\mathrm{R_{eff}}(uv)$ is defined as the potential difference between $u, v \in V$ when shipping one unit of flow from u to v, i.e., injecting flow $z_{uv} = e_u - e_v$ (as the vector x). Then the vertex potentials are given by $\phi = L_G^{-1} z_{uv}$, and

$$\mathrm{R_{eff}}(uv) = \phi_u - \phi_v = (e_u - e_v)^\intercal \phi = z_{uv}^\intercal L_G^{-1} z_{uv}. \tag{3}$$

Matrix powering and pseudo-inverse: Let M be a symmetric matrix, and recall we can always write it as $M = U\Lambda U^\intercal$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Given $\alpha \in \mathbb{R}$, we can define the matrix power by essentially powering each eigenvalue separately, i.e.,

$$M^\alpha := U \,\mathrm{diag}(\lambda_1^\alpha, \ldots, \lambda_n^\alpha) U^\intercal.$$

It generalizes the usual matrix powers (for natural α), e.g., $M \cdot M = (U\Lambda U^\intercal)(U\Lambda U^\intercal) = U\Lambda^2 U^\intercal = M^2$.

For us, the really important values of α are $\{-1, 1/2, -1/2\}$. For $\alpha = -1$, the only problem is with zero eigenvalues $\lambda_i = 0$, in which case just we leave them intact (not inverting these eigenvalues). This is called the *Moore-Penrose pseudo-inverse*, denote $M^\dagger$. Observe that M and $M^\dagger$ have the same kernel.

For $\alpha = 1/2$, we basically restrict attention to PSD matrices, i.e., all $\lambda_i \geq 0$, and then there is no problem. For $\alpha = -1/2$, we combine both, i.e., restrict attention to PSD matrices (e.g., a Laplacian L_G), and power only the positive eigenvalues.

Observe that using these definitions, $(L_G^{1/2})^2 = L_G$ and that $L_G^{-1} L_G$ operates like the identity on every $x \perp \vec{1}$.

Corollary 6:

$$\forall u, v \in V, \quad \mathrm{R_{eff}}(uv) = z_{uv}^\intercal L_G^\dagger z_{uv}.$$

Cyclic property of Trace: For every matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times n}$, we have $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

The proof follows easily by expanding and changing order of summation. Alternatively, $\mathrm{Tr}(AB)$ is just the inner product of the "flattened" A with the "flattened" $B^\intercal$, and is thus the same as $\mathrm{Tr}(AB)$.

Connection to importance sampling: Lemma 7 below shows that $w_{uv}\,\mathrm{R_{eff}}(u,v)$ for an edge $uv \in E$ is precisely the maximum possible (over all x) relative contribution of this edge to $x^\intercal L_G x = \sum_{ij\in E} w_{ij}(x_i - x_j)^2$. Thus, the sampling probability p_e of an edge is proportional to its worst-case relative contribution to $x^\intercal L_G x$. (Why proportionally and not exactly? because the values $w_{uv}\,\mathrm{R_{eff}}(u,v)$ could sum up to more than 1.)

We could thus apply the importance sampling theorem with $\lambda = n-1$ for any specific $x \in \mathbb{R}^V$. However, this would still not prove Theorem 3, because the importance sampling theorem provides only weak concentration, which is not strong enough to take a union bound over all $x \in \mathbb{R}^V$.

Lemma 7:

$$\forall uv \in E, \quad \mathrm{R_{eff}}(u,v) = \max_{x\in\mathbb{R}^V} \frac{(x_u - x_v)^2}{x^\intercal L_G x}.$$

Observe that we can think of x as a vector of potentials $\phi \in \mathbb{R}^V$, and restate the lemma as an analogue of Thomson's principle (minimizing energy, but now for potentials):

$$\forall uv \in E, \quad \mathrm{R_{eff}}(u,v) = \big[\min_{\phi_u - \phi_v = 1} \phi^\intercal L_G \phi\big]^{-1}.$$

Exer: Prove Lemma 7.

Hint: Consider a minimizer ϕ. First show that every ϕ_i for $i \neq u,v$ is the weighted average of ϕ_j over its neighbors $j \in N(i)$. Then use this minimizer ϕ to define an electrical flow f, and use this flow to express each side, $\mathrm{R_{eff}}(u,v)$ and $\phi^\intercal L_G \phi$.

We will continue next class with the proofs of Lemma 5 and Theorem 3.