Spectral Sparsification by Effective Resistance Sampling

Guanghao Ye
ghye@uw.edu

Abstract

In this paper, we will discuss graph sparsification. In particular, we study the construction of spectral sparsifiers by Spielman and Srivastava [7]. In their paper, they show by sampling edges proportional to the effective resistance, they are able to get a sparsifiers with size $O(n \log n/\varepsilon^2)$ in nearly-linear time.

Instead of directly following their proof, we give a proof without argument of sampling with replacement. A key ingredient of our proof is a matrix version of Chernoff bound which we believe is of independent interest.

Contents

1 Introduction

Given a weighted graph $G = (V, E, w)$ where $n = |V|$, $m = |E|$, and $w : E \mapsto \mathbb{R}_+$, the sparsification of such graph is finding a sparser graph $H = (V, E', w')$ to approximate the original graph while preserving some characteristics. We can speed up algorithms whose runtime highly depend upon m by doing such sparsification.

The characteristic we are interested in is the spectral one, which we consider two graph is similar if their Laplacian matrices are close as linear operators. Here, we follow the notion of spectral sparsification introduced by Spielman and Teng [9]:

Definition 1.1. *The graph H is a κ-spectral approximation of G if for all $x \in \mathbb{R}^n$,*

$$x^\top L_G x \le x^\top L_H x \le \kappa x^\top L_G x,$$

where L_G and L_H are Laplacian matrices of G and H.

The main idea of this paper is constructing a random subgraph H of graph G by including each edge of G in the new graph H with probability proportional to its effective resistance. To define effective resistance, we can view the graph as a resistor network: Give a graph $G = (V, E, w)$, we replace each edge e with weights $w(e)$ as an resistor with conductance $w(e)$, i.e., a resistor with resistance $1/w(e)$. Then, the effective resistance of edge e is the resistance of this network when unit current is injected at one end of the edge and extracted at the other end. The algorithm could be stated as follows.

procedure SPARSIFY(G, q)
    Let H be a empty graph.
    for edges $e \in G$ **do**
        Add edge e to H with probability $p_e = \min(1, R_{\text{eff}}(e) w_e q)$ and weight w_e/p_e.
    end for
    return H
end procedure

We modify the weight of sampled edges in this way to preserve the Laplacian matrix in expectation. Let L_e denote the elementary Laplacian on edge e, then

$$L_G = \sum_{e \in E} w(e) L_e.$$

Thus,

$$\mathbb{E}[L_H] = \sum_{e \in E} p_e (w_e/p_e) L_e = L_G.$$

Now, we formally state the main theorem.

Theorem 1. *Given a weighted connected graph $G = (V, E, w)$ and H is the subgraph of G constructed by the way described above, L_G and L_H be the Laplacian matrix of G and H respectively, and $1/\sqrt{n} < \varepsilon \le 1$. If $q = C \log n/\varepsilon^2$, where C is the constant in section 4.3 and if n is sufficiently large, then with probability at least $1/2$*

$$\forall x \in \mathbb{R}^n \quad (1 - \varepsilon) x^\top L_G x \le x^\top L_H x \le (1 + \varepsilon) x^\top L_G x.$$

It is worth noting that L_H is also a cut sparsifier by letting $x \in \{0, 1\}^n$.

2 Preliminaries

2.1 Basic Linear Algebra

Definition 2.1. *A symmetric matrix A is positive semi-definite (PSD) if $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^n$. We write $A \succeq B$ if $A - B$ is a PSD matrix.*

Theorem 2.2. *If M_1, M_2 be two arbitrary symmetric matrices, if $M_1 \succeq M_2$, then for any matrix A, we have $A^\top M_1 A \succeq A^\top M_2 A$.*

Proof. For any vector $x \in \mathbb{R}^n$,

$$x^\top A^\top M_1 Ax - x^\top A^\top M_2 Ax = (x^\top A^\top)M_1(Ax) - (x^\top A^\top)M_2(Ax) = y^\top M_1 y - y^\top M_2 y.$$

where $y = Ax$. Since $M_1 \succeq M_2$, we have

$$y^\top M_1 y - y^\top M_2 y \geq 0.$$

Thus,

$$x^\top A^\top M_1 Ax - x^\top A^\top M_2 Ax \geq 0$$

for any $x \in \mathbb{R}^n$, which implies $A^\top M_1 A \succeq A^\top M_2 A$. □

Theorem 2.3 (Spectral Theorem)**.** *For any symmetric matrix A, there are eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, with corresponding eigenvectors $u_1, u_2, \ldots, u_n$ which are orthonormal. We can then wrtite*

$$A = \sum_{i=1}^{n} \lambda_1 u_i u_i^\top = UDU^\top,$$

where U has u_i as its ith column and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

Definition 2.4. *The trace of a square matrix A, denoted as $\operatorname{Tr}(A)$, is defined to be the sum of elements on the main diagonal of A.*

Theorem 2.5. *The trace of symmetric matrix $A \in \mathbb{R}^{n\times n}$ is equal to the sum of its eigenvalues.*

Proof. By definition of trace

$$\operatorname{Tr}(A) = \sum_{i=1}^{n} e_i^\top A e_i,$$

where e_i is the ith vector of standard basis. Using spectral theorem, we can write

$$\begin{aligned}
\operatorname{Tr}(A) &= \sum_{i=1}^{m} e_i^\top \left(\sum_{j=1}^{n} \lambda_j u_j^\top u_j \right) e_i \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j e_1^\top u_j u_j^\top e_i \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j (e_i^\top u_j)^2 \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} (e_i^\top u_j)^2 \\
&= \sum_{j=1}^{n} \lambda_j.
\end{aligned}$$

The last identity uses the fact that for any vector u_j, $\sum_{i=1}^{n}(e_i^\top u_j)^2 = \|u_j\|^2$. $\square$

2.2 The Incidence Matrix and the Laplacian

Let $G = (V, E, w)$ be a connected weighted undirected graph with n vertices and m edges and edge weights $w_e > 0$. If we orient the edges of G arbitrarily, we can write its Laplacian as $L = B^\top W B$, where $B_{m\times n}$ is the *signed edge-vertex incidence matrix*, given by

$$B(e, v) = \begin{cases} 1 & \text{if } v \text{ is } e\text{'s head} \\ -1 & \text{if } v \text{ is } e\text{'s tail} \\ 0 & \text{otherwise} \end{cases}$$

and $W_{m\times m}$ is the diagonal matrix with $W(e, e) = w_e$. It is immediate that L is a PSD matrix since:

$$\begin{aligned} x^\top L x &= x^\top B^\top W B x = \|W^{1/2} B x\|_2^2 \\ &= \sum_{(u,v)\in E} w_{u,v}(x_u - x_v)^2 \geq 0, \qquad \text{for every } x \in \mathbb{R}^n. \end{aligned}$$

and that G is connected if and only if $\ker(L) = \ker(W^{1/2}B) = \operatorname{span}(\mathbf{1})$. Then, $\operatorname{rank}(L) = n - 1$, which implies L has $n - 1$ nonzero eigenvalues.

2.3 The Pseudoinverse and Square Root of Matrix

Since L is symmetric we can diagonalize it and write

$$L = \sum_{i=1}^{n-1} \lambda_i u_i u_i^\top$$

where $\lambda_1, \ldots, \lambda_{n-1}$ are the nonzero eigenvalues of L and $u_1, \ldots, u_{n-1}$ are a corresponding set of orthonormal eigenvectors. The *Moore-Penrose Pseudoinverse* of L is then defined as

$$L^+ = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} u_i u_i^\top.$$

Notice that $\ker(L) = \ker(L^+)$ and that

$$LL^+ = L^+ L = \sum_{i=1}^{n-1} u_i u_i^\top,$$

which is simply the projection onto the span of the nonzero eigenvectors of L (which are also the eigenvectors of L^+). Thus, $LL^+ = L^+L$ is the identity on $\operatorname{im}(L) = \ker(L)^\perp$.

We can define the square root of matrix in a similar way: for a symmetric matrix L, which can be diagonalized as $L = \sum_{i=1}^{n-1} \lambda_i u_i u_i^\top$, we define its square root as

$$L^{1/2} = \sum_{i=1}^{n-1} \lambda_i^{1/2} u_i u_i^\top$$

2.4 Electrical Flow

We describe the electrical flow of graphs by the following notations: for vector $\mathbf{i}_{\text{ext}}(u)$ of currents injected at the vertices, let $\mathbf{i}(e)$ be the currents induced in the edges and $\mathbf{v}(u)$ the potential induced on the vertices. By Kirchoff's law, the currents entering a vertex is equal to the amount injected at the vertex:

$$B^\top \mathbf{i} = \mathbf{i}_{\text{ext}}.$$

By Ohm's law, the current flow in an edge is equal to potential difference across its ends times its conductance:

$$\mathbf{i} = WB\mathbf{v}.$$

Combining these two facts, we have

$$\mathbf{i}_{\text{ext}} = B^\top (WB\mathbf{v}) = L\mathbf{v}.$$

Then, if $\mathbf{i}_{\text{ext}} \perp \text{span}(\mathbf{1})$, we can write

$$\mathbf{v} = L^+ \mathbf{i}_{\text{ext}}.$$

Definition 2.6. *The effective resistance between two vertices u and v is defined as the potential difference between u and v when a unit is injected at one and extracted at another. We denote the effective resistance across edge by*

$$R_{\text{eff}}(a,b) = (\delta_a - \delta_b)^\top L_G^+ (\delta_a - \delta_b),$$

where δ_k is the unit vector which have 1 in kth coordinate.

3 Matrix Chernoff Bounds

In the original paper [7], Spielman and Srivastava use a concentration bound of Rudelson [6], which requires an argument of sampling with replacement. Here, we avoid the replacement by using the an matrix analogy of the Chernoff bound [10].

Theorem 2. *(Matrix Chernoff Bound) Let $\mathbf{X}_1, \ldots, \mathbf{X}_m$ be independent random n-dimensional symmetric positive semidefinite matrices so that $\|\mathbf{X}_i\| \le R$ 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,

$$\Pr\left[\lambda_{\min}(\sum_i \mathbf{X}_i) \le (1-\varepsilon)\mu_{\min}\right] \le n\left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\mu_{\min}/R}, \qquad \textit{for } 0 < \varepsilon < 1 \textit{ and}$$

$$\Pr\left[\lambda_{\max}(\sum_i \mathbf{X}_i) \ge (1+\varepsilon)\mu_{\max}\right] \le n\left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu_{\max}/R}, \qquad \textit{for } 0 < \varepsilon.$$

3.1 Preliminary of Matrix Analysis

Before we start the proof of Matrix Chernoff Bound, we need some basic concepts and facts.

Definition 3.1. *(Spectral mapping) Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a function. We extend f to a new function $f(A)$ on symmetric matrices by applying f to eigenvalues of A. Recall that we can diagonalize A into*

$$A = UDU^\top,$$

Define $f(A) = Uf(D)U^\top$, where $f(D)$ is the diagonal matrix with $f(D)_{i,i} = f(D_{i,i})$.

We can extends concepts of monotonicity and concavity to matrices as follows.

Definition 3.2. *A function $f : \mathbb{R} \mapsto \mathbb{R}$ is*

- **Operator monotone** *is $f(A) \succeq f(B)$ where $A \succeq B$.*
- **Operator concave** *if $f((1-\alpha)A + \alpha B) \succeq (1-\alpha)f(A) + \alpha f(B)$ for all $\alpha \in [0,1]$ and all A, B.*

It's worthy note that function f monotone does not imply that f is operator monotone. For a counter example, consider

$$f(X) = x^2, \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix},$$

where f is monotone on $\mathbb{R}_{\geq 0}$ and $A \preceq B$, but $f(A) \not\preceq f(B)$. Similarly, f concave does not imply that f is operator concave. For a counter example, consider $f(x) = -x^3$, the same matrices A and B provides a counter example with $x = 0.5$.

To get around these problems, we using the following inequalities that are known to hold.

Theorem 3.3. *Let $f : \mathbb{R} \mapsto \mathbb{R}$ and $g : \mathbb{R} \mapsto \mathbb{R}$ satisfy $f(x) \leq g(x)$ for all $x \in [l, u]$. Suppose A is symmetric and all the eigenvalues of A are in the interval $[l, u]$. Then $f(A) \preceq g(A)$.*

Proof. Let $A = UDU^\top$, then

$$g(A) - f(A) = Ug(D)U^\top - Uf(D)U^\top = U(g(D) - f(D))U^\top.$$

Since the diagonal entries of D are exactly eigenvalues of A, which falls in the interval $[l, u]$, then $g(D_{i,i}) \geq f(D_{i,i})$ for all i. Therefore, the diagonal matrix $g(D) - f(D)$ has non-negative entries on the diagonal, which implies $g(D) - f(D)$ is PSD. By Theorem 2.2, then $U(g(D) - f(D))U^\top$ is also PSD. Hence $f(A) \preceq g(A)$. □

Theorem 3.4. *If X and Y are random matrices and $X \preceq Y$, then $\mathbb{E}[X] \preceq \mathbb{E}[Y]$.*

Proof. This is a easy cor of linearity of expectation. If $X \preceq Y$, then $Y - X$ is a PSD matrix, then $\mathbb{E}[Y - X] = \mathbb{E}[Y] - \mathbb{E}[X]$ is also PSD. □

Theorem 3.5 (Weyl's Monotonicity Theorem)**.** *Suppose A and B are symmetric, $n \times n$ matrices. Let $\lambda_i(A)$ be the ith largest eigenvalue of A. If $A \preceq B$, then $\lambda_i(A) \leq \lambda_i(B)$ for all i.*

Proof. [1] We use the variational characterization of eigenvalues for symmetric matrices:

$$\lambda_i(A) = \max\{\min\{R_A(x) \mid x \in U \setminus \{0\}\} \mid U \subset V, \dim(U) = i\},$$
$$\text{where } R_A(x) = \frac{x^T A x}{x^T x}$$

To see this, consider the decomposition of $\mathbb{R}^n$ into the eigenspaces $E_1, \ldots, E_n$ of A, where $E_j = \operatorname{span}\{v_j\}$, and v_j is a unit eigenvector of A with eigenvalue $\lambda_j(A)$. By taking $U = S_i = \sum_{j=1}^{i} E_j$, we see the RHS above is $\geq R_A(v_i) = \frac{\lambda_i v_i^T v_i}{v_i^T v_i} = \lambda_i$, since v_i minimizes $R_A(x)$ for $x \in S_i$.
On the other hand, let P be the orthogonal projection onto S_i, let U be any subspace with dimension i and consider $P|_U$, the restriction of P to U. If $P|_U$ has trivial kernel, then $rank(P|_U) = dim(U) = rank(P)$, so we conclude $U = \operatorname{im}(P) = S_i$. Otherwise, say $x \in \operatorname{kernel}(P|_U), x \neq 0$. Then x is a linear combination of eigenvectors v_j with $j > i$, so

[1] I follow the proof of [2]

$$\frac{x^T A x}{x^T x} = \frac{\sum_{j=i+1}^{n} \alpha_j^2 \lambda_j}{\sum_{j=i+1}^{n} \alpha_j^2} \leq \frac{\sum_{j=i+1}^{n} \alpha_j^2 \lambda_{i+1}}{\sum_{j=i+1}^{n} \alpha_j^2} = \lambda_{i+1} \leq \lambda_i,$$

thus the minimum of R_A over U is less than or equal to the minimum of R_A over S_i, and the RHS above is $\leq \lambda_i$.

Now we prove the claim. Suppose S_A maximizes the expression $\min\{R_A(x) \mid x \in S_A \setminus \{0\}\}$ among all subspaces with dimension i, and S_B is similiarly the maximizer for B. We have:

$$\begin{aligned}
&\lambda_i(B) - \lambda_i(A) \\
&= \min\{R_B(x) \mid x \in S_B \setminus \{0\}\} - \min\{R_A(x) \mid x \in S_A \setminus \{0\}\} \\
&\geq \min\{R_B(x) \mid x \in S_A \setminus \{0\}\} - \min\{R_A(x) \mid x \in S_A \setminus \{0\}\} \\
&\overset{(1)}{\geq} \min\{R_B(x) - R_A(x) \mid x \in S_A \setminus \{0\}\} \\
&\overset{(2)}{=} \min\{R_{B-A}(x) \mid x \in S_A \setminus \{0\}\} \\
&\geq \min\{R_{B-A}(x) \mid x \in \mathbb{R}^n \setminus \{0\}\} \\
&= \lambda_n(B - A) \\
&\overset{(3)}{\geq} 0
\end{aligned}$$

To obtain (1), say x_A, x_B are the minimizers for A, B in S_A respectively. Then $R_A(x_A) \leq R_A(x_B)$, so

$$R_B(x_B) - R_A(x_A) \geq R_B(x_B) - R_A(x_B) \geq \min\{R_B(x) - R_A(x) \mid x \in S_A \setminus \{0\}\},$$

establishing (1).
For (2), note

$$R_B(x) - R_A(x) = \frac{x^T B x}{x^T x} - \frac{x^T A x}{x^T x} = \frac{x^T (B - A) x}{x^T x} = R_{B-A}(x)$$

(3) follows by the fact that $B - A$ is PSD. □

Corollary 3.6. *If f is monotone, then $\operatorname{Tr} f$ is monotone.*

Proof. This follows directly from Theorem 3.5. Suppose symmetric matrices $A \preceq B$, then

$$\operatorname{Tr} f(A) = \sum_{i=1}^{n} f(\lambda_i(A)) \leq \sum_{i=1}^{n} f(\lambda_i(B)) = \operatorname{Tr} f(B).$$

□

Theorem 3.7 (Löwner-Heinz Theorem)**.** [2] *log is operator concave.*

Another annoyance of matrix analysis is that matrix multiplication is not commutative, hence we define a commutative multiplication.

Definition 3.8. *If A, B are positive definite matrices, then we define $A \odot B = \exp(\log(A) + \log(B))$.*

[2]See Theorem 2.6 in [3] for the proof

Note that the logarithm and exponential of matrix are defined by power series, where

$$\exp(A) \equiv \sum_{n=0}^{\infty} \frac{A^n}{n!}, \qquad \log(A) \equiv \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(A-I)^n}{n}.$$

Theorem 3.9 (Lieb's Theorem [5]). *Fix any symmetric H. The map $A \mapsto \operatorname{Tr}\exp\big(\log(A) + H\big)$ is concave on positive definite matrices.*

Corollary 3.10. $\operatorname{Tr}(A \odot B)$ *is concave in A.*

Proof. By definition of $\odot$, $\operatorname{Tr}(A \odot B) = \operatorname{Tr}\exp(\log A + \log B)$, then it suffices to apply Lieb's theorem with $H = \log B$. $\square$

Corollary 3.11. *Let B be fixed, and A a random matrix. Then $\mathbb{E}[\operatorname{Tr}(A \odot B)] \le \operatorname{Tr}(\mathbb{E}[A \odot B])$.*

Proof. Since $\operatorname{Tr}(A \odot B)$ is concave in A, then apply Jensen's inequality, we have

$$\mathbb{E}[\operatorname{Tr}(A \odot B)] \le \operatorname{Tr}(\mathbb{E}[A \odot B]).$$

$\square$

Corollary 3.12. *Let $A_1, \ldots, A_k$ be independent random positive definite matrices. Then*

$$\mathbb{E}[\operatorname{Tr}(A_1 \odot \ldots \odot A_k)] \le \operatorname{Tr}(\mathbb{E}[A_1] \odot \ldots \odot \mathbb{E}[A_k])$$

Proof. This could be proved by induction and combining Corollaries 3.10 and 3.11. $\square$

3.2 Proof of Matrix Chernoff Bound

Lemma 3.13.

$$\Pr\left[\lambda_{\max}(\sum_{i=1}^{k} \mathbf{X}_i) \ge t\right] \le e^{-\theta t} \cdot \operatorname{Tr}\left[\bigodot_{i=1}^{k} \mathbb{E}[\exp(\theta \mathbf{X}_i)]\right]$$

Proof.

$$\begin{aligned}
\Pr\left[\lambda_{\max}(\sum_{i=1}^{k} \mathbf{X}_i) \ge t\right] &= \Pr\left[\lambda_{\max}(\sum_{i=1}^{k} \theta\mathbf{X}_i) \ge \theta t\right] \\
&= \Pr\left[\exp\lambda_{\max}(\sum_{i=1}^{k} \theta\mathbf{X}_i) \ge \exp(\theta t)\right] \\
&\le e^{-\theta t} \cdot \mathbb{E}[\exp\lambda_{\max}(\sum_{i=1}^{k} \theta\mathbf{X}_i)].
\end{aligned}$$

The last inequality follows by the Markov's inequality. $\square$

Notice that

$$\exp\lambda_{\max}(\sum_{i=1}^{k} \theta\mathbf{X}_i) = \lambda_{\max}(\exp\sum_{i=1}^{k} \theta\mathbf{X}_i) \le \operatorname{Tr}\left(\exp\sum_{i=1}^{k} \theta\mathbf{X}_i\right).$$

The inequality follows by Theorem 2.5 and the sum of eigenvalues is greater than the maximum eigenvalue.

Thus, we have

$$
\begin{aligned}
\Pr\left[\lambda_{\max}(\sum_{i=1}^{k}\mathbf{X}_i) \geq t\right] &\leq e^{-\theta t}\cdot \mathbb{E}[\operatorname{Tr}\exp(\sum_{i=1}^{k}\theta\mathbf{X}_i)] \\
&= e^{-\theta t}\cdot \mathbb{E}[\operatorname{Tr}\exp(\sum_{i=1}^{k}\log(\exp\theta\mathbf{X}_i))] \\
&= e^{-\theta t}\cdot \mathbb{E}[\operatorname{Tr}(e^{-\theta\mathbf{X}_1}\odot e^{-\theta\mathbf{X}_2}\odot\ldots\odot e^{-\theta\mathbf{X}_k})] \qquad \text{by definition of } \odot \\
&\leq e^{-\theta t}\cdot \operatorname{Tr}\left[\bigodot_{i=1}^{k}\mathbb{E}[\exp(\theta\mathbf{X}_i)]\right] \qquad \text{bf Corollary 3.12}
\end{aligned}
$$

Lemma 3.14. *Let X be a random symmetric matrix such that $0 \preceq \mathbf{X} \preceq I$, then*

$$\mathbb{E}[e^{\theta\mathbf{X}}] \preceq I + (e^{\theta}-1)\mathbb{E}[X].$$

Proof. For $x \in [0,1]$, $e^{\theta x} \leq 1 + (e^{\theta}-1)x$ by convexity. Since $0 \preceq \mathbf{X} \preceq I$, $\lambda_{\min}(\mathbf{X}) \geq 0$ and $\lambda_{\max}(\mathbf{X}) \leq 1$, then by Theorem 3.3,

$$e^{\theta\mathbf{X}} \preceq I + (e^{\theta}-1)\mathbf{X}.$$

Then, by Theorem 3.4,

$$\mathbb{E}[e^{\theta\mathbf{X}}] \preceq I + (e^{\theta}-1)\mathbb{E}[\mathbf{X}].$$

$\square$

Proof of Theorem 2. We only prove the upper bound. WLOG, we assume $R = 1$. By the operator concavity of the log function (Löwner-Heinz Theorem),

$$\sum_{i=1}^{k}\log\mathbb{E}[e^{\theta\mathbf{X}_i}] = k\sum_{i=1}^{k}\frac{1}{k}\log\mathbb{E}[e^{\theta\mathbf{X}_i}] \preceq \log\left(\sum_{i=1}^{k}\frac{1}{k}\mathbb{E}[e^{\theta\mathbf{X}_i}]\right). \tag{1}$$

Then, we have

$$
\begin{aligned}
&\operatorname{Tr}\left[\bigodot_{i=1}^{k}\mathbb{E}[\exp(\theta\mathbf{X}_i)]\right] \\
&= \operatorname{Tr}\left(\exp\sum_{i=1}^{k}\log(\mathbb{E}[e^{\theta\mathbf{X}_i}])\right) && \text{by definition of } \odot \\
&\leq \operatorname{Tr}\left(\exp\left(k\log\left(\sum_{i=1}^{k}\frac{1}{k}\mathbb{E}[\exp(\theta\mathbf{X}_i)]\right)\right)\right) && \text{by inequality 1 and Corollary 3.6} \\
&\leq d\cdot\lambda_{\max}\left(\exp\left(k\log\left(\sum_{i=1}^{k}\frac{1}{k}\mathbb{E}[\exp(\theta\mathbf{X}_i)]\right)\right)\right) && \text{by } \sum\lambda_i \leq d\lambda_{\max} \\
&= d\cdot\exp\left(k\log\lambda_{\max}\left(\sum_{i=1}^{k}\frac{1}{k}\mathbb{E}[\exp(\theta\mathbf{X}_i)]\right)\right) \\
&\leq d\cdot\exp\left(k\log\lambda_{\max}\left(I+(e^{\theta}-1)\sum_{i=1}^{k}\frac{1}{k}\mathbb{E}[\exp(\theta\mathbf{X}_i)]\right)\right) && \text{by 3.14 and Weyl's Monotonicity Theorem} \\
&= d\cdot\exp\left(k\log\left(1+\frac{1}{k}(e^{\theta}-1)\lambda_{\max}\left(\sum_{i=1}^{k}\mathbb{E}[\mathbf{X}_i]\right)\right)\right) \\
&\leq d\exp[(e^{\theta}-1)\mu_{\max}] && \text{by } \log(1+x)<x.
\end{aligned}
$$

Thus, by plugging the inequality above to Lemma 3.13 and let $t=(1+\delta)\mu_{\max}$ and $\theta=\log(1+\delta)$, we have

$$\Pr\left[\lambda_{\max}(\sum_{i=1}^{k}\mathbf{X}_i)\geq(1+\delta)\mu_{\max}\right]\leq d\cdot\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right).$$

□

4 The Main Result

4.1 Reduction to A Projection Matrix

Theorem 4.1. *For positive semidefinite matrices A and B, we have*

$$A\preccurlyeq(1+\varepsilon)B\Longleftrightarrow B^{-1/2}AB^{-1/2}\preccurlyeq(1+\varepsilon)I.$$

Proof. It's a easy corollary of Theorem 2.2. □

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

$$L_H\preccurlyeq(1+\varepsilon)L_G\Longleftrightarrow L_G^{+/2}L_HL_G^{+/2}\preceq(1+\varepsilon)L_G^{+/2}L_GL_G^{+/2},$$

where $L_G^{+/2}$ is the square root of the Pseudoinverse of L_G. Let

$$\Pi=L_G^{+/2}L_GL_G^{+/2},$$

which is the projection onto the range of L_G.

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

$$\mathbb{E}\left[L_G^{+/2} L_H L_G^{+/2}\right] = L_G^{+/2}(\mathbb{E}L_H)L_G^{+/2} = L_G^{+/2} L_G L_G^{+/2} = \Pi.$$

Hence, it suffices to show

$$\forall x \in \mathbb{R}^n \quad (1-\varepsilon)x^\top \Pi x \le x^\top L_G^{+/2} L_H L_G^{+/2} x \le (1+\varepsilon) x^\top \Pi x,$$

with probability at least 0.5.

Lemma 4.2 (Lemma 3 in [7])**.** *The eigenvalues of* Π *are* 1 *with multiplicity* $n-1$ *and* 0 *with multiplicity* $m-n+1$.

Proof. First, we show that $\Pi^2 = \Pi$. Notice that we can rewrite the matrix Π as $W^{1/2}BL_G^{+}B^TW^{1/2}$. Then,

$$\begin{aligned}
\Pi^2 &= (W^{1/2}BL_G^{+}B^TW^{1/2})(W^{1/2}BL_G^{+}B^TW^{1/2}) \\
&= W^{1/2}BL_G^{+}B^TWBL_G^{+}B^TW^{1/2} \\
&= W^{1/2}BL_G^{+}L_GL_G^{+}B^TW^{1/2} \\
&= W^{1/2}BL_G^{+}B^TW^{1/2} \\
&= \Pi.
\end{aligned}$$

Then, we have

$$\mathrm{im}(\Pi) = \mathrm{im}(W^{1/2}BL^{+}B^TW^{1/2}) \subseteq \mathrm{im}(W^{1/2}B).$$

To show the inclusion of another direction, assume $y \in \mathrm{im}(W^{1/2}B)$. Then we can choose $x \perp \ker(L_G)$ such that $W^{1/2}Bx = y$. Then,

$$\begin{aligned}
\Pi y &= W^{1/2}BL_G^{+}B^TW^{1/2}W^{1/2}Bx \\
&= W^{1/2}BL_G^{+}L_Gx \\
&= W^{1/2}Bx \quad x \in \mathrm{im}(L_G^{+}) \\
&= y.
\end{aligned}$$

This means $y \in \mathrm{im}(\Pi)$. Thus, we have

$$\dim(\mathrm{im}(\Pi)) = \dim(\mathrm{im}(W^{1/2}B)) = n - \dim(\ker(W^{1/2}B)).$$

Since we show $\dim(\ker(W^{1/2}B)) = 1$ in section 2.2, $\dim(\mathrm{im}(\Pi)) = n-1$. We have $\Pi^2 = \Pi$, which means eigenvalues of Π are either 1 or 0, so the eigenvalues of Π are 1 with multiplicity $n-1$ and 0 with multiplicity $m-n+1$. ☐

Now, we try to give a bound on the expected number of edges Y in the subgraph H.

4.2 Number of Edges in H

We have

$$
\begin{aligned}
\mathbb{E}[Y] &= \sum_{e \in E} p_e \\
&= \sum_{e \in E} \min(1, q w_e R_{\text{eff}}(e)) \\
&\leq \sum_{e \in E} q w_e R_{\text{eff}}(e) \\
&= q \sum_{e \in E} w_e R_{\text{eff}}(e) \\
&= q \sum_{(u,v) \in E} w_{(u,v)} (\delta_u - \delta_v)^T L_G^+ (\delta_u - \delta_v) \\
&= q \sum_{(u,v) \in E} w_{(u,v)} \operatorname{Tr}(L_G^+ (\delta_u - \delta_v)(\delta_u - \delta_v)^T) \\
&= q \operatorname{Tr} \left(\sum_{(u,v) \in E} w_{(u,v)} L_G^+ (\delta_u - \delta_v)(\delta_u - \delta_v)^T \right) \\
&= q \operatorname{Tr} \left(L_G^+ \sum_{e \in E} w_e L_e \right) \\
&= q \operatorname{Tr}(L_G^+ L_G) \\
&= q \operatorname{Tr}(\Pi) \\
&= q(n-1) \quad \text{(by Lemma 4.2)} \\
&= \frac{C \log n (n-1)}{\varepsilon^2}.
\end{aligned}
$$

This shows graph H only have $Cn \log n \varepsilon^{-2}$ many edges in the expectation. Then, we can use the Chernoff bound (for the real number) to show that the number of edges in graph H will almost be multiple of its expectation.

$$
\Pr[Y > k\mathbb{E}[Y]] \leq \exp\left(-\frac{(k-1)^2 \mathbb{E}[Y]}{k+1} \right) \leq \exp\left(\frac{Ckn \log n}{\varepsilon^2} \right)
$$

By let $k = 10$, we have $Y < 10\mathbb{E}[Y]$ with probability less than 10^{-3}.

4.3 The Analysis of Deviation

We define

$$
\mathbf{X}_e = \begin{cases} w_e/p_e L_G^{+/2} L_e L_G^{+/2} & \text{with probability } p_e \\ 0 & o.w. \end{cases}
$$

First, we show that $\mathbf{X}_e$ has small norms, so that we could apply matrix Chernoff bound later.

Similar to the equation above, we have

$$\begin{aligned}
\left\|X_{(u,v)}\right\| &= (w_{(u,v)}/p_{(u,v)}) \left\|L_G^{+/2} L_{(u,v)} L_G^{+/2}\right\| \\
&= (w_{(u,v)}/p_{(u,v)}) \operatorname{Tr}(L_G^{+}(\delta_u - \delta_v)(\delta_u - \delta_v)^\top) \\
&= \frac{w_{(u,v)} R_{\mathrm{eff}}(u,v)}{p_{(u,v)}}.
\end{aligned}$$

Recall that we define

$$p_e = \min(1, q w_e R_{\mathrm{eff}}(e)),$$

we have

$$\left\|X_{(u,v)}\right\| \leq q^{-1} = \frac{\varepsilon^2}{C \log n}. \tag{2}$$

It is easy to find that

$$\mathbb{E}\left[\sum_{e \in E} X_e\right] = \sum_{e \in E} \mathbb{E}[X_e] = \Pi.$$

Before using the matrix Chernoff bound, we use the following approximation to make life easier.

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

Then, we can obtain the following bound by combining matrix Chernoff bound (Theorem 2) and the approximation above:

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

Then, using the bound of inequality 2, we get

$$\Pr[\sum_{e \in E} X_e \succeq (1+\varepsilon)\Pi] \leq n \exp(-\varepsilon^2 C \varepsilon^{-2} \log n/3) = n^{-C/3+1},$$

and

$$\Pr[\sum_{e \in E} X_e \preceq (1-\varepsilon)\Pi] \leq n \exp(-\varepsilon^2 C \varepsilon^{-2} \log n/2) = n^{-C/2+1}.$$

We can let $C = 20$, which gives us probability less than 10^{-4} for both sides.

For the edges which have $p_e = 1$, we can deal with them by split the corresponding matrix into K copies for some large K. This will not change the expectation, and since each copy X_e/K has a small norm, we can apply the matrix Chernoff bound again.

5 Computing the Resistance

Here, we show how to compute the approximate effective resistance in nearly-linear time by using Spielman and Tengs's fast Laplacian solvers [8] and the Johnson-Lindenstrauss dimension reduction [4].

Theorem 3 (theorem 2 in [7])**.** *There is an $\widetilde{O}(m(\log r)\varepsilon^{-2})$ time algorithm which on input $\varepsilon > 0$ and $G = (V, E, w)$ with $r = w_{\max}/w_{\min}$ computes a $(24 \log n\varepsilon^{-2}) \times n$ matrix $\widetilde{Z}$ such that with probability at least $1 - 1/n$*

$$(1-\varepsilon)R_{\text{eff}}(u,v) \leq \left\|\widetilde{Z}(\delta_u - \delta_v)\right\|^2 \leq (1+\varepsilon)R_{\text{eff}}(u,v)$$

for every $(u, v) \in V \times V$.

Notice that we can rewrite the effective resistance as

$$R_{\text{eff}}(u,v) = (\delta_u - \delta_v)^T L_G^+(\delta_u - \delta_v).$$

Then, we have

$$\begin{aligned} R_{\text{eff}}(u,v) &= (\delta_u - \delta_v)^T L_G^+(\delta_u - \delta_v) \\ &= (\delta_u - \delta_v)^T L_G^+ L_G L_G^+(\delta_u - \delta_v) \\ &= (\delta_u - \delta_v)^T L_G^+(B^T W B) L_G^+(\delta_u - \delta_v) \\ &= ((\delta_u - \delta_v)^T L_G^+ B^T W^{1/2})(W^{1/2} B L_G^+(\delta_u - \delta_v)) \\ &= \left\|W^{1/2} B L_G^+(\delta_u - \delta_v)\right\|_2^2 \end{aligned}$$

This shows we can treat effective resistance as square of the pairwise Euclidean distance after do the projection into $W^{1/2}BL_G^+$. By Johnson-Lindenstrauss Lemma, we can preserve the pairwise Euclidean distance by projecting these vector into a subspace spanned by $O(\log n)$ random vectors. Here, we using a modified version of Johnson-Lindenstrauss Lemma [1].

Lemma 5.1. *Given fixed vectors $v_1, \ldots, v_n \in \mathbb{R}^d$ and $\varepsilon > 0$, let $Q_{k\times d}$ be a random $\pm 1/\sqrt{k}$ matrix (i.e. independent Bernoulli entries) with $k > 24 \log n\varepsilon^{-2}$. Then with probability at least $1 - 1/n$*

$$(1-\varepsilon)\left\|v_i - v_j\right\|_2^2 \leq \left\|Qv_i - Qv_j\right\|_2^2 \leq (1+\varepsilon)\left\|v_i - v_j\right\|_2^2$$

for all pairs $i, j \leq n$.

Then, instead of computing the projection matrix, we are interested in computing the project $\{QW^{1/2}BL_G^+\delta_v\}$. Actually, we will use the fast Laplacian solver of Spielman and Teng [8] to find the projection.

Theorem 5.2. *There is an algorithm $x = \mathsf{STSolve}(L, y, \delta)$ which takes a Laplacian matrix L, a column vector y, and an error parameter $\delta > 0$, and returns a column vector x satisfying*

$$\left\|x - L^+ y\right\|_L \leq \varepsilon \left\|L^+ y\right\|_L,$$

where $\|y\|_L = \sqrt{y^T L y}$. The algorithm runs in expected time $\widetilde{O}(\text{nnz}(L)\log(1/\delta))$, where $\text{nnz}(L)$ is the number of nonzero entries of L.

Let $Z = QW^{1/2}BL_G^+$, we compute its approximation $\widetilde{Z}$ by using $\mathsf{STSolve}$ to approximate row of Z. For matrix X, we use x_i to denote the ith row of matrix X. Now, we construct the matrix $\widetilde{Z}$ in the following steps:

1. Let Q be a random $\pm 1/\sqrt{k}$ matrix of dimension $k \times n$ where $k = 24 \log n\varepsilon^{-2}$.
2. Compute $Y = QW^{1/2}B$.
3. For $1 \le i \le k$, compute $\tilde{z}_i = \mathsf{STSolve}(L, y_i, \delta)$.

[3]It suffices to call STSolve with

$$\delta = \frac{\varepsilon}{3}\sqrt{\frac{2(1-\varepsilon)w_{\min}}{(1+\varepsilon)n^3 w_{\max}}}.$$

Note the second steps only takes $2m \times 24 \log n\varepsilon^{-2} + m$ time, since B only has $2m$ entries and W is an diagonal matrix.

Thus, the construction of $\widetilde{Z}$ takes $\widetilde{O}(m \log(1/\delta)\varepsilon^{-2}) = \widetilde{O}(m \log r\varepsilon^{-2})$. Then,to query the approximate effective resistance, we only need to subtract the corresponding column in $O(\log n\varepsilon^{-2})$ time.

References

[1] Dimitris Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, June 2003.

[2] Rajendra Bhatia. *Matrix Analysis*. Springer, 1997.

[3] Eric Carlen. Trace inequalities and quantum entropy: An introductory course. 529, 01 2010.

[4] William Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemp. Math.*, 26:189–206, 1984.

[5] Elliott Lieb. Convex trace functions and the wigner-yanase-dyson conjecture. *Advances in Mathematics*, 11(3):267–288, 1 1973.

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

[7] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40:1913–1926, 2011.

[8] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing*, STOC '04, pages 81–90, New York, NY, USA, 2004. ACM.

[9] Daniel A. Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM J. Comput.*, 40:981–1025, 2011.

[10] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12:389–434, 2012.

[3]See Lemma 9 in [7] for the proof