Abstract

We present the first almost-linear time algorithm for constructing linear-sized spectral sparsification for graphs. This improves all previous constructions of linear-sized spectral sparsification, which requires $\Omega(n^2)$ time [1], [2], [3]. A key ingredient in our algorithm is a novel combination of two techniques used in literature for constructing spectral sparsification: Random sampling by effective resistance [4], and adaptive constructions based on barrier functions [1], [3].

Keywords

algorithmic spectral graph theory; spectral sparsification

I. INTRODUCTION

Graph sparsification is the procedure of approximating a graph $G$ by a sparse graph $G'$ such that certain quantities between $G$ and $G'$ are preserved. For instance, spanners are defined between two graphs in which the distances between any pair of vertices in these two graphs are approximately the same [5]; cut sparsifiers are reweighted sparse graphs of the original graphs such that the weights of every cut between the sparsifiers and the original graphs are approximately the same [6]. Since both storing and processing large-scale graphs are expensive, graph sparsification is one of the most fundamental building blocks in designing fast graph algorithms, including solving Laplacian systems [7], [8], [9], [10], [11], [12], designing approximation algorithms for the maximum flow problem [6], [13], [14], and solving streaming problems [15], [16]. Beyond graph problems, techniques developed for spectral sparsification are widely used in randomized linear algebra [17], [18], [19], sparsifying linear programs [20], and various pure mathematics problems [21], [22], [23], [24].

In this work, we study spectral sparsification introduced by Spielman and Teng [25]: A spectral sparsifier is a reweighted sparse subgraph of the original graph such that, for all real vectors, the Laplacian quadratic forms between that subgraph and the original graph are approximately the same [6]. Formally, for any undirected and weighted graph $G = (V, E, w)$ with $n$ vertices and $m$ edges, we call a subgraph $G'$ of $G$, with proper reweighting of the edges, a $(1 + \varepsilon)$-spectral sparsifier if it holds for any $x \in \mathbb{R}^n$ that

$$(1 - \varepsilon) x^T L_G x \leq x^T L_{G'} x \leq (1 + \varepsilon) x^T L_G x,$$

where $L_G$ and $L_{G'}$ are the respective graph Laplacian matrices of $G$ and $G'$.

Spielman and Teng [25] presented the first algorithm for constructing spectral sparsification. For any undirected graph $G$ of $n$ vertices, their algorithm runs in $O(n \log^c n/\varepsilon^2)$ time, for some big constant $c$, and produces a spectral sparsifier with $O(n \log^c n/\varepsilon^2)$ edges for some $c' \geq 2$. Since then, there has been a wealth of work on spectral sparsification. For instance, Spielman and Srivastava [4] presented a nearly-linear time algorithm for constructing a spectral sparsifier of $O(n \log n/\varepsilon^2)$ edges. Batson, Spielman and Srivastava [1] presented an algorithm for constructing spectral sparsifiers with $O(n/\varepsilon^2)$ edges, which is optimal up to a constant. However,
all previous constructions either require $\Omega(n^{2+\varepsilon})$ time in order to produce linear-sized sparsifiers [1], [2], [3], or $O(n \log^{O(1)} n/\varepsilon^2)$ time but the number of edges in the sparsifiers is sub-optimal.

In this paper we present the first almost-linear time algorithm for constructing linear-sized spectral sparsification for graphs. Our result is summarized as follows:

**Theorem I.1.** Given any integer $q \geq 10$ and $0 < \varepsilon \leq 1/120$. Let $G = (V, E, w)$ be an undirected and weighted graph with $n$ vertices and $m$ edges. Then, there is an algorithm that outputs a $(1 + \varepsilon)$-spectral sparsifier of $G$ with $O\left(\frac{qn}{\varepsilon^2} \frac{1}{\varepsilon}\right)$ edges. The algorithm runs in $\tilde{O}\left(\frac{m \cdot n^{1+1/q}}{\varepsilon^2}\right)$ time.

Graph sparsification is known as a special case of sparsifying sums of rank-1 positive semi-definite (PSD) matrices [1], [4], and our algorithm works in this general setting as well. Our result is summarized as follows:

**Theorem I.2.** Given any integer $q \geq 10$ and $0 < \varepsilon \leq 1/120$. Let $I = \sum_{i=1}^{m} v_i v_i^\top$ be the sum of $m$ rank-1 PSD matrices. Then, there is an algorithm that outputs scalars $\{s_i\}_{i=1}^{m}$ with $|\{s_i : s_i \neq 0\}| = O\left(\frac{qn}{\varepsilon^2}\right)$ such that

$$(1 - \varepsilon) \cdot I \preceq \sum_{i=1}^{m} s_i v_i v_i^\top \preceq (1 + \varepsilon) \cdot I.$$ 

The algorithm runs in $\tilde{O}\left(\frac{m \cdot n^{1+1/q}}{\varepsilon^2}\right)$ time, where $\omega$ is the matrix-multiplication constant.

A key ingredient in our algorithm is a novel combination of two techniques used in literature for constructing spectral sparsification: Random sampling by effective resistance of edges [4], and adaptive construction based on barrier functions [1], [3]. We will present an overview of the algorithm, and the intuitions behind it in Section II.

**Preliminaries:** Let $G = (V, E, w)$ be a connected, undirected and weighted graph with $n$ vertices and $m$ edges, and weight function $w : V \times V \rightarrow \mathbb{R}_{\geq 0}$. The Laplacian matrix of $G$ is an $n$ by $n$ matrix $L$ defined by

$$L_G(u, v) = \begin{cases} -w(u, v) & \text{if } u \sim v, \\ \deg(u) & \text{if } u = v, \\ 0 & \text{otherwise,} \end{cases}$$

where $\deg(u) = \sum_{v \sim u} w(u, v)$. It is easy to see that

$$x^\top L_G x = \sum_{u \sim v} w_{u,v} (x_u - x_v)^2 \geq 0,$$

for any $x \in \mathbb{R}^n$.

For any matrix $A$, let $\lambda_{\text{max}}(A)$ and $\lambda_{\text{min}}(A)$ be the maximum and minimum eigenvalues of $A$. The condition number of matrix $A$ is defined by $\lambda_{\text{max}}(A)/\lambda_{\text{min}}(A)$. For any two matrices $A$ and $B$, we write $A \preceq B$ to represent $B - A$ is positive semi-definite (PSD), and $A \prec B$ to represent $B - A$ is positive definite. For any two matrices $A$ and $B$ of equal dimensions, let $A \bullet B \triangleq \text{tr}(A^\top B)$. For any function $f$, we write $\tilde{O}(f) \triangleq O(f \cdot \log^{O(1)} f)$. For matrices $A$ and $B$, we write $A \approx_\varepsilon B$ if $(1 - \varepsilon) \cdot A \preceq B \preceq (1 + \varepsilon) A$.

**II. Algorithm**

We study the algorithm of sparsifying the sum of rank-1 PSD matrices in this section. Our goal is to, for any vectors $v_1, \cdots, v_m$ with $\sum_{i=1}^{m} v_i v_i^\top = I$, find scalars $\{s_i\}_{i=1}^{m}$ satisfying

$$|\{s_i : s_i \neq 0\}| = O\left(\frac{qn}{\varepsilon^2}\right),$$

such that

$$(1 - \varepsilon) \cdot I \preceq \sum_{i=1}^{m} s_i v_i v_i^\top \preceq (1 + \varepsilon) \cdot I.$$ 

We will use this algorithm to construct graph sparsifiers in Section III.
A. Overview of Our Approach

Our construction is based on a probabilistic view of the algorithm presented in Batson et al. [1]. We refer their algorithm BSS for short, and give a brief overview of the BSS algorithm at first.

At a high level, the BSS algorithm proceeds by iterations, and adds a rank-1 matrix $c \cdot v_i v_i^\top$ with some scaling factor $c$ to the currently constructed matrix $A_j$ in iteration $j$. To control the spectral properties of matrix $A_j$, the algorithm maintains two barrier values $u_j$ and $\ell_j$, and initially $u_0 > 0$, $\ell_0 < 0$. It was proven that one can always find a vector in $\{v_i\}_{i=1}^n$ and update $u_j, \ell_j$ in a proper manner in each iteration, such that the invariant

$$\ell_j I < A_j < u_j I$$  \hspace{1cm} (1)

always holds, [1]. To guarantee this, Batson et al. [1] introduces a potential function

$$\Phi_{u,\ell}(A) \triangleq \text{tr}(uI - A)^{-1} + \text{tr}(A - \ell I)^{-1}$$  \hspace{1cm} (2)

to measure “how far the eigenvalues of $A$ are from the barriers $u$ and $\ell$”, since a small value of $\Phi_{u,\ell}(A)$ implies that no eigenvalue of $A$ is close to $u$ or $\ell$. With the help of the potential function, it was proven that, after $k = \Theta\left(\frac{n}{\varepsilon^2}\right)$ iterations, it holds that $\ell_k \geq cu_k$ for some constant $c$, implying that the resulting matrix $A_k$ is a linear-sized and $A_k \approx_{O(\varepsilon)} I$.

The original BSS algorithm is deterministic, and in each iteration the algorithm finds a rank-1 matrix which maximizes certain quantities. To informally explain our algorithm, let us look at the following randomized variant of the BSS algorithm: In each iteration, we choose a vector $v_i$ with probability $p_i$, and add a rank-1 matrix

$$\Delta_A \triangleq \frac{\varepsilon}{\ell} \cdot \frac{1}{p_i} \cdot v_i v_i^\top$$

to the current matrix $A$. See Algorithm 1 for formal description.

\begin{algorithm}
\begin{algorithmic}
1: $j = 0$;
2: $\ell_0 = -8n/\varepsilon$, $u_0 = 8n/\varepsilon$;
3: $A_0 = 0$;
4: \textbf{while} $u_j - \ell_j < 8n/\varepsilon$ \textbf{do}
5: \hspace{1cm} Let $t = \text{tr}\left((u_j I - A_j)^{-1} + (A_j - \ell_j I)^{-1}\right)$;
6: \hspace{1cm} Sample a vector $v_i$ with probability $p_i \triangleq \left(\frac{v_i^\top (u_j I - A_j)^{-1} v_i + v_i^\top (A_j - \ell_j I)^{-1} v_i}{t}\right)$;
7: \hspace{1cm} $A_{j+1} = A_j + \frac{\varepsilon}{t} \cdot \frac{1}{p_i} \cdot v_i v_i^\top$;
8: \hspace{1cm} $u_{j+1} = u_j + \frac{\varepsilon}{\ell(1-\varepsilon)}$ and $\ell_{j+1} = \ell_j + \frac{\varepsilon}{\ell(1+\varepsilon)}$;
9: \hspace{1cm} $j \leftarrow j + 1$;
10: \textbf{Return} $A_j$;
\end{algorithmic}
\end{algorithm}

Let us look at any fixed iteration $j$, and analyze how the added $\Delta_A$ impacts the potential function. We drop the subscript representing the iteration $j$ for simplicity. After adding $\Delta_A$, the first-order approximation of $\Phi_{u,\ell}(A)$ gives that

$$\Phi_{u,\ell}(A + \Delta_A) \sim \Phi_{u,\ell}(A) + (uI - A)^{-2} \cdot \Delta_A - (A - \ell I)^{-2} \cdot \Delta_A.$$  \hspace{1cm} (3)

Since

$$\mathbb{E} [\Delta_A] = \sum_{i=1}^m p_i \cdot \left(\frac{\varepsilon}{t} \cdot \frac{1}{p_i} \cdot v_i v_i^\top\right) = \frac{\varepsilon}{t} \cdot \sum_{i=1}^m v_i v_i^\top = \frac{\varepsilon}{t} I,$$
we have that

$$
\mathbb{E}[\Phi_{u,\ell}(A + \Delta_A)] \sim \Phi_{u,\ell}(A) + \frac{\epsilon}{t} \cdot (uI - A)^{-2} \cdot I - \frac{\epsilon}{t} \cdot (A - \ell I)^{-2} \cdot I
$$

$$
= \Phi_{u,\ell}(A) + \frac{\epsilon}{t} \cdot \text{tr}((uI - A)^{-2} - \frac{\epsilon}{t} \cdot \text{tr}(A - \ell I)^{-2})
$$

$$
= \Phi_{u,\ell}(A) - \frac{\epsilon}{t} \cdot \frac{d}{du}\Phi_{u,\ell}(A) - \frac{\epsilon}{t} \cdot \frac{d}{d\ell}\Phi_{u,\ell}(A).
$$

Notice that if we increase $u$ by $\frac{\epsilon}{\ell}$ and $\ell$ by $\frac{\epsilon}{\ell}$, $\Phi_{u,\ell}$ approximately increases by

$$
\frac{\epsilon}{t} \cdot \frac{d}{du}\Phi_{u,\ell}(A) + \frac{\epsilon}{\ell} \cdot \frac{d}{d\ell}\Phi_{u,\ell}(A).
$$

Hence, comparing $\Phi_{u+\epsilon/t,\ell+\epsilon/\ell}(A + \Delta_A)$ with $\Phi_{u,\ell}(A)$, the increase of the potential function due to the change of barrier values is approximately compensated by the drop of the potential function by the effect of $\Delta_A$. For a more rigorous analysis, we need to look at the higher-order terms and increase $u$ slightly more than $\ell$ to compensate that. Batson et al. [1] gives the following estimate:

**Lemma II.1** ([1], proof of Lemma 3.3 and 3.4). Let $A \in \mathbb{R}^{n \times n}$, and $u, \ell$ be parameters satisfying $\ell I \prec A \prec uI$. Suppose that $w \in \mathbb{R}^n$ satisfies $ww^\top \preceq \delta(uI - A)$ and $ww^\top \preceq \delta(A - \ell I)$ for some $0 < \delta < 1$. Then, it holds that

$$
\Phi_{u,\ell}(A + ww^\top) \leq \Phi_{u,\ell}(A) + \frac{w^\top(uI - A)^{-2}w}{1 - \delta} - \frac{w^\top(A - \ell I)^{-2}w}{1 + \delta}.
$$

The estimate above shows that the first-order approximation (3) is good if $ww^\top \preceq \delta(uI - A)$ and $ww^\top \preceq \delta(A - \ell I)$ for small $\delta$. It is easy to check that, by setting $\delta = \epsilon$, the added matrix $\Delta_A$ satisfies these two conditions, since

$$
\frac{\epsilon}{t} \cdot \frac{1}{p_i} \cdot v_i v_i^\top = \frac{\epsilon}{t} \cdot v_i v_i^\top \leq \frac{\epsilon}{t} \cdot (uI - A)^{-1} v_i \leq \frac{\epsilon}{t} \cdot (A - \ell I)^{-1} v_i \leq \epsilon(uI - A),
$$

where we used the fact that $vv^\top \preceq (v^\top B^{-1} v)B$ for any vector $v$ and PSD matrix $B$. Similarly, we have that

$$
\frac{\epsilon}{t} \cdot \frac{1}{p_i} \cdot v_i v_i^\top \leq \epsilon(A - \ell I).
$$

Hence, if $\Phi_{u,\ell}(A)$ is small initially, our crude calculations above gives a good approximation and $\Phi_{u,\ell}(A)$ is small throughout the executions of the whole algorithm. Up to a constant factor, this gives the same result as [1], and therefore Algorithm 1 constructs an $\Theta(n^{1/2})$-sized $(1 + O(\epsilon))$-spectral sparsifier.

Our algorithm follows the same framework as Algorithm 1. However, to construct a spectral sparsifier in almost-linear time, we expect that the sampling probability $\{p_i\}_{i=1}^m$ of vectors (i) can be approximately computed fast, and (ii) can be further “reused” for a few iterations.

For fast approximation of the sampling probabilities, we adopt the idea proposed in [3]: Instead of defining the potential function by (2), we define the potential function by

$$
\Phi_{u,\ell}(A) \triangleq \text{tr}(uI - A)^{-q} + \text{tr}(A - \ell I)^{-q}.
$$

Since $q$ is a large constant, the value of the potential function becomes larger when some eigenvalue of $A$ is close to $u$ or $\ell$. Hence, a bounded value of $\Phi_{u,\ell}(A)$ insures that the eigenvalues of $A$ never get too close to $u$ or $\ell$, which further allows us to compute the sampling probabilities $\{p_i\}_{i=1}^m$ efficiently simply by Taylor expansion. Moreover, by defining the potential function based on $\text{tr}(\cdot)^{-q}$, one can prove a similar result as Lemma II.1. This gives an alternative analysis of the algorithm presented in [3], which is the first almost-quadratic time algorithm for constructing linear-sized spectral sparsifiers.

To “reuse” the sampling probabilities, we re-compute $\{p_i\}_{i=1}^m$ after every $\Theta(n^{1-1/q})$ iterations: We show that
as long as the sampling probability satisfies
\[ p_i \geq C \cdot \frac{v_i^T (uI - A)^{-1} v_i + v_i^T (A - \ell I)^{-1} v_i}{\sum_{i=1}^{m} (v_i^T (uI - A)^{-1} v_i + v_i^T (A - \ell I)^{-1} v_i)} \]
for some constant \( C > 0 \), we can still sample \( v_i \) with probability \( p_i \) and get the same guarantee on the potential function. The reason is as follows: Assume that \( \Delta_A = \sum_{i=1}^{T} \Delta_{A,i} \) is the sum of the sampled matrices within \( T = O(n^{1-1/\varepsilon}) \) iterations. If a randomly chosen matrix \( \Delta_{A,i} \) satisfies \( \Delta_{A,i} \preceq \frac{1}{Cq} (uI - A) \), then by the matrix Chernoff bound \( \Delta_A \preceq \frac{1}{2} (uI - A) \) holds with high probability. By scaling every sampled rank-1 matrix \( q \) times smaller, the sampling probability only changes by a constant factor within \( T \) iterations. Since we choose \( \Theta(n/\varepsilon^2) \) vectors in total, our algorithm only recomputes the sampling probabilities \( \Theta\left(n^{1/\varepsilon}/\varepsilon^2\right) \) times. Hence, our algorithm runs in almost-linear time if \( q \) is a large constant.

B. Algorithm Description

The algorithm follows the same framework as Algorithm 1, and proceeds by iterations. Initially, the algorithm sets
\[
\begin{align*}
u_0 & \triangleq (2n)^{1/q}, \\
\ell_0 & \triangleq -2n)^{1/q}, \\
A_0 & \triangleq 0.
\end{align*}
\]
After iteration \( j \) the algorithm updates \( u_j, \ell_j \) by \( \Delta_{u,j}, \Delta_{\ell,j} \) respectively, i.e.,
\[
\begin{align*}
u_{j+1} & \triangleq u_j + \Delta_{u,j}, \\
\ell_{j+1} & \triangleq \ell_j + \Delta_{\ell,j},
\end{align*}
\]
and updates \( A_j \) with respect to the chosen matrix in iteration \( j \). The choice of \( \Delta_{u,j} \) and \( \Delta_{\ell,j} \) insures that
\[
\ell_j I < A_j < u_j I
\]
holds for any \( j \). In iteration \( j \), the algorithm computes the relative effective resistance of vectors \( \{v_i\}_{i=1}^{m} \) defined by
\[
R_i(A_j, u_j, \ell_j) \triangleq v_i^T (u_j I - A_j)^{-1} v_i + v_i^T (A_j - \ell_j I)^{-1} v_i,
\]
and samples \( N_j \) vectors independently with replacement, where vector \( v_i \) is chosen with probability proportional to \( R_i(A_j, u_j, \ell_j) \), and
\[
N_j \triangleq \frac{1}{n^{2/q}} \left( \sum_{i=1}^{m} R_i(A_j, u_j, \ell_j) \right) \min \{\lambda_{\min}(u_j I - A_j), \lambda_{\min}(A_j - \ell_j I)\}.
\]
The algorithm sets \( A_{j+1} \) to be the sum of \( A_j \) and sampled \( v_i v_i^T \) with proper reweighting. For technical reasons, we define \( \Delta_{u,j} \) and \( \Delta_{\ell,j} \) by
\[
\begin{align*}
\Delta_{u,j} & \triangleq (1 + 2\varepsilon) \cdot \frac{\varepsilon \cdot N_j}{q \cdot \sum_{i=1}^{m} R_i(A_j, u_j, \ell_j)}, \\
\Delta_{\ell,j} & \triangleq (1 - 2\varepsilon) \cdot \frac{\varepsilon \cdot N_j}{q \cdot \sum_{i=1}^{m} R_i(A_j, u_j, \ell_j)}.
\end{align*}
\]
See Algorithm 2 for formal description.

We remark that, although exact values of \( N_j \) and relative effective resistances are difficult to compute in almost-linear time, we can use approximated values of \( R_i \) and \( N_j \) instead. It is easy to see that in each iteration an over estimate of \( R_i \), and an under estimate of \( N_j \) with constant-factor approximation suffice for our purpose.

III. Analysis

We analyze Algorithm 2 in this section. To make the calculation less messy, we assume the following:

Assumption III.1. We always assume that \( 0 < \varepsilon \leq 1/120 \), and \( q \) is an integer satisfying \( q \geq 10 \).
Algorithm 2 Algorithm for constructing spectral sparsifiers

Require: $\epsilon \leq 1/120$, $q \geq 10$
1: $j = 0$;
2: $\ell_0 = -(2n)^{1/q}$, $u_0 = (2n)^{1/q}$, $A_0 = 0$;
3: while $u_j - \ell_j < 4 \cdot (2n)^{1/q}$ do
4: \quad $W_j = 0$;
5: \quad Compute $R_i(A_j, u_j, \ell_j)$ for all vectors $v_i$;
6: \quad Sample $N_j$ vectors independently with replacement, where every $v_i$ is chosen with probability proportional to $R_i(A_j, u_j, \ell_j)$. For every sampled $v$, add $\epsilon/q \cdot (R_i(A_j, u_j, \ell_j))^{-1} \cdot vv^\top$ to $W_j$;
7: \quad $A_{j+1} = A_j + W_j$;
8: \quad $u_{j+1} = u_j + \Delta_{u,j}$, $\ell_{j+1} = \ell_j + \Delta_{\ell,j}$;
9: \quad $j = j + 1$;
10: Return $A_j$;

Our analysis is based on a potential function $\Phi_{u,\ell}$ with barrier values $u, \ell \in \mathbb{R}$. Formally, for a symmetric matrix $A \in \mathbb{R}^{n \times n}$ with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and parameters $u, \ell$ satisfying $\ell I < A < uI$, let

$$
\Phi_{u,\ell}(A) \triangleq \text{tr}(uI - A)^{-q} + \text{tr}(A - \ell I)^{-q} = \sum_{i=1}^{n} \left( \frac{1}{u - \lambda_i} \right)^q + \sum_{i=1}^{n} \left( \frac{1}{\lambda_i - \ell} \right)^q.
$$

(4)

We will show how the potential function evolves after each iteration in Section III-A. Combining this with the ending condition of the algorithm, we will prove in Section III-B that the algorithm outputs a linear-sized spectral sparsifier. We will prove Theorem I.1 and Theorem I.2 in Section III-C.

A. Analysis of a Single Iteration

We analyze the sampling scheme within a single iteration, and drop the subscript representing the iteration $j$ for simplicity. Recall that in each iteration the algorithm samples $N$ vectors independently from $\mathcal{V} = \{v_i\}_{i=1}^{m}$ satisfying $\sum_{i=1}^{m} v_i v_i^\top = I$, where every vector $v_i$ is sampled with probability $\frac{R_i(A, u, \ell)}{\sum_{j=1}^{N} R_j(A, u, \ell)}$. We use $v_1, \cdots, v_N$ to denote these $N$ sampled vectors, and define the reweighted vectors by

$$
w_i \triangleq \sqrt{\frac{\epsilon}{q \cdot R_i(A, u, \ell)} \cdot v_i},
$$

for any $1 \leq i \leq N$. Let

$$
W \triangleq \sum_{i=1}^{N} w_i w_i^\top,
$$

and we use $W \sim \mathcal{D}(A, u, \ell)$ to represent that $W$ is sampled in this way with parameters $A, u$ and $\ell$. We will show that with high probability matrix $W$ satisfies $0 \leq W \leq \frac{1}{2}(uI - A)$. We first recall the following Matrix Chernoff Bound.

Lemma III.2 (Matrix Chernoff Bound, [26]). Let $\{X_k\}$ be a finite sequence of independent, random, and self-adjoint matrices with dimension $n$. Assume that each random matrix satisfies $X_k \succeq 0$, and $\lambda_{\max}(X_k) \leq D$. Let $\mu \geq \lambda_{\max}(\sum_k \mathbb{E}[X_k])$. Then, it holds for any $\delta \geq 0$ that

$$
\mathbb{P} \left[ \lambda_{\max} \left( \sum_k X_k \right) \geq (1 + \delta)\mu \right] \leq n \cdot \left( \frac{e^{\delta}}{(1 + \delta)^{1+\delta}} \right)^{\mu/D}.
$$
Lemma III.3. Assume that the number of samples satisfies

\[ N < \frac{2}{q^{2/2}} \left( \sum_{i=1}^{m} R_i(A, u, \ell) \right) \cdot \lambda_{\min}(uI - A). \]

Then, it holds that

\[ \mathbb{E}[W] = \frac{\varepsilon}{q} \cdot \frac{N}{\sum_{i=1}^{m} R_i(A, u, \ell)} \cdot I, \]

and

\[ \mathbb{P}\left[ 0 \leq W \leq \frac{1}{2} \cdot (uI - A) \right] \geq 1 - \frac{\varepsilon^2}{100qn}. \]

Proof: By the description of the sampling procedure, it holds that

\[ \mathbb{E}[w_i w_i^\top] = \sum_{j=1}^{m} \sum_{t=1}^{m} \frac{R_j(A, u, \ell) \cdot \varepsilon}{q} \cdot \frac{v_j v_j^\top}{R_j(A, u, \ell)} = \frac{\varepsilon}{q} \cdot \frac{1}{\sum_{t=1}^{m} R_t(A, u, \ell)} \cdot I, \]

and

\[ \mathbb{E}[W] = \mathbb{E}\left[ \sum_{i=1}^{N} w_i w_i^\top \right] = \frac{\varepsilon}{q} \cdot \frac{N}{\sum_{i=1}^{m} R_i(A, u, \ell)} \cdot I, \]

which proves the first statement.

Now for the second statement. Let

\[ z_i = (uI - A)^{-1/2}w_i. \]

It holds that

\[ \text{tr}(z_i z_i^\top) = \text{tr}\left((uI - A)^{-1/2}w_i w_i^\top(uI - A)^{-1/2}\right) \]

\[ = \frac{\varepsilon}{q} \cdot \text{tr}\left((uI - A)^{-1/2}v_i v_i^\top(uI - A)^{-1/2}\right) \]

\[ \leq \frac{\varepsilon}{q} \cdot v_i^\top(uI - A)^{-1}v_i \]

\[ \leq \frac{\varepsilon}{q}, \]

and \( \lambda_{\max}(z_i z_i^\top) \leq \frac{\varepsilon}{q} \). Moreover, it holds that

\[ \mathbb{E}\left[ \sum_{i=1}^{N} z_i z_i^\top \right] = \frac{\varepsilon}{q} \cdot \frac{N}{\sum_{t=1}^{m} R_t(A, u, \ell)} \cdot (uI - A)^{-1} \]

\[ \leq \frac{\varepsilon}{q} \cdot \frac{N}{\sum_{t=1}^{m} R_t(A, u, \ell)} \cdot \lambda_{\max}\left(\frac{1}{uI - A}\right) \cdot I. \]

This implies that

\[ \lambda_{\max}\left(\mathbb{E}\left[ \sum_{i=1}^{N} z_i z_i^\top \right]\right) \leq \frac{\varepsilon}{q} \cdot \frac{N}{\sum_{t=1}^{m} R_t(A, u, \ell)} \cdot \lambda_{\max}\left(\frac{1}{uI - A}\right) \cdot I. \]

By setting

\[ \mu = \frac{\varepsilon}{q} \cdot \frac{N}{\sum_{i=1}^{m} R_i(A, u, \ell)} \cdot \lambda_{\max}\left(\frac{1}{uI - A}\right), \]
it holds by the Matrix Chernoff Bound (cf. Lemma III.2) that
\[
P \left[ \lambda_{\max} \left( \sum_{i=1}^{N} z_i z_i^\top \right) \geq (1 + \delta) \mu \right] \leq n \cdot \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\mu q / \varepsilon}.
\]
Set the value of \(1 + \delta\) to be
\[
1 + \delta = \frac{1}{2 \mu} = \frac{q}{2 \varepsilon N} \cdot \left( \sum_{j=1}^{m} R_j(A, u, \ell) \right) \cdot \frac{1}{\lambda_{\max} \left( \sum_{i=1}^{N} z_i z_i^\top \right)} \cdot \lambda_{\min}(uI - A)
\]
\[
\geq \frac{q}{4 \varepsilon} \cdot \frac{n^{2/q}}{q},
\]
where the last inequality follows from the condition on \(N\). Hence, with probability at least
\[
1 - n \cdot \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\mu q / \varepsilon} \geq 1 - n \cdot \left( \frac{e}{1 + \delta} \right)^{(1+\delta) \cdot \mu q / \varepsilon} \geq 1 - n \cdot \left( \frac{e}{1 + \delta} \right)^{q} \geq 1 - \frac{\varepsilon^2}{100 q n},
\]
we have that
\[
\lambda_{\max} \left( \sum_{i=1}^{N} z_i z_i^\top \right) \leq (1 + \delta) \cdot \mu = \frac{1}{2},
\]
which implies that \(0 \preceq \sum_{i=1}^{N} z_i z_i^\top \leq \frac{1}{2} \cdot I\) and \(0 \preceq W \preceq \frac{1}{2} \cdot (uI - A)\).

Now we analyze the change of the potential function after each iteration, and show that the expected value of the potential function decreases over time. By Lemma III.3, with probability at least \(1 - \frac{\varepsilon^2}{100 q n}\), it holds that
\[
0 \preceq W \preceq \frac{1}{2} \cdot (uI - A).
\]

We define
\[
\mathbb{E}[f(W)] = \sum_{W \sim D(A, u, \ell)} \mathbb{P}[W \text{ is chosen and } W \leq \frac{1}{2}(uI - A)] \cdot f(W).
\]

Lemma III.4 below shows how the potential function changes after each iteration, and plays a key role in our analysis. This lemma was first proved in [1] for the case of \(q = 1\), and was extended in [3] to general values of \(q\). For completeness, we include the proof of the lemma in the appendix.

**Lemma III.4** ([3]). Let \(q \geq 10\) and \(\varepsilon \leq 1/10\). Suppose that \(w^T(uI - A)^{-1}w \leq \frac{\varepsilon}{q}\) and \(w^T(A - \ell I)^{-1}w \leq \frac{\varepsilon}{q}\). It holds that
\[
\text{tr}(A + w w^T - \ell I)^{-q} \leq \text{tr}(A - \ell I)^{-q} - q(1 - \varepsilon) w^T(A - \ell I)^{-q+1} w,
\]
and
\[
\text{tr}(uI - A - w w^T)^{-q} \leq \text{tr}(uI - A)^{-q} + q(1 + \varepsilon) w^T(uI - A)^{-q+1} w.
\]

**Lemma III.5.** Let \(j\) be any iteration. It holds that
\[
\mathbb{E}[\Phi_{u_{j+1}, \ell_{j+1}}(A_{j+1})] \leq \Phi_{u_j, \ell_j}(A_j).
\]
Proof: Let \( w_1 w_1^T, \ldots, w_N w_N^T \) be the matrices picked in iteration \( j \), and define for any \( 0 \leq i \leq N_j \) that
\[
B_i = A_j + \sum_{t=1}^i w_t w_t^T.
\]
We study the change of the potential function after adding a rank-1 matrix within each iteration. For this reason, we use
\[
\Delta_u = \frac{\Delta u_j}{N_j} = (1 + 2\varepsilon) \cdot \frac{\varepsilon}{q \cdot \sum_{t=1}^m R_t(A_j, u_j, \ell_j)},
\]
and
\[
\Delta_\ell = \frac{\Delta \ell_j}{N_j} = (1 - 2\varepsilon) \cdot \frac{\varepsilon}{q \cdot \sum_{t=1}^m R_t(A_j, u_j, \ell_j)}
\]
to express the average change of the barrier values \( \Delta u_j \) and \( \Delta \ell_j \). We further define for \( 0 \leq j \leq N_j \) that
\[
h_i = u_j + i \cdot \Delta u, \quad \ell_i = \ell_j + i \cdot \Delta \ell.
\]
Assuming \( 0 \leq W_j \leq \frac{1}{2} (u_j I - A_j) \), we claim that
\[
w_i w_i^T \preceq \frac{2\varepsilon}{q} \cdot (h_i I - B_{i-1}) \text{ and } w_i w_i^T \preceq \frac{2\varepsilon}{q} \cdot \left( B_{i-1} - \ell_i I \right),
\]
for any \( 1 \leq i \leq N_j \). Based on this, we apply Lemma III.4 and get that
\[
\mathbb{E}\left[ \Phi_{h_i, \ell_i} (B_{i-1} + w_i w_i^T) \right] \leq \Phi_{h_i, \ell_i} (B_{i-1}) + q(1 + 2\varepsilon)\text{tr}\left( (h_i I - B_{i-1})^{-(q+1)} \mathbb{E} \left[ w_i w_i^T \right] \right) - q(1 - 2\varepsilon)\text{tr}\left( (B_{i-1} - \ell_i I)^{-(q+1)} \mathbb{E} \left[ w_i w_i^T \right] \right) = \Phi_{h_i, \ell_i} (B_{i-1}) + q \cdot \Delta u \cdot \text{tr}\left( (h_i I - B_{i-1})^{-(q+1)} \right) - q \cdot \Delta \ell \cdot \text{tr}\left( (B_{i-1} - \ell_i I)^{-(q+1)} \right). \tag{7}
\]
We define a function \( f_i \) by
\[
f_i(t) = \text{tr}\left( (h_{i-1} + t \cdot \Delta u) I - B_{i-1} \right)^{-q} + \text{tr}\left( B_{i-1} - \left( \ell_{i-1} + t \cdot \Delta \ell \right) I \right)^{-q}.
\]
Notice that
\[
\frac{df_i(t)}{dt} = -q \cdot \Delta u \cdot \text{tr}\left( (h_{i-1} + t \cdot \Delta u) I - B_{i-1} \right)^{-(q+1)} + q \cdot \Delta \ell \cdot \text{tr}\left( B_{i-1} - \left( \ell_{i-1} + t \cdot \Delta \ell \right) I \right)^{-(q+1)}.
\]
Since \( f \) is convex, we have that
\[
\frac{df_i(t)}{dt} \bigg|_{t=1} \geq f_i(1) - f_i(0) = \Phi_{h_i, \ell_i} (B_{i-1}) - \Phi_{h_{i-1}, \ell_{i-1}} (B_{i-1}). \tag{8}
\]
Putting (7) and (8) together, we have that
\[
\mathbb{E}\left[ \Phi_{h_i, \ell_i} (B_i) \right] \leq \Phi_{h_i, \ell_i} (B_{i-1}) - \frac{df_i(t)}{dt} \bigg|_{t=1} \leq \Phi_{h_{i-1}, \ell_{i-1}} (B_{i-1}).
\]
Repeat this argument, we have that
\[
\mathbb{E}\left[ \Phi_{u_{j+1}, \ell_{j+1}} (A_{j+1}) \right] = \mathbb{E}\left[ \Phi_{u_{N_j}, \ell_{N_j}} (B_{N_j}) \right] \leq \Phi_{u_0, \ell_0} (B_0) = \Phi_{u_j, \ell_j} (A_j),
\]
which proves the statement.

So, it suffices to prove the claim (6). Since \( v v^T \preceq (v^T B^{-1} v) B \) for any vector \( v \) and PSD matrix \( B \), we have
Lemma III.6. The output matrix $A_k$ has condition number at most $1 + O(\varepsilon)$.

Proof: Since the condition number of $A_k$ is at most

$$\frac{u_k}{\ell_k} = \left(1 - \frac{u_k - \ell_k}{u_k}\right)^{-1},$$

it suffices to prove that $(u_k - \ell_k)/u_k = O(\varepsilon)$.

Since the increase rate of $\Delta_{u,j} - \Delta_{\ell,j}$ with respect to $\Delta_{u,j}$ for any iteration $j$ is

$$\frac{\Delta_{u,j} - \Delta_{\ell,j}}{\Delta_{u,j}} = \frac{(1 + 2\varepsilon) - (1 - 2\varepsilon)}{1 + 2\varepsilon} = \frac{4\varepsilon}{1 + 2\varepsilon} \leq 4\varepsilon,$$

we have that

$$\frac{u_k - \ell_k}{u_k} \leq \frac{2 \cdot (2n)^{1/q} + \sum_{j=0}^{k-1} (\Delta_{u,j} - \Delta_{\ell,j})}{(2n)^{1/q} + \sum_{j=0}^{k-1} \Delta_{u,j}} \leq \frac{2 \cdot (2n)^{1/q} + \sum_{j=0}^{k-1} (\Delta_{u,j} - \Delta_{\ell,j})}{(2n)^{1/q} + (4\varepsilon)^{-1} \sum_{j=0}^{k-1} (\Delta_{u,j} - \Delta_{\ell,j})}.$$

By the ending condition of the algorithm, it holds that $u_k - \ell_k \geq 4 \cdot (2n)^{1/q}$, i.e.

$$\sum_{j=0}^{k-1} (\Delta_{u,j} - \Delta_{\ell,j}) \geq 2 \cdot (2n)^{1/q}.$$ 

Hence, it holds that

$$\frac{u_k - \ell_k}{u_k} \leq \frac{2 \cdot (2n)^{1/q} + 2 \cdot (2n)^{1/q}}{(2n)^{1/q} + (4\varepsilon)^{-1} 2 \cdot (2n)^{1/q}} \leq 8\varepsilon,$$
which finishes the proof.

Now we prove that the algorithm finishes in \(O\left(\frac{qn^{3/2}}{\epsilon^2}\right)\) iterations, and picks \(O\left(\frac{qn}{\epsilon^2}\right)\) vectors in total.

**Lemma III.7.** The following statements hold:

- With probability at least 4/5, the algorithm finishes in \(10\frac{qn^{3/2}}{\epsilon^2}\) iterations.
- With probability at least 4/5, the algorithm chooses at most \(10\frac{qn}{\epsilon^2}\) vectors.

**Proof:** Notice that after iteration \(j\) the barrier gap \(u_j - \ell_j\) is increased by

\[
\Delta u_j - \Delta \ell_j = \frac{4\epsilon^2}{q} \frac{N_j}{\sum_{i=1}^{m} R_i(A_j, u_j, \ell_j)} = \frac{4\epsilon^2}{q} \frac{1}{n^{2/2}} \cdot \min \{\lambda_{\min}(u_j I - A_j), \lambda_{\min}(A_j - \ell_j I)\} \geq \frac{4\epsilon^2}{q} \frac{1}{n^{2/2}} \cdot \left(\Phi_{u_j, \ell_j}(A_j)\right)^{-1/2}.
\]

Since the algorithm finishes within \(k\) iterations if

\[
\sum_{j=0}^{k-1} (\Delta u_j - \Delta \ell_j) \geq 2 \cdot (2n)^{1/2},
\]

it holds that

\[
P[\text{algorithm finishes within } k \text{ iterations}] \geq P\left[\sum_{j=0}^{k-1} (\Delta u_j - \Delta \ell_j) \geq 2 \cdot (2n)^{1/2}\right] \geq P\left[\sum_{j=0}^{k-1} \frac{4\epsilon^2}{q n^{2/2}} \cdot \left(\Phi_{u_j, \ell_j}(A_j)\right)^{-1/2} \geq 2 \cdot (2n)^{1/2}\right] \geq P\left[\sum_{j=0}^{k-1} \frac{4\epsilon^2}{q n^{2/2}} \cdot \left(\Phi_{u_j, \ell_j}(A_j)\right)^{-1/2} \geq 2 \cdot (2n)^{1/2}\right] \geq P\left[\sum_{j=0}^{k-1} \frac{4\epsilon^2}{q n^{2/2}} \cdot \left(\Phi_{u_j, \ell_j}(A_j)\right)^{-1/2} \geq 2 \cdot (2n)^{1/2}\right],
\]

where the last inequality follows from the fact that

\[
\left(\sum_{j=0}^{k-1} \left(\Phi_{u_j, \ell_j}(A_j)\right)^{-1/2}\right) \cdot \left(\sum_{j=0}^{k-1} \left(\Phi_{u_j, \ell_j}(A_j)\right)^{1/2}\right) \geq k^2.
\]

By Lemma III.3, every picked matrix \(W_j\) in iteration \(j\) satisfies

\[
0 \preceq W_j \preceq \frac{1}{2} \cdot (u_j I - A)
\]

with probability at least \(1 - \frac{\epsilon^2}{100q^3}\), and with probability \(9/10\) all matrices picked in \(k = \frac{10q^3}{\epsilon^2}\) iterations satisfy the condition above. Also, by Lemma III.5 we have that

\[
\mathbb{E} \left[\sum_{j=0}^{k-1} \left(\Phi_{u_j, \ell_j}(A_j)\right)^{1/2}\right] = \sum_{j=0}^{k-1} \mathbb{E} \left[\left(\Phi_{u_j, \ell_j}(A_j)\right)^{1/2}\right] \leq \sum_{j=0}^{k-1} \left(\mathbb{E} \left[\Phi_{u_j, \ell_j}(A_j)\right]\right)^{1/2} \leq k,
\]

(9)
since the initial value of the potential function is at most 1. Therefore, it holds that
\[ P \{ \text{algorithm finishes in more than } k \text{ iterations} \} \]
\[ \leq P \left[ \sum_{j=0}^{k-1} \left( \Phi_{u_j, \ell_j} (A_j) \right)^{1/q} \geq 2 \cdot \frac{k^2 \varepsilon^2}{q} \cdot \left( \frac{1}{2n^3} \right)^{1/q} \right] \]
\[ \leq P \left[ \sum_{j=0}^{k-1} \left( \Phi_{u_j, \ell_j} (A_j) \right)^{1/q} \geq 2 \cdot \frac{k^2 \varepsilon^2}{q} \cdot \left( \frac{1}{2n^3} \right)^{1/q} \right. \]
\[ + \left. P \left[ \exists j : W_j \leq \frac{1}{2} (u_j I - A_j) \right] \right] \]
\[ \leq \frac{q}{2} \cdot \frac{4 \varepsilon^2}{k^2} \cdot (2n^3)^{1/q} + 1/10 \leq 1/5, \]
where the second last inequity follows from Markov’s inequality and (9), and the last inequality follows by our choice of \( k \). This proves the first statement.

Now for the second statement. Notice that for every vector chosen in iteration \( j \), the barrier gap \( \Delta_{u_j} - \Delta_{\ell_j} \) is increased on average by
\[ \frac{\Delta_{u_j} - \Delta_{\ell_j}}{N_j} = \frac{4 \varepsilon^2}{q} \sum_{i=1}^{m} R_i (A_j, u_j, \ell_j). \]

To bound \( R_i (A_j, u_j, \ell_j) \), let the eigenvalues of matrix \( A_j \) be \( \lambda_1, \ldots, \lambda_n \). Then, it holds that
\[ \sum_{i=1}^{m} R_i (A_j, u_j, \ell_j) = \sum_{i=1}^{m} v_i^T (u_j I - A_j)^{-1} v_i + \sum_{i=1}^{m} v_i^T (A_j - \ell_j I)^{-1} v_i \]
\[ = \sum_{i=1}^{m} \frac{1}{u_j - \lambda_i} + \sum_{i=1}^{m} \frac{1}{\lambda_i - \ell_j} \]
\[ \leq \left( \sum_{i=1}^{n} (u_j - \lambda_i)^{-q} + \sum_{i=1}^{n} (\lambda_i - \ell_j)^{-q} \right)^{1/q} (2n)^{1-1/q} \]
\[ = \left( \Phi_{u_j, \ell_j} (A_j) \right)^{1/q} \cdot (2n)^{1-1/q}. \]

Therefore, we have that
\[ \frac{\Delta_{u_j} - \Delta_{\ell_j}}{N_j} \geq \frac{4 \varepsilon^2}{q} \cdot \frac{1}{(2n)^{1-1/q} \cdot \left( \Phi_{u_j, \ell_j} (A_j) \right)^{1/q}}. \quad (10) \]

Let \( v_1, \ldots, v_z \) be the vectors sampled by the algorithm, and \( v_j \) is picked in iteration \( \tau_j \), where \( 1 \leq j \leq z \). We first assume that the algorithm could check the ending condition after adding every single vector. In such case, it holds that
\[ P \{ \text{algorithm finishes after choosing } z \text{ vectors} \} \geq P \left[ \sum_{j=1}^{z} \frac{4 \varepsilon^2}{q} \cdot \frac{1}{(2n)^{1-1/q} \cdot \left( \Phi_{u_{\tau_j}, \ell_{\tau_j}} (A_{\tau_j}) \right)^{1/q}} \geq 2 \cdot (2n)^{1/q} \right] \]
\[ = P \left[ \sum_{j=1}^{z} (\Phi_{u_{\tau_j}, \ell_{\tau_j}} (A_{\tau_j}))^{-1/q} \geq qn \varepsilon^2 \right]. \]

Following the same proof as the first part and noticing that in the final iteration the algorithm chooses at most \( O(n) \) extra vectors, we obtain the second statement. \( \blacksquare \)
C. Proof of the Main Results

Now we analyze the runtime of the algorithm, and prove the main results. We first analyze the algorithm for
sparsifying sums of rank-1 PSD matrices, and prove Theorem I.2.

Proof of Theorem I.2: By Lemma III.7, with probability at least $4/5$ the algorithm chooses at most $\frac{10qn}{\varepsilon^2}$ vectors, and by Lemma III.6 the condition number of $A_k$ is at most $1 + O(\varepsilon)$, implying that the matrix $A_k$ is a $(1 + O(\varepsilon))$-approximation of $I$. These two results together prove that $A_k$ is a linear-sized spectral sparsifier.

For the runtime, Lemma III.7 proves that the algorithm finishes in $\frac{10qn^3}{\varepsilon^2}$ iterations, and it is easy to see that all the required quantities in each iteration can be approximately computed in $\tilde{O}(m \cdot n^{\omega-1})$ time using fast matrix multiplication. Therefore, the total runtime of the algorithm is $\tilde{O} \left( \frac{qm}{\varepsilon^2} \cdot n^{\omega-1+3/q} \right)$.

Next we show how to apply our algorithm in the graph setting, and prove Theorem I.1. Let $L = \sum_{i=1}^{m} u_i u_i^\top$ be the Laplacian matrix of an undirected graph $G$, where $u_i u_i^\top$ is the Laplacian matrix of the graph consisting of a single edge $e_i$. By setting

$$v_i = L^{-1/2} u_i$$

for $1 \leq i \leq m$, it is easy to see that constructing a spectral sparsifier of $G$ is equivalent to sparsifying the matrix $\sum_{i=1}^{m} v_i v_i^\top$. We will present in the appendix almost-linear time algorithms to approximate the required quantities

$$\lambda_{\min}(u_j I - A_j), \lambda_{\min}(A_j - \ell_j I), v_i^\top (u_j I - A_j)^{-1} v_i, \text{ and } v_i^\top (A_j - \ell_j I)^{-1} v_i$$

in each iteration, and this gives Theorem I.1.

Proof of Theorem I.1: By applying the same analysis as in the proof of Theorem I.2, we know that the output matrix $A_k$ is a linear-sized spectral sparsifier, and it suffices to analyze the runtime of the algorithm.

By Lemma III.3 and the Union Bound, with probability at least $9/10$ all the matrices picked in $k = \frac{10qn^{3/q}}{\varepsilon^2}$ iterations satisfy

$$W_{\ell_j} \preceq \frac{1}{2} (u_j I - A_j).$$

Conditioning on the event, with constant probability $\mathbb{E} \left[ \Phi_{u_j, \ell_j}(A_j) \right] \leq 2$ for all iterations $j$, and by Markov’s inequality with high probability it holds that $\Phi_{u_j, \ell_j}(A_j) = \tilde{O} \left( \frac{m}{\varepsilon^2} \right)$ for all iterations $j$.

On the other hand, notice that it holds for any $1 \leq j \leq n$ that

$$(u - \lambda_j)^{-q} \leq \sum_{i=1}^{n} (u - \lambda_i)^{-q} < \Phi_{u, \ell}(A),$$

which implies that $\lambda_j < u - (\Phi_{u, \ell}(A))^{-1/q}$. Similarly, it holds that $\lambda_j > \ell + (\Phi_{u, \ell}(A))^{-1/q}$ for any $1 \leq j \leq n$. Therefore, we have that

$$\ell_j + O \left( \left( \frac{\varepsilon^2}{qn} \right)^{1/q} \right) I \prec A_j \prec \left( u_j - O \left( \left( \frac{\varepsilon^2}{qn} \right)^{1/q} \right) \right) I.$$

Since both of $u_j$ and $\ell_j$ are of the order $O(n^{1/q})$, we set $\eta = O \left( (\varepsilon/n)^2/q \right)$ and obtain that

$$\ell_j + |\ell_j| \eta I \prec A_j \prec (1 - \eta) u_j I.$$

Hence, we apply Lemma A.5 and Lemma A.6 to compute all required quantities in each iteration up to constant approximation in time

$$\tilde{O} \left( \frac{m}{\varepsilon^2 \cdot \eta} \right) = \tilde{O} \left( \frac{m \cdot n^{2/q}}{\varepsilon^{2+2/q}} \right).$$

Since by Lemma III.7 the algorithm finishes in $\frac{10qn^{3/q}}{\varepsilon^2}$ iterations with probability at least $4/5$, the total runtime
of the algorithm is
\[ \tilde{O}\left(\frac{q \cdot m \cdot n^{5/q}}{\varepsilon^{4+4/q}}\right). \]

**ACKNOWLEDGMENT**

This work was partially supported by NSF awards 0843915 and 1111109. Part of this work was done while both authors were visiting the Simons Institute for the Theory of Computing, UC Berkeley, and the second author was affiliated with the Max Planck Institute for Informatics, Germany. We thank Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia for sending us their manuscript of [3] and the inspiring talk Zeyuan Allen-Zhu gave at the Simons Institute for the Theory of Computing. Finally, we thank Michael Cohen for pointing out a gap in a previous version of the paper and his fixes for the gap, as well as Lap-Chi Lau for many insightful comments on improving the presentation of the paper.

**REFERENCES**


**APPENDIX**

A. _Estimates of the Potential Functions_

In this subsection we prove Lemma III.4. We first list the following two lemmas, which will be used in our proof.
Lemma A.1 (Sherman-Morrison Formula). Let $A \in \mathbb{R}^{n \times n}$ be an invertible matrix, and $u, v \in \mathbb{R}^n$. Suppose that $1 + v^\top A^{-1} u \neq 0$. Then it holds that

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1} u}.$$  

Lemma A.2 (Lieb-Thirring Inequality, [27]). Let $A$ and $B$ be positive definite matrices, and $q \geq 1$. Then it holds that

$$\text{tr}(BAB)^q \leq \text{tr}(B^q A^q B^q).$$

Proof of Lemma III.4: Let $Y = A - \ell I$. By the Sherman-Morrison Formula (Lemma A.1), it holds that

$$\text{tr}(Y + ww^\top)^{-q} = \text{tr}\left(Y^{-1} - \frac{Y^{-1}ww^\top Y^{-1}}{1 + w^\top Y^{-1} w}\right)^q. \quad (11)$$

By the assumption of $w^\top Y^{-1} w \leq \frac{\varepsilon}{q}$, we have that

$$\text{tr}(Y + ww^\top)^{-q} \leq \text{tr}\left(Y^{-1} - \frac{Y^{-1}ww^\top Y^{-1}}{1 + \varepsilon/q}\right)^q \quad (12)$$

$$= \text{tr}\left(Y^{-1/2} \left(I - \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1 + \varepsilon/q}\right) Y^{-1/2}\right)^q$$

$$\leq \text{tr}\left(Y^{-q/2} \left(I - \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1 + \varepsilon/q}\right)^q Y^{-q/2}\right) \quad (13)$$

$$= \text{tr}\left(Y^{-q} \left(I - \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1 + \varepsilon/q}\right)^q\right), \quad (14)$$

where (12) uses the fact that $A \preceq B$ implies that $\text{tr}(A^q) \leq \text{tr}(B^q)$, (13) follows from the Lieb-Thirring inequality (Lemma A.2), and (14) uses the fact that the trace is invariant under cyclic permutations.

Let

$$D = \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1 + \varepsilon/q}.$$  

Note that $0 \preceq D \preceq \frac{\varepsilon}{q} I$, and

$$(I - D)^q \preceq I - qD + \frac{q(q - 1)}{2} D^2 \preceq I - \left(q - \frac{\varepsilon(q - 1)}{2}\right) D$$

Therefore, we have that

$$\left(I - \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1 + \varepsilon/q}\right)^q \preceq \left(I - \left(q - \frac{\varepsilon(q - 1)}{2}\right) \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1 + \varepsilon/q}\right)$$

$$\preceq I - \left(q - \frac{\varepsilon(q - 1)}{2}\right) \left(1 - \frac{\varepsilon}{q}\right) Y^{-1/2}ww^\top Y^{-1/2}$$

$$\preceq I - q \left(1 - \frac{\varepsilon(q + 1)}{2q}\right) Y^{-1/2}ww^\top Y^{-1/2}$$

$$\preceq I - q \left(1 - \varepsilon\right) Y^{-1/2}ww^\top Y^{-1/2}.$$
This implies that
\[ \text{tr}(Y + w^\top w)^{-q} \leq \text{tr} \left( Y^{-q} \left( I - q(1 - \varepsilon)Y^{-1/2}ww^\top Y^{-1/2} \right) \right) \leq \text{tr} \left( Y^{-q} \right) - q(1 - \varepsilon) w^\top Y^{-(q+1)} w, \]
which proves the first statement.

Now for the second inequality. Let \( Z = uI - A \). By the Sherman-Morrison Formula (Lemma A.1), it holds that
\[ \text{tr}(Z - ww^\top)^{-q} = \text{tr} \left( Z^{-1} + \frac{Z^{-1}ww^\top Z^{-1}}{1 - w^\top Z^{-1}w} \right)^q. \] (15)

By the assumption of \( w^\top Z^{-1}w \leq \frac{\varepsilon}{q} \), it holds that
\[ \text{tr}(Z - ww^\top)^{-q} \leq \text{tr} \left( Z^{-1} + \frac{Z^{-1}ww^\top Z^{-1}}{1 - \varepsilon/q} \right)^q \]
\[ = \text{tr} \left( Z^{-1/2} \left( I + \frac{Z^{-1/2}ww^\top Z^{-1/2}}{1 - \varepsilon/q} \right) Z^{-1/2} \right)^q \]
\[ \leq \text{tr} \left( Z^{-q/2} \left( I + \frac{Z^{-1/2}ww^\top Z^{-1/2}}{1 - \varepsilon/q} \right)^q Z^{-q/2} \right) \] (17)
\[ = \text{tr} \left( Z^{-q} \left( I + \frac{Z^{-1/2}ww^\top Z^{-1/2}}{1 - \varepsilon/q} \right)^q \right), \] (18)
where (16) uses the fact that \( A \preceq B \) implies that \( \text{tr} (A^q) \leq \text{tr} (B^q) \), (17) follows from the Lieb-Thirring inequality (Lemma A.2), and (18) uses the fact that the trace is invariant under cyclic permutations.

Let
\[ E = Z^{-1/2}ww^\top Z^{-1/2}. \]

Combining \( E \preceq \frac{\varepsilon}{q} I \) with the assumption that \( q \geq 10 \) and \( \varepsilon \leq 1/10 \), we have that
\[ \left( I + \frac{E}{1 - \varepsilon/q} \right)^q \preceq I + \frac{qE}{1 - \varepsilon/q} + \frac{q(q - 1)}{2} \left( 1 + \frac{\varepsilon/q}{1 - \varepsilon/q} \right)^{q-2} \left( \frac{E}{1 - \varepsilon/q} \right)^2 \]
\[ \preceq I + q \left( 1 + \frac{1.1}{q} \right) E + 1.4 \frac{q(q - 1)}{2} E^2 \]
\[ \preceq I + q (1 + 0.3\varepsilon) E + 0.7\varepsilon q E \]
\[ \preceq I + q (1 + \varepsilon) E. \]

Therefore, we have that
\[ \text{tr}(Z - ww^\top)^{-q} \leq \text{tr} \left( Z^{-q} \right) + q(1 + \varepsilon) w^\top Z^{-(q+1)} w, \]
which proves the second statement.

\[ \blacksquare \]

B. Implementation of the Algorithm

In this section, we show that the algorithm for constructing graph sparsification runs in almost-linear time. Based on previous discussion, we only need to prove that, for any iteration \( j \), the number of samples \( N_j \) and \( \{ R_i(A_j, u_j, \ell_j) \}_{i=1}^m \) can be approximately computed in almost-linear time. By definition, it suffices to compute \( \lambda_{\min}(u_j I - A_j), \lambda_{\min}(A_j - \ell_j I), v_i^\top (u_j I - A_j)^{-1} v_i, \) and \( v_i^\top (A_j - \ell_j I)^{-1} v_i \) for all \( i \). For simplicity we drop the subscript \( j \) expressing the iterations in this subsection. We will assume that the following assumption holds on \( A \). We remark that an almost-linear time algorithm for computing similar quantities was shown in [3].

**Assumption A.3.** Let \( L \) and \( \bar{L} \) be the Laplacian matrices of graph \( G \) and its subgraph after reweighting. Let
\[ A = L^{-1/2}LL^{-1/2}, \text{ and assume that} \]
\[ (\ell + |\ell|\eta) \cdot I < A < (1 - \eta)u \cdot I \]
holds for some \( 0 < \eta < 1 \).

**Lemma A.4.** Under Assumption A.3, the following statements hold:

- We can construct a matrix \( S_u \) such that
  \[ S_u \approx \varepsilon/10 \left( uI - A \right)^{-1/2}, \]
  and \( S_u = p(A) \) for a polynomial \( p \) of degree \( O \left( \frac{\log(1/\varepsilon \eta)}{\eta} \right) \).
- We can construct a matrix \( S_\ell \) such that
  \[ S_\ell \approx \varepsilon/10 \left( A - \ell I \right)^{-1/2}. \]
Moreover, \( S_\ell \) is of the form \( (A')^{-1/2}q((A')^{-1}) \), where \( q \) is a polynomial of degree \( O \left( \frac{\log(1/\varepsilon \eta)}{\eta} \right) \) and \( A' = L^{-1/2}L'LL^{-1/2} \) for some Laplacian matrix \( L' \).

**Proof:** By Taylor expansion, it holds that
\[ (1 - x)^{-1/2} = 1 + \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} \left( j + \frac{1}{2} \right) \frac{x^k}{k!}. \]
We define for any \( T \in \mathbb{N} \) that
\[ p_T(x) = 1 + \sum_{k=1}^{T} \prod_{j=0}^{k-1} \left( j + \frac{1}{2} \right) \frac{x^k}{k!}. \]
Then, it holds for any \( 0 < x < 1 - \eta \) that
\[ p_T(x) \leq (1 - x)^{-1/2} = p_T(x) + \sum_{k=T+1}^{\infty} \prod_{j=0}^{k-1} \left( j + \frac{1}{2} \right) \frac{x^k}{k!} \]
\[ \leq p_T(x) + \sum_{k=T+1}^{\infty} x^k \]
\[ \leq p_T(x) + \frac{(1 - \eta)^{T+1}}{\eta}. \]
Hence, it holds that
\[ (uI - A)^{-1/2} = u^{-1/2}(I - u^{-1}A)^{-1/2} \geq u^{-1/2}p_T(u^{-1}A), \]
and
\[ (uI - A)^{-1/2} \leq u^{-1/2} \left( p_T(u^{-1}A) + \frac{(1 - \eta)^{T+1}}{\eta} \cdot I \right), \]
since \( u^{-1}A \preceq (1 - \eta)I \). Notice that \( u^{-1/2}I \preceq (uI - A)^{-1/2} \), and therefore
\[ (uI - A)^{-1/2} \leq u^{-1/2}p_T(u^{-1}A) + \frac{(1 - \eta)^{T+1}}{\eta} \cdot (uI - A)^{-1/2}. \]
Setting \( T = \frac{c \log(1/(\varepsilon \eta))}{\eta} \) for some constant \( c \) and defining \( S_u = u^{-1/2}p_T(u^{-1}A) \) gives us that
\[ S_u \approx \varepsilon/10 (uI - A)^{-1/2}. \]
Now for the second statement. Our construction of $S_\ell$ is based on the case distinction ($\ell > 0$, and $\ell \leq 0$).

Case (1): $\ell > 0$. Notice that
\[(A - \ell I)^{-1/2} = A^{-1/2}(I - \ell A^{-1})^{-1/2},\]
and
\[p_T(\ell A^{-1}) \leq (I - \ell A^{-1})^{-1/2} \leq p_T(\ell A^{-1}) + \frac{(1 - \eta/2)^{T+1}}{\eta/2} \cdot I.\]

Using the same analysis as before, we have that
\[A^{-1/2}(I - \ell A^{-1})^{-1/2} \approx_{\varepsilon/10} A^{-1/2}p_T(\ell A^{-1}).\]

By defining $S_\ell = A^{-1/2}p_T(\ell A^{-1})$, i.e., $A' = A$ and $q((A')^{-1}) = p_T(\ell A^{-1})$, we have that
\[S_\ell \approx_{\varepsilon/10} (A - \ell I)^{-1/2}.\]

Case (2): $\ell \leq 0$. We look at the matrix
\[A - \ell I = L^{-1/2}\bar{L}L^{-1/2} - \ell I = L^{-1/2}(\bar{L} - \ell L)L^{-1/2}.\]

Notice that $\bar{L} - \ell L$ is a Laplacian matrix, and hence this reduces to the case of $\ell = 0$, for which we simply set $S_\ell = (A - \ell I)^{-1/2}$. Therefore, we can write $S_\ell$ as a desired form, where $A' = A - \ell I$ and polynomial $q = 1$. 

Lemma A.5 below shows how to estimate $v_i^T(uI - A)^{-1}v_i$, and $v_i^T(A - \ell I)^{-1}v_i$, for all $v_i$ in nearly-linear time.

**Lemma A.5.** Let $A = \sum_{i=1}^m v_i v_i^T$, and suppose that $A$ satisfies Assumption A.3. Then, we can compute $\{r_i\}_{i=1}^m$ and $\{t_i\}_{i=1}^m$ in $O\left(\frac{m}{\varepsilon^2 \eta}\right)$ time such that
\[(1 - \varepsilon)r_i \leq v_i^T(uI - A)^{-1}v_i \leq (1 + \varepsilon)r_i,\]
and
\[(1 - \varepsilon)t_i \leq v_i^T(A - \ell I)^{-1}v_i \leq (1 + \varepsilon)t_i.\]

**Proof:** Define $u_i = L^{1/2}v_i$ for any $1 \leq i \leq m$. By Lemma A.4, we have that
\[v_i^T(uI - A)^{-1}v_i \approx_{3\varepsilon/10} \|p(A)v_i\|^2 = \left\|p\left(L^{-1/2}\bar{L}L^{-1/2}\right)L^{-1/2}u_i\right\|^2 = \left\|L^{1/2}p\left(L^{-1}\bar{L}\right)L^{-1}u_i\right\|^2.\]

Let $L = B^\top B$ for some $B \in \mathbb{R}^{m \times n}$. Then, it holds that
\[v_i^T(uI - A)^{-1}v_i \approx_{3\varepsilon/10} \left\|Bp\left(L^{-1}\bar{L}\right)L^{-1}u_i\right\|^2.\]

We invoke the Johnson-Lindenstrauss Lemma and find a random matrix $Q \in \mathbb{R}^{O(\log n/\varepsilon^2) \times m}$: With high probability, it holds that
\[v_i^T(uI - A)^{-1}v_i \approx_{4\varepsilon/10} \left\|QBp\left(L^{-1}\bar{L}\right)L^{-1}u_i\right\|^2.\]

We apply a nearly-linear time Laplacian solver to compute $\|QBp\left(L^{-1}\bar{L}\right)L^{-1}u_i\|^2$ for all $\{u_i\}_{i=1}^m$ up to $(1 \pm \varepsilon/10)$-multiplicative error in time $O\left(\frac{m}{\varepsilon^2 \eta}\right)$. This gives the desired $\{r_i\}_{i=1}^m$. 


The computation for \( \{t_i\}_{i=1}^m \) is similar. By Lemma A.4, it holds for any \( 1 \leq i \leq m \) that

\[
v_i^T (A - \ell I)^{-1} v_i \approx 3\varepsilon/10 \| (A')^{-1/2} q((A')^{-1}) v_i \|^2
\]

\[
= \| (A')^{-1/2} q \left( L^{1/2} (L')^{-1} L^{1/2} \right) L'^{-1/2} u_i \|^2
\]

\[
= \| (A')^{-1/2} L^{-1/2} q(L(L')^{-1}) u_i \|^2.
\]

Let \( L' = (B')^\top (B') \) for some \( B' \in \mathbb{R}^{m \times n} \). Then, it holds that

\[
v_i^T (A - \ell I)^{-1} v_i \approx 3\varepsilon/10 \| (L')^{-1/2} q \left( L(L')^{-1} \right) u_i \|^2
\]

\[
= \| (L')^{1/2} (L')^{-1} q \left( L(L')^{-1} \right) u_i \|^2
\]

\[
= \| B'(L')^{-1} q \left( L(L')^{-1} \right) u_i \|^2.
\]

We invoke the Johnson-Lindenstrauss Lemma and a nearly-linear time Laplacian solver as before to obtain required \( \{t_i\}_{i=1}^m \). The total runtime is \( \widetilde{O} \left( \frac{m}{n\varepsilon^2} \right) \).

Lemma A.6 shows that how to approximate \( \lambda_{\min}(uI - A) \) and \( \lambda_{\min}(A - \ell I) \) in nearly-linear time.

**Lemma A.6.** Under Assumption A.3, we can compute values \( \alpha, \beta \) in \( \widetilde{O} \left( \frac{m}{\varepsilon^2} \right) \) time such that

\[
(1 - \varepsilon) \alpha \leq \lambda_{\min}(uI - A) \leq (1 + \varepsilon) \alpha
\]

and

\[
(1 - \varepsilon) \beta \leq \lambda_{\min}(A - \ell I) \leq (1 + \varepsilon) \beta.
\]

**Proof:** By Lemma A.4, we have that \( S_u \approx \varepsilon/10 \ (uI - A)^{-1/2} \). Hence, \( \lambda_{\max}(S_u)^{-2} \approx 3\varepsilon/10 \lambda_{\min}(uI - A) \), and it suffices to estimate \( \lambda_{\max}(S_u) \). Since

\[
\lambda_{\max}(S_u) \leq \left( \text{tr} \left( S_u^{2k} \right) \right)^{1/2k} \leq n^{1/2k} \lambda_{\max}(S_u),
\]

by picking \( k = \log n/\varepsilon \) we have that \( \left( \text{tr} \left( S_u^{2k} \right) \right)^{1/2k} \approx \varepsilon/2 \lambda_{\max}(S_u) \). Notice that

\[
\text{tr} \left( S_u^{2k} \right) = \text{tr} \left( p^{2k} \left( L^{-1/2} \tilde{L} L^{-1/2} \right) \right) = \text{tr} \left( p^{2k} \left( L^{-1/2} \tilde{L} L^{-1/2} \right) \right).
\]

Set \( \tilde{L} = \tilde{B}^\top \tilde{B} \) for some matrix \( \tilde{B} \in \mathbb{R}^{m \times n} \), and we have that \( \text{tr} \left( S_u^{2k} \right) = \text{tr} \left( p^{2k} \left( \tilde{B} L^{-1} \tilde{B}^\top \right) \right) \). Since we can apply \( p^k \left( \tilde{B} L^{-1} \tilde{B}^\top \right) \) to vectors in \( \widetilde{O} \left( \frac{m}{\eta\varepsilon} \right) \) time, we invoke the Johnson-Lindenstrauss Lemma and approximate \( \text{tr} \left( S_u^{2k} \right) \) in \( \widetilde{O} \left( \frac{m}{\eta\varepsilon^2} \right) \) time.

We approximate \( \lambda_{\min}(A - \ell I) \) in a similar way. Notice that

\[
\text{tr} \left( S_u^{4k} \right) = \text{tr} \left( (A')^{-1/2} q((A')^{-1}) (A')^{4k} \right)
\]

\[
= \text{tr} \left( q((A')^{-1}) (A')^{-1} q((A')^{-1}) \right)^{2k}.
\]

Let \( z \) be a polynomial defined by \( z(x) = xq^2(x) \) and \( L' = (B')^\top(B') \). Then, we have that

\[
\text{tr} \left( S_u^{4k} \right) = \text{tr} \left( z^{2k}((A')^{-1}) \right) = \text{tr} \left( z^{2k} \left( L^{1/2}(L')^{-1} L^{1/2} \right) \right).
\]

Applying the same analysis as before, we can estimate the trace in \( \widetilde{O} \left( \frac{m}{\eta\varepsilon^2} \right) \) time. ■