

A Spectral Approach to Network Design

Lap Chi Lau*, Hong Zhou†

Abstract

We present a spectral approach to design approximation algorithms for network design problems. We observe that the underlying mathematical questions are the spectral rounding problems, which were studied in spectral sparsification and in discrepancy theory. We extend these results to incorporate additional non-negative linear constraints, and show that they can be used to significantly extend the scope of network design problems that can be solved. Our algorithm for spectral rounding is an iterative randomized rounding algorithm based on the regret minimization framework. In some settings, this provides an alternative spectral algorithm to achieve constant factor approximation for the classical survivable network design problem, and partially answers a question of Bansal about survivable network design with concentration property. We also show many other applications of the spectral rounding results, including weighted experimental design and additive spectral sparsification.

*School of Computer Science, University of Waterloo. Supported by NSERC Discovery Grant 2950-120715 and NSERC Accelerator Supplement 2950-120719. Email: lapchi@uwaterloo.ca

†School of Computer Science, University of Waterloo. Supported by NSERC Discovery Grant 2950-120715 and NSERC Accelerator Supplement 2950-120719. Email: h76zhou@uwaterloo.ca

1 Introduction

Network design is a central topic in combinatorial optimization, approximation algorithms and operations research. The general setting of network design is to find a minimum cost subgraph satisfying certain requirements. The most well-studied problem is the survivable network design problem [40, 1, 41, 35], where the requirement is to have at least a specified number $f_{u,v}$ of edge-disjoint paths between every pair of vertices u, v . A seminal work of Jain [43] introduced the iterative rounding method for linear programming to design a 2-approximation algorithm for the survivable network design problem, and this method has been extended to various more general settings [30, 34, 23, 46, 47, 28, 32, 49, 10]. There are also other linear programming based algorithms such as randomized rounding [70, 36, 19, 9, 42] to obtain important algorithmic results for network design. It is widely recognized that linear programming is the most general and powerful approach in designing approximation algorithms for network design problems.

In the past decade, spectral techniques have been developed to make significant progress in designing graph algorithms [68, 24, 15, 3, 6, 64]. One striking example is the spectral sparsification problem introduced by Spielman and Teng [69], where the objective is to find a sparse edge-weighted graph H to approximate the input graph G so that $(1 - \epsilon)L_G \preceq L_H \preceq (1 + \epsilon)L_G$ where L_G and L_H are the Laplacian matrices of the graph G and H . The spectral condition $(1 - \epsilon)L_G \preceq L_H \preceq (1 + \epsilon)L_G$ implies that H is also a cut sparsifier of G such that the total weight on each cut in H is approximately the same as that in G . Batson, Spielman, Srivastava [15] proved that every graph G has a spectral sparsifier H with only $O(n/\epsilon^2)$ edges. This improves upon the influential result of Benczúr and Karger [16] that every graph G has a cut sparsifier H with $O(n \log n/\epsilon^2)$ edges, which has many applications in designing fast algorithms for graph problems. From a technical perspective, the spectral approach introduces linear algebraic concepts and continuous optimization techniques in solving graph problems, and the results in spectral sparsification [15, 3, 6] show that it is algorithmically more convenient to control the spectral properties of the graph in order to control its combinatorial properties.

Inspired by these developments, we are motivated to study whether there is a spectral approach to design approximation algorithms for network design problems. The general way to designing approximation algorithms is to solve a convex program to obtain a fractional solution x in polynomial time, and then to round x into an integral solution z that well approximates x (with respect to the constraints and the objective function) as an approximate solution. We observe that the following spectral rounding question, where the objective is to approximate the spectral properties of x , underlies a large class of problems including the survivable network design problem.

Question 1.1 (Spectral Rounding). *For each edge e in a graph, let L_e be the Laplacian matrix of e and c_e be its cost. Given $x_e \in \mathbb{R}_+$ for each edge e , characterize when we can find $z_e \in \mathbb{Z}_+$ for each e such that*

$$\sum_e x_e L_e \approx \sum_e z_e L_e \quad \text{and} \quad \sum_e c_e x_e \approx \sum_e c_e z_e.$$

When spectral rounding is possible, we notice that the integral solution z not only approximately preserves the cost and the pairwise edge connectivity properties of x as required by the survivable network design problem, but also many other properties of x including pairwise effective resistances,

the graph expansion, and degree constraints. This would significantly extend the scope of useful properties that a network designer could control simultaneously to design better networks.

1.1 General Survivable Network Design

The main conceptual contribution of this paper is to show that the techniques in spectral graph theory and discrepancy theory can be used to significantly extend the scope of network design problems that can be solved.

In network design, we are given a graph $G = (V, E)$ where each edge has a cost c_e , and the objective is to find a minimum cost subgraph that satisfies certain requirements. In survivable network design [40, 43], the requirements are pairwise edge-connectivities, that every pair of vertices u, v should have at least f_{uv} edge-disjoint paths for $u, v \in V$. This captures several classical problems as special cases, including minimum Steiner tree [19], minimum Steiner forest [1, 41], and minimum k -edge-connected subgraph [35]. Jain introduced the iterative rounding method for linear programming to design a 2-approximation algorithm for the survivable network design problem [43]. His proof exploits the nice structures of the connectivity constraints to show that there is always a variable x_e with value at least $\frac{1}{2}$ in any extreme point solution to the linear program. His work leads to many subsequent developments in network design [30, 23, 34, 35, 22], and the iterative rounding algorithm is still the only known constant factor approximation algorithm for the survivable network design problem.

Motivated by the need of more realistic models for the design of practical networks, researchers study generalizations of survivable network design problems where we can incorporate additional useful constraints. One well-studied problem is the degree-constrained survivable network design problem, where there is a degree upper bound d_v on each vertex v to control its workload. There is a long line of work on this problem [62, 63, 39, 46, 28, 32, 49] and the iterative rounding method has been extended to incorporate degree constraints into survivable network design successfully. In the general setting [46, 53, 49], there is a polynomial time algorithm to find a subgraph that violates the cost and the degree constraints by a multiplicative factor of at most 2. For interesting special cases such as finding a spanning tree [39, 66] or a Steiner tree [48, 49], there is a polynomial time algorithm that returns a solution that violates the degree constraint by an additive constant.

More generally, one can consider to add linear packing constraints and linear covering constraints into survivable network design [17, 13, 60, 52], but not as much is known about how to approximately satisfy these constraints simultaneously especially when the linear constraints are unstructured.

Another natural constraint is to control the shortest path distance between pairs of vertices, but unfortunately this is proved to be computationally hard [27] to incorporate into network design.

In [20], together with Chan, Schild, and Wong, we propose to incorporate the effective resistance metric into network design, as an interpolation of shortest path distance and edge-connectivity between vertices. Incorporating effective resistances can also allow one to control some natural quantities about random walks on the resulting subgraph, such as the commute time between vertices [21] and the cover time [58, 26]. We note that effective resistances have interesting connections to many other graph problems, including spectral sparsification [68], maximum flow computation [24], asym-

metric traveling salesman problem [6], and random spanning tree generation [55, 64]. We believe that it is a useful property to be incorporated into network design.

There are many other natural constraints that could help in designing better networks, including total effective resistances [38], algebraic connectivity (and graph expansion) [37], and the mixing time of random walks [18]. These constraints are also well-motivated and were studied individually before (without taking other constraints together into consideration, e.g. connectivity requirements), but not much is known about approximation algorithms with nontrivial approximation guarantees for these constraints (see Section 4.3).

It would be ideal if a network designer can control all of these properties simultaneously to design a good network that suits their need. We can write a convex programming relaxation for this general network design problem incorporating all these constraints.

$$\begin{array}{llll}
\text{cp} := \min_x \langle c, x \rangle & & & \\
x(\delta(S)) \geq f(S) & \forall S \subseteq V & \text{(connectivity constraints)} & \\
x(\delta(v)) \leq d_v & \forall v \in V & \text{(degree constraints)} & \\
Ax \leq a & A \in \mathbb{R}_+^{p \times m}, a \in \mathbb{R}_+^p & \text{(linear packing constraints)} & \\
Bx \geq b & B \in \mathbb{R}_+^{q \times m}, b \in \mathbb{R}_+^q & \text{(linear covering constraints)} & \\
\text{Reff}_x(u, v) \leq r_{uv} & \forall u, v \in V & \text{(effective resistance constraints)} & \text{(CP)} \\
L_x \succcurlyeq M & M \succcurlyeq 0 & \text{(spectral constraints)} & \\
\lambda_2(L_x) \geq \lambda & & \text{(algebraic connectivity constraint)} & \\
0 \leq x_e \leq 1 & \forall e \in E & \text{(capacity constraints)} &
\end{array}$$

The connectivity constraints are specified by a function f on vertex subsets, e.g. in survivable network design $f(S) := \max_{u,v} \{f_{u,v} \mid u \in S, v \notin S\}$. The matrix L_x is the Laplacian matrix of the fractional solution x . More explanations about this convex program can be found in Section 4.1.1.

Our first result for network design is the following approximation algorithm for this general problem. We remark that the degree constraints are not handled in the following result.

Theorem 1.2 (Informal). *Suppose we are given an optimal solution x to the convex program (CP). There is a polynomial time randomized algorithm to return an integral solution z to (CP) that simultaneously satisfies all the connectivity constraints, the effective resistance constraints, the spectral constraints, the algebraic connectivity constraint and the capacity constraints exactly with high probability. The objective value of the integral solution z is*

$$\langle c, z \rangle \leq (1 + O(\epsilon)) \cdot \text{cp} + O\left(\frac{nc_\infty}{\epsilon}\right)$$

with high probability, where n is the number of vertices in the graph and $c_\infty := \|c\|_\infty$ is the maximum cost of an edge. Furthermore, the linear packing constraints and the linear covering constraints are satisfied approximately with high probability (see Theorem 1.8 and Theorem 4.2 for the approximation guarantees for these constraints).

Note that this provides a $(1 + O(\epsilon))$ -approximation algorithm if $\text{cp} \gtrsim nc_\infty/\epsilon^2$, and a constant factor approximation algorithm if $\text{cp} \gtrsim nc_\infty$. We remark that, for survivable network design, the $(1 + O(\epsilon))$ -approximation algorithm does not improve on the 2-approximation algorithm of Jain's result, as Jain's algorithm always returns a solution with cost at most $\text{cp} + 2nc_\infty$.

The main advantage of the spectral approach is that it significantly extends the scope of useful properties that can be incorporated into network design, while previously there are no known non-trivial approximation algorithms even for some individual constraints. We demonstrate the use of Theorem 1.2 with one concrete setting.

Example 1.3. *Suppose the connectivity requirement satisfies $f_{u,v} \geq k$ for all $u, v \in V$ (e.g. to find a k -edge-connected subgraph). Assume the cost c_e of each edge e satisfies $1 \leq c_e \leq O(k)$. Then Theorem 1.2 provides a constant factor approximation algorithm for this survivable network design problem. To our knowledge, the only known constant factor approximation algorithm even restricted to this special case is Jain's iterative rounding algorithm. The algorithm in Theorem 1.2 provides a completely different spectral algorithm to achieve constant factor approximation in this special case.*

Furthermore, the constant factor approximation algorithm can be achieved while incorporating additional effective resistance constraints (e.g. to upper bound commute times between pairs of vertices), spectral constraints (e.g. to dominate another graph/topology in terms of the number of edges in cuts), algebraic connectivity constraint (e.g. to lower bound graph expansion). Also, additional linear packing and covering constraints can be satisfied approximately, even when they are unstructured. See Section 4.1 for a more in-depth discussion.

Recently, Bansal [10] designed a rounding technique that achieves the guarantees by iterative rounding and randomized rounding simultaneously, and he showed various interesting applications of his techniques. However, he left it as an open question whether there is an $O(1)$ -approximation algorithm for survivable network design while satisfying some concentration property of the output. Theorem 1.2 provides some progress towards his question (e.g. in the setting in Example 1.3), as the guarantees on the linear packing and linear covering constraints satisfy some concentration property as shown in Theorem 4.2. We defer to Section 4.1.5 for details.

Our second result for network design is a strong upper bound on the integrality gap of the convex program that incorporates degree constraints as well, assuming the fractional solution x satisfies some additional properties.

Theorem 1.4 (Informal). *Suppose we are given a solution x to the convex program (CP). Assume that $\text{Reff}_x(u, v) \leq \epsilon^2$ for every $uv \in E$ and $c_\infty \leq \epsilon^2 \langle c, x \rangle$ for some $\epsilon \in [0, 1]$. Then, there exists an integral solution z that approximately satisfies all the connectivity constraints, degree constraints, effective resistance constraints, spectral constraints, algebraic connectivity constraints, and capacity constraints with $\langle c, z \rangle \leq (1 + O(\epsilon)) \langle c, x \rangle$.*

We remark that Theorem 1.4 does not provide a polynomial time algorithm to find such an integral solution, as it is proved using the non-constructive results in discrepancy theory. Also, we note that Theorem 1.4 does not handle linear covering and packing constraints. The assumption $\text{Reff}_x(u, v) \leq \epsilon^2$ for every $uv \in E$ may not be satisfied in applications, and we will explain in Section 4.1.4 when it will be satisfied and show that it is not too restrictive.

1.2 Previous Work on Spectral Rounding

The most relevant works for spectral rounding are from spectral sparsification and discrepancy theory. There are two previous theorems that imply non-trivial results for spectral rounding.

1.2.1 Spectral Sparsification

There are various algorithms for spectral sparsifications, by random sampling [68], by barrier functions [15], by regret minimization [3, 65], and by some combinations of these ideas [51, 50]. Most of these algorithms need to work with arbitrary weights and cannot guarantee that the output subgraph has only integral weights. There are some algorithms which guarantee that the output has only integral weights, but they only achieve considerably weaker spectral approximation [7, 3, 14].

Allen-Zhu, Li, Singh, and Wang [5] formulated and proved the following spectral rounding theorem, using the framework of regret minimization developed for spectral sparsification [3].

Theorem 1.5 ([5]). *Let $v_1, v_2, \dots, v_m \in \mathbb{R}^n$, $x \in [0, 1]^m$ and $k = \sum_{i=1}^m x_i$. Suppose $\sum_{i=1}^m x_i v_i v_i^T = I_n$ and $k \geq 5n/\epsilon^2$ for some $\epsilon \in (0, \frac{1}{3}]$. Then there is a polynomial time algorithm to return a subset $S \subseteq [m]$ with*

$$|S| \leq k \quad \text{and} \quad \sum_{i \in S} v_i v_i^T \succcurlyeq (1 - 3\epsilon)I_n.$$

Theorem 1.5 can be understood as a one-sided spectral rounding result, where the fractional solution x is rounded to a zero-one solution while the budget constraint is satisfied and the spectral lower bound is approximately satisfied. Through a general reduction, this theorem implies near-optimal approximation algorithms for a large class of experimental design problems [5].

We remark that Theorem 1.5 can be modified to prove similar but more restrictive results for network design as in Theorem 1.2, when the objective function c is the all-one vector and there are no linear covering and packing constraints. This already extends the scope of unweighted network design significantly, but this connection was not made before. For network design, it is desirable to have different costs on edges, and these weighted problems are usually more difficult to solve than the unweighted problems (e.g. minimum k -edge-connected subgraphs [35] vs [43], minimum bounded degree spanning trees [33] vs [39], etc).

1.2.2 Discrepancy Theory

The techniques in spectral sparsification have been extended greatly to prove discrepancy theorems in spectral settings [57, 6, 45], most notably in the solution to Weaver’s conjecture that resolves the Kadison-Singer problem [56, 57] and its extension and surprising application to the asymmetric traveling salesman problem [6]. The following recent result by Kyng, Luh, and Song [45] provides the most refined formulation in the discrepancy setting, using the method of interlacing polynomials and the barrier arguments developed in [57, 6].

Theorem 1.6 ([45]). *Let $v_1, \dots, v_m \in \mathbb{R}^n$, and ξ_1, \dots, ξ_m be independent random scalar variables with finite support. There exists a choice of outcomes $\epsilon_1, \dots, \epsilon_m$ in the support of ξ_1, \dots, ξ_m such that*

$$\left\| \sum_{i=1}^m \mathbb{E}[\xi_i] v_i v_i^T - \sum_{i=1}^m \epsilon_i v_i v_i^T \right\|_{\text{op}} \leq 4 \left\| \sum_{i=1}^m \text{Var}[\xi_i] (v_i v_i^T)^2 \right\|_{\text{op}}^{1/2}.$$

We note that Theorem 1.6 implies the following two-sided spectral rounding result, which is very similar to Corollary 1.7 in [45] but with a weaker assumption, where we only need $\|\sum_{i=1}^m x_i v_i v_i^T\|_{\text{op}} \leq 1$ instead of $\|\sum_{i=1}^m v_i v_i^T\|_{\text{op}} \leq 1$ as in [45]. The proof will be presented in Section 3.2 in a more general setting.

Corollary 1.7. *Let $v_1, \dots, v_m \in \mathbb{R}^n$ and $x \in [0, 1]^m$. Suppose $\sum_{i=1}^m x_i v_i v_i^T = I_n$ and $\|v_i\| \leq \epsilon$ for all $i \in [m]$. Then there exists a subset $S \subseteq [m]$ satisfying*

$$(1 - O(\epsilon))I_n \preceq \sum_{i \in S} v_i v_i^T \preceq (1 + O(\epsilon))I_n.$$

Comparing to Theorem 1.5, the advantage of Corollary 1.7 is that it provides a two-sided spectral approximation. On the other hand, Corollary 1.7 requires the assumption that all vectors are short, and it has no guarantee on the size of S . Also, it is important to point out that the proof of Corollary 1.7 does not provide a polynomial time algorithm to find such a subset.

1.3 Our Technical Contributions

We extend the previous results on spectral rounding to incorporate non-negative linear constraints and to satisfy the requirements for network design problems. These results have interesting applications in many other problems besides network design; see Section 1.4 and Section 4.

Our main technical result considers one-sided spectral rounding.

Theorem 1.8. *Suppose we are given $v_1, \dots, v_m \in \mathbb{R}^n$ and $x \in [0, 1]^m$ such that $\sum_{i=1}^m x_i v_i v_i^T = I_n$. For any $\epsilon \in (0, \frac{1}{4})$, there is a polynomial time randomized algorithm that returns a solution $z \in \{0, 1\}^m$ such that*

$$\sum_{i=1}^m z_i v_i v_i^T \succcurlyeq I_n$$

with probability at least $1 - \exp(-\Omega(n))$. Furthermore, for any $c \in \mathbb{R}_+^m$, the solution z satisfies the upper bound

$$\langle c, z \rangle \leq (1 + 6\epsilon) \langle c, x \rangle + \frac{15nc_\infty}{\epsilon}$$

with probability at least $1 - \exp(-\Omega(n))$, and the solution z satisfies the lower bound

$$\langle c, z \rangle \geq \langle c, x \rangle - \delta nc_\infty$$

with probability at least $1 - \exp(-\Omega(\min\{\epsilon\delta, \epsilon\delta^2\} \cdot n))$ for $\delta > 0$.

The main advantage of Theorem 1.8 over Theorem 1.5 is that we can prove that $\langle c, z \rangle$ is not too far from $\langle c, x \rangle$ for an arbitrary vector $c \in \mathbb{R}_+^m$ with high probability. This allows us to bound the cost of the returned solution to network design problems, and when $nc_\infty \lesssim \langle c, x \rangle$ we can conclude that z is a constant factor approximate solution. Note that the guarantee on linear constraints can be applied to up to exponentially many constraints. This allows us to incorporate additional linear packing and covering constraints into network design and have some non-trivial guarantees. Another advantage is that we construct a solution that satisfies the spectral lower bound exactly, by allowing the solution to choose more than $k = \sum_{i=1}^m x_i$ vectors. This is important in network design problems where we would like to construct a solution that satisfies all the constraints (instead of approximately satisfying all the constraints), by allowing the cost of the solution to be higher than the cost of the optimal solutions.

We note that there are examples showing that the additive error term $O(nc_\infty/\epsilon)$ in Theorem 1.8 is tight up to a constant factor (see Section 3.3).

Using the proof techniques in Theorem 1.8, we can strengthen a recent deterministic algorithm by Bansal, Svensson and Trevisan [14] to construct unweighted spectral sparsifiers, to ensure that there will be no parallel edges in the sparsifier. See Section 4.4 for details.

For two-sided spectral rounding, we show that Corollary 1.7 can be extended to incorporate one given non-negative linear constraint.

Theorem 1.9. *Let $v_1, \dots, v_m \in \mathbb{R}^n$, $x \in [0, 1]^m$ and $c \in \mathbb{R}_+^m$. Suppose $\sum_{i=1}^m x_i v_i v_i^T = I_n$, $\|v_i\| \leq \epsilon < \frac{1}{8}$ for all $i \in [m]$ and $c_\infty \leq \epsilon^2 \langle c, x \rangle$. Then there exists $z \subseteq \{0, 1\}^m$ such that*

$$(1 - 8\epsilon)I_n \preceq \sum_{i=1}^m z_i v_i v_i^T \preceq (1 + 8\epsilon)I_n \quad \text{and} \quad (1 - 8\epsilon)\langle c, x \rangle \leq \langle c, z \rangle \leq (1 + 8\epsilon)\langle c, x \rangle.$$

Note that the linear constraint c in Theorem 1.9 is required to be given as part of the input, while it is not required so in Theorem 1.8. Theorem 1.9 is useful in bounding the integrality gap for convex programs for network design problems, showing strong approximation results when the assumptions are satisfied (see Section 4.1.4). Also, we will show in Section 4.4 that it can be used in the study of additive unweighted spectral sparsification [14], proving an optimal existential result.

1.3.1 Techniques

The main technical contribution is an iterative randomized rounding algorithm for Theorem 1.8. Our algorithm is based on the regret minimization framework developed in [3, 5] for spectral sparsification and one-sided spectral rounding. Let us first review the previous work. To prove Theorem 1.5, Allen-Zhu, Li, Singh, and Wang [5] analyzed a local search algorithm where they start from an arbitrary subset S_0 of k vectors, and in each iteration $t \geq 1$ they find a pair of vectors $i \in S_{t-1}$ and $j \notin S_{t-1}$ so that roughly speaking $\lambda_{\min}(\sum_{l \in S_{t-1}-i+j} v_l v_l^T) > \lambda_{\min}(\sum_{l \in S_{t-1}} v_l v_l^T)$, and then they set $S_t = S_{t-1} - i_t + j_t$. Using the framework of regret minimization, with the $l_{1/2}$ -regularizer introduced in [3], they proved that the task of finding a pair to improve the minimum

eigenvalue can be reduced to finding a pair $i_t \in S_{t-1}$ and $j_t \notin S_{t-1}$ so that

$$\frac{\langle v_{j_t} v_{j_t}^T, A_t \rangle}{1 + 2\alpha \langle v_{j_t} v_{j_t}^T, A_t^{1/2} \rangle} - \frac{\langle v_{i_t} v_{i_t}^T, A_t \rangle}{1 - 2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle} \geq \Delta > 0, \quad (1.1)$$

where A_t is the matrix defined in (2.1) based on the current solution S_{t-1} . Using a delicate argument, they proved that if $i_t \in S_{t-1}$ (subjecting to the restriction that $2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle < 1$) is chosen to minimize $\langle v_{i_t} v_{i_t}^T, A_t \rangle / (1 - 2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle)$ and $j_t \notin S_{t-1}$ is chosen to maximize $\langle v_{j_t} v_{j_t}^T, A_t \rangle / (1 + 2\alpha \langle v_{j_t} v_{j_t}^T, A_t^{1/2} \rangle)$, then this pair i_t, j_t satisfies the above inequality with $\Delta = \epsilon/k$ as long as $\lambda_{\min}(\sum_{l \in S_{t-1}} v_l v_l^T) \leq 1 - 3\epsilon$. This implies, by the regret minimization framework, that the local search algorithm will succeed to find a solution S_τ with $\lambda_{\min}(\sum_{l \in S_\tau} v_l v_l^T) \geq 1 - 3\epsilon$ within $\tau \leq k/\epsilon$ iterations. We will review more about the regret minimization framework in Section 2.3.

To incorporate non-negative linear constraints, our idea is to turn the deterministic local search algorithm into an iterative randomized rounding algorithm. In this randomized rounding algorithm, we first construct an initial solution S_0 by adding each i into S_0 with probability x_i independently. This will ensure that $c(S_0) \approx \langle c, x \rangle$ with high probability. In each iteration $t \geq 1$, based on the current solution S_{t-1} , we construct a probability distribution to sample a vector v_{i_t} to be removed from S_{t-1} , and a probability distribution to sample a vector v_{j_t} to be added to S_{t-1} . To maintain $c(S_t) \approx \langle c, x \rangle$, the basic idea is to remove a vector v_i with probability proportional to $1 - x_i$ and add a vector v_j with probability proportional to x_j , but doing so will not satisfy the spectral lower bound with high probability. Instead, we prove that if we recompute the sampling probability so that a vector v_i is removed with probability proportional to $(1 - x_i)(1 - 2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle)$ and a vector v_j is added with probability proportional to $x_j(1 + 2\alpha \langle v_j v_j^T, A_t^{1/2} \rangle)$, then (1.1) is satisfied with expected progress $\mathbb{E}[\Delta] \geq \epsilon/k$ as long as $\lambda_{\min}(\sum_{l \in S_{t-1}} v_l v_l^T) \leq 1 - 2\epsilon$. Informally, a vector pointing to a direction that is not well covered by the current solution is more likely to be added and less likely to be removed, to ensure that the spectral lower bound will be satisfied. However, this changes the expectation on the linear constraint, but we can bound the error by the additive term $O(nc_\infty/\epsilon)$. Note that there are examples showing that this additive error is unavoidable if our goal is to satisfy the spectral lower bound exactly (see Section 3.3), so our analysis is tight up to a constant factor. Compared to the deterministic approach in [5], this randomized approach uses the fractional solution x more crucially in the rounding procedure, and we note that it can be used to give a simpler proof of the deterministic local search algorithm in [5] (see Remark 3.6).

The advantage of the randomized approach is that we can prove that the random variables are concentrated around their expected values, so that we can handle multiple non-negative linear constraints simultaneously. Since the sampling probabilities change over time based on the previous samples, the random variables that we consider are not a sum of independent random variables and thus Chernoff type bounds cannot be applied. For the spectral lower bound, we will define a martingale and use Freedman's inequality to prove that the total progress we make in (1.1) is concentrated around its expected value. For the non-negative linear constraints, we show that they satisfy an interesting "self-adjusting" property, such that if $c(S_t) - \langle c, x \rangle$ is (more) positive then $\mathbb{E}[c(S_{t+1})] - c(S_t)$ is (more) negative and vice versa, so intuitively $c(S_t) \approx \langle c, x \rangle$ with high probability for any t . This sequence of random variables is not a martingale and so Freedman's

inequality cannot be applied. Instead, we prove a new concentration inequality for this self-adjusting process that provides a quantitative bound similar to that in Freedman’s inequality. We note that the iterative randomized rounding algorithm does not even need to know the linear constraint c in advance in order to return a solution S with $c(S) \approx \langle c, x \rangle$. This property is quite similar to that of a recent rounding algorithm by Bansal [10] combining iterative rounding and randomized rounding as we will discuss in Section 4.1.5.

We remark that our approach to turn a deterministic algorithm into a randomized algorithm is inspired by the fast algorithm for spectral sparsification by Lee and Sun [51], where they turned the deterministic algorithm by Batson, Spielman and Srivastava [15] into a randomized algorithm that recomputes the sampling probabilities in different phases. In their algorithm, the advantage of the randomized algorithm is to sample many vectors in parallel instead of carefully choosing one vector at a time as in [15]. In our algorithm, the advantage of the randomized algorithm is to approximately preserves many linear constraints simultaneously using arguments about expectation and concentration, while it is not clear how to modify the proofs in the deterministic local search algorithm in [5] to prove that there is always a pair of vectors v_i, v_j which makes enough progress in (1.1) and at the same time $c_j - c_i$ is small, even if there is just have one constraint c and it is given in advance. We believe that this probabilistic approach will be useful in designing algorithms using the regret minimization framework.

1.4 Other Applications

The spectral rounding results are quite general and have many other applications besides network design. We mention some of these results and defer the details to Section 4.

1.4.1 Weighted Experimental Design

Experimental design is an important class of problems in statistics and has found new applications in machine learning [8, 61]. The one-sided spectral rounding result of Allen-Zhu, Li, Singh and Wang [5] was used to give near optimal approximation algorithms for many well-known experimental design problems. We will explain these previous work in Section 4.2, and show that our results can be used to design approximation algorithms for the more general setting where different experiments may have different costs while incorporating some additional linear constraints; see Theorem 4.11 and the discussions thereafter.

Theorem 1.10 (Informal). *We are given m design points that are represented by n -dimensional vectors $v_1, \dots, v_m \in \mathbb{R}^n$, a cost vector $c \in \mathbb{R}_+^m$ and a cost budget $C \in \mathbb{R}_+$. For any $\epsilon \in (0, \frac{1}{2}]$, if $C \geq 15nc_\infty/\epsilon^2$, there is a randomized polynomial time algorithm that returns a subset of vectors with total cost at most C so that the objective value of A/D/E/V/G-design is at most $(1 + O(\epsilon))$ times of that of the optimal solution.*

1.4.2 Spectral Network Design

There are several previous work on network design problems with spectral requirements, including maximizing algebraic connectivity [37, 44], minimizing total effective resistances [38], and network design for s - t effective resistances [20]. In Section 4.3, we will see that these problems are special cases of the general network design problem and the weighted experimental design problem in Section 4.1 and Section 4.2, and our results provide improved approximation algorithms for these problems and also generalize these problems to incorporate many additional constraints.

We provide the first non-trivial approximation algorithm for the problem of maximizing algebraic connectivity subject to a knapsack constraint, proposed by Ghosh and Boyd [37].

Theorem 1.11. *Let $G = (V, E)$ be a graph where each edge has cost c_e and C be a given cost budget. Suppose $C \geq 15|V|c_\infty/\epsilon^2$ for some $\epsilon \leq 1/2$. There is a randomized polynomial time algorithm which returns a subgraph H of G with*

$$\sum_{e \in H} c_e \leq C \quad \text{and} \quad \lambda_2(L_H) \geq (1 - O(\epsilon))\lambda_{\text{opt}},$$

where λ_{opt} is the maximum λ_2 that can be achieved by a solution with cost at most C .

We also provide a similar result for the problem of minimizing total effective resistance, proposed by Ghosh, Boyd and Saberi [38].

Theorem 1.12. *Let $G = (V, E)$ be a graph where each edge has cost c_e and C be a given cost budget. Suppose $C \geq 15|V|c_\infty/\epsilon^2$ for some $\epsilon \leq 1/2$. There is a randomized polynomial time algorithm which returns a subgraph H of G with*

$$\sum_{e \in H} c_e \leq C \quad \text{and} \quad \sum_{u,v} \text{Reff}_H(u, v) \leq (1 + O(\epsilon))\text{opt},$$

where opt is the minimum total effective resistance that can be achieved by a solution with cost at most C .

These results can be extended to incorporate additional constraints (e.g. connectivity constraints). See Section 4.3 for details about these results, including the related work [44, 59].

1.4.3 Additive Spectral Sparsification

Recently, Bansal, Svensson and Trevisan [14] study whether there is a non-trivial notion of unweighted spectral sparsification with which linear-sized spectral sparsification is always possible. They provide randomized and deterministic algorithms to construct “additive” unweighted spectral sparsifiers, a notion suggested by Oveis Gharan. In Section 4.4, we will explain their results and show that our spectral rounding results can be applied to this problem. Using Theorem 1.9, we prove an optimal existential result for the problem.

Theorem 1.13. *Suppose we are given a graph $G = (V, E)$ with n vertices, m edges, and maximum degree d . Let $\tilde{m} = n/\epsilon^2$. For any $\epsilon \in (0, 1)$, there exists a subset of edges $F \subseteq E$ with $|F| \leq 8n/\epsilon^2$ such that*

$$-8\sqrt{2}\epsilon d I_n \preceq L_G - \frac{m}{\tilde{m}} \sum_{e \in F} b_e b_e^T \preceq 8\sqrt{2}\epsilon d I_n.$$

Using the proof techniques in Theorem 1.8, we provide an improved deterministic algorithm to construct additive unweighted spectral sparsifiers with no parallel edges (where the result in [14] may produce parallel edges).

Theorem 1.14. *Given a graph $G = (V, E)$ with n vertices, m edges, maximum degree d , and $\epsilon \in (0, 1)$, there is a polynomial time deterministic algorithm that finds a subset F of edges with size $\tilde{m} = |F| = O(n/\epsilon^2)$ such that $\tilde{G} = (V, F)$ satisfies*

$$\frac{2m}{\tilde{m}} D_{\tilde{G}} - 2D_G - \epsilon d I \preceq \frac{m}{\tilde{m}} L_{\tilde{G}} - L_G \preceq \epsilon d I,$$

where D_G and $D_{\tilde{G}}$ are the diagonal degree matrix of G and \tilde{G} respectively.

2 Preliminaries

We review some basic linear algebra and spectral graph theory in Section 2.1 and Section 2.2. Then we review the regret minimization framework for one-sided spectral rounding in Section 2.3, and state some concentration inequalities for the analysis of our randomized algorithm in Section 2.4.

2.1 Linear Algebra

We write \mathbb{R} and \mathbb{R}_+ as the sets of real numbers and non-negative real numbers, and \mathbb{Z} and \mathbb{Z}_+ as the sets of integers and non-negative integers.

All the vectors in this paper only have real entries. Let \mathbb{R}^n denote the n -dimensional Euclidean space. We write $\vec{1}_n$ as the n -dimensional all-one vector. Given a vector x , we write $\|x\|$ as its ℓ_2 -norm, $\|x\|_1$ as its ℓ_1 -norm, and $\|x\|_\infty$ as its ℓ_∞ -norm. A vector $v \in \mathbb{R}^n$ is a column vector, and its transpose is denoted by v^T . Given two vectors $x, y \in \mathbb{R}^n$, the inner product is defined as $\langle x, y \rangle := \sum_{i=1}^n x_i y_i$. The Cauchy-Schwarz inequality says that $\langle x, y \rangle \leq \|x\| \|y\|$.

We write I_n as the $n \times n$ identity matrix, and J_n as the $n \times n$ all-one matrix. All matrices considered in this paper are real symmetric matrices. It is a fundamental result that any $n \times n$ real symmetric matrix has n real eigenvalues $\lambda_1 \leq \dots \leq \lambda_n$ and an orthonormal basis of eigenvectors. We write $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ as the maximum and the minimum eigenvalue of a matrix M . The trace of a matrix M , denoted by $\text{tr}(M)$, is defined as the sum of the diagonal entries of M . It is well-known that $\text{tr}(M) = \sum_{i=1}^n \lambda_i(M)$ where $\lambda_i(M)$ denotes the i -th eigenvalue of M .

A matrix M is a positive semidefinite (PSD) matrix, denoted as $M \succeq 0$, if M is symmetric and all the eigenvalues are nonnegative, or equivalently, the quadratic form $x^T M x \geq 0$ for any vector

x . We use $A \succcurlyeq B$ to denote $A - B \succcurlyeq 0$ for matrices A and B . We write \mathbb{S}_+^n as the set of all n -dimensional PSD matrices. Let $M \succcurlyeq 0$ be a PSD matrix with eigendecomposition $M = \sum_i \lambda_i v_i v_i^T$, where $\lambda_i \geq 0$ is the i -th eigenvalue and v_i is the corresponding eigenvector. The square root of M is $M^{1/2} := \sum_i \sqrt{\lambda_i} v_i v_i^T$.

Given two matrices A and B of the same size, the Frobenius inner product of A, B is denoted as $\langle A, B \rangle := \sum_{i,j} A_{ij} B_{ij} = \text{tr}(A^T B)$. The following are two standard facts

$$A, B \succcurlyeq 0 \implies \langle A, B \rangle \geq 0 \quad \text{and} \quad A \succcurlyeq 0, B \succcurlyeq C \succcurlyeq 0 \implies \langle A, B \rangle \geq \langle A, C \rangle.$$

We write $\|M\|_{\text{op}} := \max_{\|x\|=1} \|Mx\|$ as the operator norm of a matrix M . For symmetric matrices, the operator norm is just the largest absolute value of its eigenvalues. For positive semidefinite matrices, the operator norm is just its largest eigenvalue.

2.2 Graphs and Laplacian Matrices

Let $G = (V, E)$ be an undirected graph with edge weight $x_e \geq 0$ on each edge $e \in E$. The number of vertices and the number of edges are denoted by $n := |V|$ and $m := |E|$. For a subset of edges $F \subseteq E$, the total weight of edges in F is $x(F) := \sum_{e \in F} x_e$. For a subset of vertices $S \subseteq V$, the set of edges with one endpoint in S and one endpoint in $V - S$ is denoted by $\delta(S)$. For a vertex v , the set of edges incident on a vertex v is $\delta(v) := \delta(\{v\})$, and the weighted degree of v is $\deg(v) := x(\delta(v))$. The expansion of a set $\phi(S) := |\delta(S)|/|S|$ is defined as the ratio of the number of edges on the boundary of S to the size of S . The expansion of a graph G is defined as $\phi(G) := \min_{0 \leq |S| \leq \frac{n}{2}} \phi(S)$.

The adjacency matrix $A \in \mathbb{R}^{n \times n}$ of the graph is defined as $A_{u,v} = x_{u,v}$ for all $u, v \in V$. The Laplacian matrix $L \in \mathbb{R}^{n \times n}$ of the graph is defined as $L = D - A$ where $D \in \mathbb{R}^{n \times n}$ is the diagonal degree matrix with $D_{u,u} = \deg(u)$ for all $u \in V$. Similarly, the signless Laplacian matrix $L^+ \in \mathbb{R}^{n \times n}$ is defined as $L^+ = D + A$. For each edge $e = uv \in E$, let $b_e := \chi_u - \chi_v$ where $\chi_u \in \mathbb{R}^n$ is the vector with one in the u -th entry and zero otherwise. The Laplacian matrix with respect to weights x can be written as

$$L_x := \sum_{e \in E} x_e b_e b_e^T.$$

Let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be the eigenvalues of L with corresponding orthonormal eigenvectors v_1, v_2, \dots, v_n so that $L = \sum_{i=1}^n \lambda_i v_i v_i^T$. It is well-known that the Laplacian matrix is positive semidefinite, $\lambda_1 = 0$ with $v_1 = \mathbf{1}/\sqrt{n}$ as the corresponding eigenvector, and $\lambda_2 > 0$ if and only if G is connected. The following fact is useful for eigenvalue maximization.

Fact 2.1 ([37]). $\lambda_2(L_x)$ is a concave function with respect to x for $x \geq 0$.

The pseudo-inverse of the Laplacian matrix L of a connected graph is defined as

$$L^\dagger = \sum_{i=2}^n \frac{1}{\lambda_i} v_i v_i^T,$$

which maps every vector b orthogonal to v_1 to a vector y such that $Ly = b$. The effective resistance between two vertices s and t on a graph G with weight x is defined as

$$\text{Reff}_x(s, t) := b_{st}^T L_x^\dagger b_{st}.$$

We will use the following fact for the formulation of the convex programming relaxation in (CP).

Fact 2.2 ([38]). *$\text{Reff}_x(s, t)$ is a convex function with respect to the weights x for $x \geq 0$.*

2.3 Regret Minimization and Spectral Rounding

We use the regret minimization framework developed by Allen-Zhu, Liao and Orecchia for spectral sparsification [3] and present the results in [3, 5]. This is an online optimization setting. In each iteration t , the player chooses an action matrix A_t from the set of density matrices $\Delta_n = \{A \in \mathbb{R}^{n \times n} \mid A \succcurlyeq 0, \text{tr}(A) = 1\}$. We can interpret the player action as choosing a probability distribution over the set of unit vectors. The player then observes a feedback matrix F_t and incurs a loss of $\langle A_t, F_t \rangle$. After τ iterations, the regret of the player is defined as

$$R_\tau := \sum_{t=1}^{\tau} \langle A_t, F_t \rangle - \inf_{B \in \Delta_n} \sum_{t=1}^{\tau} \langle B, F_t \rangle = \sum_{t=1}^{\tau} \langle A_t, F_t \rangle - \lambda_{\min} \left(\sum_{t=1}^{\tau} F_t \right),$$

which is the difference between the loss of the player actions and the loss of the best fixed action B , that can be assumed to be a rank one matrix vv^T . The objective of the player is to minimize the regret. A well-known algorithm for regret minimization is Follow-The-Regularized-Leader which plays the action

$$A_t = \operatorname{argmin}_{A \in \Delta_n} \left\{ w(A) + \alpha \cdot \sum_{l=0}^{t-1} \langle A, F_l \rangle \right\},$$

where $w(A)$ is a regularization term and α is a parameter called the learning rate that balances the loss and the regularization. Note that F_0 is an initial feedback which is given before the game started. Different choice of regularization gives different algorithm for regret minimization. One choice is the entropy regularizer $w(A) = \langle A, \log A - I \rangle$ and this gives the well-known matrix multiplicative update algorithm. The choice that we will use is the $\ell_{1/2}$ -regularizer $w(A) = -2 \text{tr}(A^{1/2})$ introduced in [3], which plays the action

$$A_t = \left(l_t I + \alpha \sum_{l=0}^{t-1} F_l \right)^{-2}, \quad (2.1)$$

where l_t is the unique constant that ensures $A_t \in \Delta_n$. Allen-Zhu, Liao and Orecchia [3] prove upper bounds on the regret of this algorithm for positive or negative semidefinite feedback matrices.

Theorem 2.3 (Theorem 3.2 and 3.3 in [3]). *Suppose $F_0 = 0$ and each feedback matrix $F_t \in \mathbb{R}^{n \times n}$ is either a positive or negative semidefinite matrix with $\alpha A_t^{1/4} F_t A_t^{1/4} \succcurlyeq -\frac{1}{4} I$ for all $t \geq 1$, and the action matrix $A_t \in \mathbb{R}^{n \times n}$ is of the form in (2.1) for some $\alpha > 0$. Then*

$$R_\tau \leq O(\alpha) \sum_{t=1}^{\tau} \langle A_t, |F_t| \rangle \cdot \left\| A_t^{1/4} F_t A_t^{1/4} \right\|_{\text{op}} + \frac{2\sqrt{n}}{\alpha}.$$

When each feedback matrix F_t is of the form $u_t u_t^T$ for some $u_t \in \mathbb{R}^n$ for all $t \geq 1$, it holds that

$$\lambda_{\min} \left(\sum_{t=1}^{\tau} u_t u_t^T \right) \geq \sum_{t=1}^{\tau} \frac{\langle u_t u_t^T, A_t \rangle}{1 + \alpha \langle u_t u_t^T, A_t^{1/2} \rangle} - \frac{2\sqrt{n}}{\alpha}. \quad (2.2)$$

For one-sided spectral rounding, the goal is to choose a subset S of vectors to maximize $\lambda_{\min}(\sum_{i \in S} v_i v_i^T)$. Using this regret minimization framework, the second part of Theorem 2.3 reduces this problem to the simpler task of finding a vector u_t that maximizes $\langle u_t u_t^T, A_t \rangle / (1 + \alpha \langle u_t u_t^T, A_t^{1/2} \rangle)$. Using the condition that $\sum_{i=1}^m x_i v_i v_i^T = I_n$ and $\sum_{i=1}^m x_i = k$, it can be shown [4] that there is always a vector v_j with $\langle v_j v_j^T, A_t \rangle / (1 + \alpha \langle v_j v_j^T, A_t^{1/2} \rangle) \geq 1/(k + \alpha\sqrt{n})$. Setting $\alpha = \sqrt{n}/\epsilon$ and $\tau = k$ and using the assumption that $k \geq n/\epsilon^2$, this gives $\lambda_{\min}(\sum_{t=1}^k u_t u_t^T) \geq 1 - 3\epsilon$ and proves Theorem 1.5 in the easier setting when a vector can be chosen more than once (i.e. the with repetition setting in experimental design). This greedy algorithm can be extended to the more difficult setting when every vector can be chosen at most once, but only achieving a $\Theta(1)$ -approximation [4].

To prove Theorem 1.5 when the output must be a zero-one solution, Allen-Zhu, Li, Singh and Wang [5] derived the following regret minimization bound for rank two feedback matrices.

Theorem 2.4 (Lemma 2.5 and 2.7 in [5]). *Suppose the action matrix $A_t \in \mathbb{R}^{n \times n}$ is of the form in (2.1) for some $\alpha > 0$. Suppose the initial feedback matrix $F_0 \in \mathbb{S}^n$ is a symmetric matrix, and for all $t \geq 1$ each feedback matrix F_t is of the form $v_{j_t} v_{j_t}^T - v_{i_t} v_{i_t}^T$ for some $v_{j_t}, v_{i_t} \in \mathbb{R}^n$ such that $\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle < \frac{1}{2}$, then*

$$\lambda_{\min} \left(\sum_{t=0}^{\tau} F_t \right) \geq \sum_{t=1}^{\tau} \left(\frac{\langle v_{j_t} v_{j_t}^T, A_t \rangle}{1 + 2\alpha \langle v_{j_t} v_{j_t}^T, A_t^{1/2} \rangle} - \frac{\langle v_{i_t} v_{i_t}^T, A_t \rangle}{1 - 2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle} \right) - \frac{2\sqrt{n}}{\alpha}.$$

With Theorem 2.4, they analyzed a deterministic local search algorithm where they start from an arbitrary solution S_0 of k vectors, and in each iteration $t \geq 1$ they find a $j_t \notin S_{t-1}$ that maximizes $\langle v_{j_t} v_{j_t}^T, A_t \rangle / (1 + 2\alpha \langle v_{j_t} v_{j_t}^T, A_t^{1/2} \rangle)$ and an $i_t \in S_{t-1}$ that minimizes $\langle v_{i_t} v_{i_t}^T, A_t \rangle / (1 - 2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle)$ subjecting to the restriction that $2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle < 1$, and define $S_t := S_{t-1} - i_t + j_t$ as the new solution. Using a delicate argument, they proved that so long as $\lambda_{\min}(\sum_{l \in S_{t-1}} v_l v_l^T) \leq 1 - 3\epsilon$, the pair i_t, j_t always satisfies

$$\frac{\langle v_{j_t} v_{j_t}^T, A_t \rangle}{1 + 2\alpha \langle v_{j_t} v_{j_t}^T, A_t^{1/2} \rangle} - \frac{\langle v_{i_t} v_{i_t}^T, A_t \rangle}{1 - 2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle} \geq \frac{\epsilon}{k}.$$

This implies by Theorem 2.4 that the local search algorithm will succeed in finding a solution S_t with $\lambda_{\min}(\sum_{l \in S_t} v_l v_l^T) \geq 1 - 3\epsilon$ within k/ϵ iterations.

One technical point used in [4, 5] is that the partial solution $Z_{t-1} := \sum_{l=0}^{t-1} F_l$ at time t and the action matrix A_t at time t have the same eigenbasis due to (2.1). This allows one to bound $\langle Z_{t-1}, A_t \rangle$ and $\langle Z_{t-1}, A_t^{1/2} \rangle$ as follows.

Lemma 2.5 (Claim 2.11 in [5]). *Let $Z \succcurlyeq 0$ be an $n \times n$ positive semidefinite matrix and $A = (\alpha Z + lI)^{-2}$ for some $\alpha > 0$ where l is the unique constant such that A is a density matrix. Then, it holds that*

$$\langle Z, A \rangle \leq \frac{\sqrt{n}}{\alpha} + \lambda_{\min}(Z) \quad \text{and} \quad \alpha \langle Z, A^{1/2} \rangle \leq n + \alpha \sqrt{n} \cdot \lambda_{\min}(Z).$$

This lemma will be used in constructing a zero-one solution for Theorem 1.8. It will also be used in strengthening the result in [14] to guarantee that the unweighted additive spectral sparsifier returned by the regret minimization algorithm has no parallel edges.

2.4 Martingale and Concentration Inequalities

A sequence of random variables Y_1, \dots, Y_τ is a martingale with respect to a sequence of random variables Z_1, \dots, Z_τ if for all $t > 0$, it holds that

1. Y_t is a function of Z_1, \dots, Z_{t-1} ;
2. $\mathbb{E}[|Y_t|] < \infty$;
3. $\mathbb{E}[Y_{t+1} | Z_1, \dots, Z_t] = Y_t$.

We will use the following theorem by Freedman to bound the probability that Y_τ is large.

Theorem 2.6 ([31, 71]). *Let $\{Y_t\}_t$ be a real-valued martingale with respect to $\{Z_t\}_t$, and $\{X_t = Y_t - Y_{t-1}\}_t$ be the difference sequence. Assume that $X_t \leq R$ deterministically for $1 \leq t \leq \tau$. Let $W_t := \sum_{j=1}^t \mathbb{E}[X_j^2 | Z_1, \dots, Z_{j-1}]$ for $1 \leq t \leq \tau$. Then, for all $\delta \geq 0$ and $\sigma^2 > 0$,*

$$\Pr(\exists t \in [\tau] : Y_t \geq \delta \text{ and } W_t \leq \sigma^2) \leq \exp\left(\frac{-\delta^2/2}{\sigma^2 + R\delta/3}\right).$$

Recently, some variants of Freedman's inequality for martingales have been used to obtain algorithmic discrepancy results [12, 11]. In this paper, for the analysis of the non-negative linear constraints, we prove another variant which applies to non-martingales with a “self-adjusting” property, that if Y_t is (more) positive then $E[Y_{t+1}] - Y_t$ is (more) negative and vice versa. With this self-adjusting property, intuitively Y_t cannot be too far away from zero, and the following theorem provides a quantitative bound that is similar to that in Freedman's inequality.

Theorem 2.7. *Let $\{Y_t\}_t$ be a sequence of random variables, and $X_t := Y_t - Y_{t-1}$ be the difference sequence. Suppose that there exist $\gamma \in (0, \frac{1}{2}]$, $\beta_u, \beta_l \geq 0$ and $\sigma > 0$ such that the following properties hold for all $t \geq 1$.*

1. (Bounded difference:) $|X_t| \leq 1$ with probability one.
2. (Self adjusting:) $-\gamma Y_{t-1} - \beta_l \leq \mathbb{E}[X_t | Y_0, \dots, Y_{t-1}] \leq -\gamma Y_{t-1} + \beta_u$.
3. (Bounded variance:) $\mathbb{E}[X_t^2 | Y_0, \dots, Y_{t-1}] \leq \gamma Y_{t-1} + \sigma$.

4. (Initial concentration:) For any $a \in [-1, 1]$, the random variable Y_0 satisfies $\mathbb{E}[e^{aY_0}] \leq e^{a^2\sigma/\gamma}$.

Then, for any $\eta > 0$ and any $t \geq 0$, it holds that

$$\Pr \left[Y_t \geq \frac{\beta_u}{\gamma} + \eta \right] \leq \exp \left[-\frac{\eta^2}{4(\sigma + \beta_u)/\gamma + 2\eta} \right]$$

and

$$\Pr \left[Y_t \leq -\frac{\beta_l}{\gamma} - \eta \right] \leq \exp \left[-\frac{\eta^2}{4\sigma/\gamma + \eta} \right].$$

The proof of Theorem 2.7 will be presented in Section 3.1.4.

3 Spectral Rounding

We will first present the iterative randomized rounding algorithm for one-sided spectral rounding in Section 3.1. Then we will present the proof of Theorem 1.9 for two-sided spectral rounding in Section 3.2, and some examples showing the tightness of our results in Section 3.3.

3.1 Iterative Randomized Rounding for One-Sided Spectral Rounding

We modify the deterministic local search algorithm in [5] to an iterative randomized rounding algorithm so as to approximately satisfy arbitrary non-negative linear constraints. In this randomized algorithm, we first construct an initial solution S_0 by adding each vector v_i into S_0 with probability x_i independently. In each iteration $t \geq 1$, based on the current solution S_{t-1} , we construct a probability distribution to sample a vector v_{i_t} to be removed from S_{t-1} , and a probability distribution to sample a vector v_{j_t} to be added to S_{t-1} . The basic idea is that a vector v_i is removed with probability proportional to $1 - x_i$ and a vector v_j is added with probability proportional to x_j , but the probability is also adjusted based on the vector's contribution to the minimum eigenvalue of the current solution. We remark that it is possible that no vector is removed and/or no vector is added in an iteration. The algorithm stops when the minimum eigenvalue of the current solution is at least $1 - 2\epsilon$. The following is the formal description of the algorithm.

Iterative Randomized Swapping Algorithm

Input: $v_1, \dots, v_m \in \mathbb{R}^n$ and $x \in [0, 1]^m$ with $\sum_{i=1}^m x_i v_i v_i^T = I_n$, and an error parameter $\epsilon \in (0, \frac{1}{2})$.

Output: a subset $S \subseteq [m]$ such that $\sum_{i \in S} v_i v_i^T \succcurlyeq (1 - 2\epsilon)I_n$ and $c(S) \approx \langle c, x \rangle$ for any $c \in \mathbb{R}_+^m$ with high probability.

1. Initialization: $t := 1$, $S_0 := \emptyset$, $\alpha := \sqrt{n}/\epsilon$, $k := m + 2n/\epsilon$.
2. Add i into S_0 independently with probability x_i for each $i \in [m]$. Let $Z_0 := \sum_{i \in S_0} v_i v_i^T$.

3. While $\lambda_{\min}(Z_{t-1}) < 1 - 2\epsilon$ do

- (a) Compute the action matrix $A_t := (\alpha Z_{t-1} - l_t I_n)^{-2}$, where $l_t \in \mathbb{R}$ is the unique value such that $A_t \succ 0$ and $\text{tr}(A_t) = 1$.
- (b) Define $S'_{t-1} := \{i \in S_{t-1} : 2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle < \frac{1}{2}\}$.
- (c) Sample i_t from the following probability distribution:

$$\Pr(i_t = i) = \frac{1}{k}(1 - x_i)(1 - 2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle) \quad \text{for } i \in S'_{t-1},$$

$$\text{and } \Pr(i_t = \emptyset) = 1 - \sum_{i \in S'_{t-1}} \Pr(i_t = i).$$

- (d) Sample j_t from the following probability distribution:

$$\Pr(j_t = j) = \frac{x_j}{k}(1 + 2\alpha \langle v_j v_j^T, A_t^{1/2} \rangle) \quad \text{for } j \in [m] \setminus S_{t-1},$$

$$\text{and } \Pr(j_t = \emptyset) = 1 - \sum_{j \in [m] \setminus S_{t-1}} \Pr(j_t = j).$$

- (e) Set $S_t := S_{t-1} \cup \{j_t\} \setminus \{i_t\}$, $Z_t := \sum_{i \in S_t} v_i v_i^T$ and $t := t + 1$.

4. Return $S = S_{t-1}$ as the solution.

Before we state the main result of this algorithm, we first check that the algorithm is well-defined.

Claim 3.1. *The probability distributions in each iteration of the iterative randomized swapping algorithm are well-defined.*

Proof. To verify that the probability distribution for sampling i_t is well-defined, we need to show that $\Pr(i_t = i) \geq 0$ for $i \in S'_{t-1}$ and $\sum_{i \in S'_{t-1}} \Pr(i_t = i) \leq 1$. Since $A_t \succ 0$ and $x_i \in [0, 1]$ and $2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle \leq 1/2$ for $i \in S'_{t-1}$, it follows that for $i \in S'_{t-1}$ we have

$$0 \leq \Pr(i_t = i) = \frac{1}{k}(1 - x_i)(1 - 2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle) \leq \frac{1}{k},$$

and this implies that $\sum_{i \in S'_{t-1}} \Pr(i_t = i) \leq |S'_{t-1}|/k \leq m/k < 1$ by the definition of k .

Next we verify that the probability distribution for sampling j_t is well-defined. It is clear that $\Pr(j_t = j) \geq 0$ as $A_t \succ 0$ and $x_j \in [0, 1]$. We claim that $\sum_{j \in [m] \setminus S_{t-1}} \Pr(j_t = j) \leq \sum_{j \in [m]} \Pr(j_t = j) \leq 1$ as

$$\sum_{j \in [m]} \Pr(j_t = j) = \frac{1}{k} \sum_{j=1}^m x_j (1 + 2\alpha \langle v_j v_j^T, A_t^{1/2} \rangle) = \frac{1}{k} \left(\sum_{j=1}^m x_j + 2\alpha \text{tr}(A_t^{1/2}) \right) \leq \frac{1}{k} \left(m + \frac{2n}{\epsilon} \right) = 1,$$

where the second equality is by the assumption that $\sum_{j=1}^m x_j v_j v_j^T = I_n$, the last equality is by the definition of k , and the inequality uses that $x_j \in [0, 1]$, $\alpha = \sqrt{n}/\epsilon$ and the bound that $\text{tr}(A_t^{1/2}) \leq \sqrt{n}$.

To see that $\text{tr}(A_t^{1/2}) \leq \sqrt{n}$, let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A_t , then

$$\text{tr}(A_t^{1/2}) = \sum_{i=1}^n \sqrt{\lambda_i} \leq \sqrt{n} \sum_{i=1}^n \lambda_i = \sqrt{n} \text{tr}(A_t) = \sqrt{n}, \quad (3.1)$$

where the inequality is by Cauchy-Schwarz and the last equality is by the definition of A_t . \square

Remark 3.2. *The reader may wonder why we do not define the probability distribution for sampling i_t by*

$$\Pr(i_t = i) = \frac{(1 - x_i)(1 - 2\alpha\langle v_i v_i^T, A_t^{1/2} \rangle)}{\sum_{j \in S'_{t-1}} (1 - x_j)(1 - 2\alpha\langle v_j v_j^T, A_t^{1/2} \rangle)} \quad \text{for } i \in S'_{t-1},$$

so that $\sum_{i \in S'_{t-1}} \Pr(i_t = i) = 1$ and likewise for sampling j_t , so that we always remove a vector from S_{t-1} and add another vector to S_{t-1} in each iteration. This is our initial approach and we believe that this should also work, but it turns out that the calculations for the linear constraints simplify considerably by having a common denominator k for these two probability distributions.

The following is the main technical result for one-sided spectral rounding.

Theorem 3.3. *Suppose we are given $v_1, \dots, v_m \in \mathbb{R}^n$, $x \in [0, 1]^m$ such that $\sum_{i=1}^m x_i v_i v_i^T = I_n$. For any $\epsilon \in (0, \frac{1}{2})$, the iterative randomized swapping algorithm returns a subset $S \subseteq [m]$ satisfying*

$$\sum_{i \in S} v_i v_i^T \succcurlyeq (1 - 2\epsilon) I_n$$

within qk/ϵ iterations with probability at least $1 - \exp(-\Omega(q\sqrt{n}))$ for $q \geq 2$. Furthermore, for any $c \in \mathbb{R}_+^m$ and any $\delta_1 \in [0, 1]$, $\delta_2 \in [0, 1]$ and $\delta_3 > 0$, the probability that the returned solution S satisfies the cost upper bound is

$$\Pr \left[c(S) \leq (1 + \delta_1) \langle c, x \rangle + \frac{15nc_\infty}{\epsilon} \right] \geq 1 - \exp \left[-\Omega \left(\frac{\delta_1 n}{\epsilon} \right) \right],$$

and the probability that the returned solution S satisfies the cost lower bound is

$$\Pr \left[c(S) \geq (1 - \delta_2) \langle c, x \rangle - \delta_3 nc_\infty \right] \geq 1 - \exp \left[-\Omega \left(\min\{\delta_2 \delta_3, \epsilon \delta_3^2\} \cdot n \right) \right].$$

Remark 3.4. *If we set $\delta_1 = \delta_2 = \epsilon$ and $\delta_3 = 1/\epsilon$, then Theorem 3.3 states that the returned solution S satisfies*

$$(1 - \epsilon) \langle c, x \rangle - \frac{nc_\infty}{\epsilon} \leq c(S) \leq (1 + \epsilon) \langle c, x \rangle + \frac{15nc_\infty}{\epsilon}$$

with probability at least $1 - \exp(-\Omega(n))$ for any $c \in \mathbb{R}_+^m$. We introduce $\delta_1, \delta_2, \delta_3$ to have a more refined control of the failure probability of the lower bound, and this will be relevant in showing that linear covering constraints can be almost satisfied.

Organization: The remainder of this subsection is organized as follows. We will first prove that the spectral lower bound will be approximately satisfied with high probability within polynomial time in Section 3.1.1, and then prove the guarantees on the linear constraints in Section 3.1.2. Then, we will use Theorem 3.3 to prove the exact one-sided spectral rounding result in Theorem 1.8 in Section 3.1.3. Finally, we provide a proof of the concentration inequality in Theorem 2.7 in Section 3.1.4.

3.1.1 Bounding the Minimum Eigenvalue

The goal in this subsection is to prove that the probability that the algorithm does not terminate within $\tau \geq qk/\epsilon$ iterations is at most $\exp(-\Omega(q\sqrt{n}))$ for $q \geq 2$.

We will bound the minimum eigenvalue of the solution using the regret minimization framework developed in [3, 5]. The initial feedback matrix is $F_0 = Z_0$, which is constructed randomly using x . In each iteration $t \geq 1$, after computing the action matrix A_t , the algorithm responds with the feedback matrix $F_t = v_{j_t} v_{j_t}^T - v_{i_t} v_{i_t}^T$. Note that $Z_\tau = \sum_{t=0}^{\tau} F_t$. Define

$$\Delta_t^+ := \frac{\langle v_{j_t} v_{j_t}^T, A_t \rangle}{1 + 2\alpha \langle v_{j_t} v_{j_t}^T, A_t^{1/2} \rangle} \quad \text{and} \quad \Delta_t^- := \frac{\langle v_{i_t} v_{i_t}^T, A_t \rangle}{1 - 2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle} \quad \text{and} \quad \Delta_t := \Delta_t^+ - \Delta_t^-.$$

Note that $2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle < \frac{1}{2} < 1$ for $1 \leq t \leq \tau$ by the definition of S'_{t-1} , and so Δ_t^- is well-defined for $1 \leq t \leq \tau$. The regret minimization Theorem 2.4 proves that

$$\lambda_{\min}(Z_\tau) = \lambda_{\min}\left(\sum_{t=0}^{\tau} F_t\right) \geq \sum_{t=1}^{\tau} \Delta_t - \frac{2\sqrt{n}}{\alpha} = \sum_{t=1}^{\tau} \Delta_t - 2\epsilon. \quad (3.2)$$

To lower bound the minimum eigenvalue, we will prove that $\sum_{t=1}^{\tau} \Delta_t \geq 1$ with high probability. In the following, we bound the expected value of $\sum_{t=1}^{\tau} \Delta_t$, and then use Freeman's martingale inequality to bound the probability that $\sum_{t=1}^{\tau} \Delta_t$ deviates significantly from its expected value.

Lemma 3.5. *Let $\lambda := \max_{0 \leq t \leq \tau} \lambda_{\min}(Z_t)$. Then*

$$\sum_{t=1}^{\tau} \mathbb{E}[\Delta_t \mid S_{t-1}] \geq \sum_{t=1}^{\tau} \frac{1}{k} (1 - \epsilon - \lambda_{\min}(Z_{t-1})) \geq \frac{\tau}{k} (1 - \epsilon - \lambda).$$

Proof. We first consider the expected gain of adding the vector j_t . By the definition of the probability distribution of j_t ,

$$\begin{aligned} \mathbb{E}[\Delta_t^+ \mid S_{t-1}] &= \frac{1}{k} \sum_{j \in [m] \setminus S_{t-1}} x_j (1 + 2\alpha \langle v_j v_j^T, A_t^{1/2} \rangle) \cdot \frac{\langle v_j v_j^T, A_t \rangle}{1 + 2\alpha \langle v_j v_j^T, A_t^{1/2} \rangle} \\ &= \frac{1}{k} \sum_{j \in [m] \setminus S_{t-1}} x_j \langle v_j v_j^T, A_t \rangle \\ &= \frac{1}{k} \left(1 - \sum_{j \in S_{t-1}} x_j \langle v_j v_j^T, A_t \rangle \right), \end{aligned} \quad (3.3)$$

where the last equality is by $\sum_{j=1}^m x_j v_j v_j^T = I_n$ and $\text{tr}(A_t) = 1$ by the definition of A_t .

Then we consider the expected loss of removing the vector i_t . By the definition of the probability

distribution of i_t ,

$$\begin{aligned}
\mathbb{E}[\Delta_t^- | S_{t-1}] &= \sum_{i \in S'_{t-1}} \frac{1}{k} (1 - x_i) (1 - 2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle) \cdot \frac{\langle v_i v_i^T, A_t \rangle}{1 - 2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle} \\
&= \frac{1}{k} \sum_{i \in S'_{t-1}} (1 - x_i) \langle v_i v_i^T, A_t \rangle \\
&\leq \frac{1}{k} \sum_{i \in S_{t-1}} (1 - x_i) \langle v_i v_i^T, A_t \rangle \\
&\leq \frac{1}{k} \left(\lambda_{\min}(Z_{t-1}) + \epsilon - \sum_{i \in S_{t-1}} x_i \langle v_i v_i^T, A_t \rangle \right),
\end{aligned} \tag{3.4}$$

where the first inequality is because $x_i \in [0, 1]$ and $\langle v_i v_i^T, A_t \rangle \geq 0$ as $A_t \succ 0$, and the last inequality follows from Lemma 2.5 that $\langle Z_{t-1}, A_t \rangle \leq \sqrt{n}/\alpha + \lambda_{\min}(Z_{t-1})$ and $\alpha = \sqrt{n}/\epsilon$.

The lemma follows by combining (3.3) and (3.4) and summing over t and using $\lambda = \max_t \lambda_{\min}(Z_t)$. \square

Remark 3.6. If we use the probability distributions stated in Remark 3.2, then we can start with a solution with $l := \sum_{i=1}^m x_i + O(n/\epsilon)$ vectors and guarantee that the solution at each iteration still has exactly l vectors. A similar statement about the expected progress as in Lemma 3.5 can be proved. This implies that there exists a good pair $i_t \in S_{t-1}$ and $j_t \in S_{t-1}$, which gives a solution set of size l satisfying the spectral lower bound approximately. Together with a preprocessing step as in Section 3.1.3, this gives a simpler proof of the deterministic algorithm of [5].

Lemma 3.7. Let $\lambda := \max_{0 \leq t \leq \tau} \lambda_{\min}(Z_t)$. Then, for any $\eta > 0$,

$$\Pr \left[\sum_{t=1}^{\tau} \Delta_t \leq \left(\sum_{t=1}^{\tau} \mathbb{E}[\Delta_t | S_{t-1}] \right) - \eta \right] \leq \exp \left(- \frac{\eta^2 k \sqrt{n} / 2}{\tau \epsilon (1 + \lambda + \epsilon) + \eta k \epsilon / 3} \right).$$

Proof. We define the following sequences of random variables where $X_t := \mathbb{E}[\Delta_t | S_{t-1}] - \Delta_t$ and $Y_t := \sum_{l=1}^t X_l$. Observe that $\{Y_t\}_t$ is a martingale with respect to $\{S_t\}_t$. We use Freedman's inequality to bound $\Pr(Y_\tau \geq \eta)$. To apply Freedman's inequality, we need to upper bound X_t and $E[X_t^2 | S_{t-1}]$. Note that

$$0 \leq \Delta_t^+ = \frac{\langle v_{j_t} v_{j_t}^T, A_t \rangle}{1 + 2\alpha \langle v_{j_t} v_{j_t}^T, A_t^{1/2} \rangle} \leq \frac{\langle v_{j_t} v_{j_t}^T, A_t \rangle}{2\alpha \langle v_{j_t} v_{j_t}^T, A_t^{1/2} \rangle} \leq \frac{1}{2\alpha},$$

where the last inequality holds as $0 \prec A_t \preccurlyeq I$. Also,

$$0 \leq \Delta_t^- = \frac{\langle v_{i_t} v_{i_t}^T, A_t \rangle}{1 - 2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle} \leq \frac{\langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle}{1 - 2\alpha \langle v_{i_t} v_{i_t}^T, A_t^{1/2} \rangle} \leq \frac{1}{2\alpha},$$

where the second last inequality is by $0 \prec A_t \preccurlyeq I$, and the first and last inequality are because i_t is chosen from the set $S'_{t-1} := \{i \mid 4\alpha \langle v_i v_i^T, A_t^{1/2} \rangle < 1\}$. (We remark that this upper bound on Δ_t^-

is exactly the reason for the definition of S'_{t-1} .) As these lower and upper bounds on Δ_t^+ and Δ_t^- hold with probability one, we have the deterministic upper bound $X_t \leq R := \epsilon/\sqrt{n}$ as

$$X_t = \mathbb{E}[\Delta_t \mid S_{t-1}] - \Delta_t \leq \mathbb{E}[\Delta_t^+ \mid S_{t-1}] + \Delta_t^- \leq \frac{1}{\alpha} = \frac{\epsilon}{\sqrt{n}} = R.$$

Next, we upper bound

$$\mathbb{E}[X_t^2 \mid S_{t-1}] \leq R \cdot \mathbb{E}[|X_t| \mid S_{t-1}] \leq \frac{\epsilon}{\sqrt{n}} \left(\mathbb{E}[\Delta_t^+ \mid S_{t-1}] + \mathbb{E}[\Delta_t^- \mid S_{t-1}] \right) \leq \frac{\epsilon}{k\sqrt{n}} (1 + \lambda + \epsilon),$$

where the last inequality follows from (3.3) and (3.4) that $\mathbb{E}[\Delta_t^+ \mid S_{t-1}] \leq 1/k$ and $\mathbb{E}[\Delta_t^- \mid S_{t-1}] \leq (\lambda + \epsilon)/k$. Therefore, $W_\tau := \sum_{t=1}^\tau \mathbb{E}[X_t^2 \mid S_{t-1}] \leq \tau\epsilon(1 + \lambda + \epsilon)/(k\sqrt{n})$. Applying Theorem 2.6 with $R = \epsilon/\sqrt{n}$ and $\sigma^2 = \tau\epsilon(1 + \lambda + \epsilon)/(k\sqrt{n})$, it follows that

$$\Pr(Y_\tau \geq \eta) \leq \exp\left(-\frac{\eta^2/2}{\sigma^2 + R\eta/3}\right) = \exp\left(-\frac{\eta^2 k \sqrt{n}/2}{\tau\epsilon(1 + \lambda + \epsilon) + \eta k \epsilon/3}\right).$$

The lemma follows as $Y_\tau \geq \eta$ is equivalent to $\sum_{t=1}^\tau \Delta_t \leq (\sum_{t=1}^\tau \mathbb{E}[\Delta_t \mid S_{t-1}]) - \eta$. \square

We are ready to prove that the algorithm terminates in a polynomial number of iterations with high probability.

Theorem 3.8. *The probability that the iterative randomized swapping algorithm does not terminate in qk/ϵ iterations for $q \geq 2$ is at most $\exp(-\Omega(q\sqrt{n}))$.*

Proof. Let $\tau = qk/\epsilon$. Suppose $\lambda = \max_{0 \leq t \leq \tau} \lambda_{\min}(Z_t) < 1 - 2\epsilon$. Then, Lemma 3.5 implies that

$$\sum_{t=1}^\tau \mathbb{E}[\Delta_t \mid S_{t-1}] \geq \frac{\tau}{k}(1 - \epsilon - \lambda) = \frac{q}{\epsilon}(1 - \epsilon - \lambda) > q,$$

and the regret minimization bound in (3.2) implies that

$$1 - 2\epsilon > \lambda_{\min}(Z_\tau) \geq \left(\sum_{t=1}^\tau \Delta_t\right) - 2\epsilon \implies \sum_{t=1}^\tau \Delta_t < 1.$$

Therefore,

$$\begin{aligned} \Pr\left[\bigcap_{t=0}^\tau \left(\lambda_{\min}(Z_t) < 1 - 2\epsilon\right)\right] &\leq \Pr\left[\sum_{t=1}^\tau \Delta_t < \left(\sum_{t=1}^\tau \mathbb{E}[\Delta_t \mid S_{t-1}]\right) - (q - 1)\right] \\ &\leq \exp\left(-\frac{(q - 1)^2 k \sqrt{n}/2}{(qk/\epsilon)\epsilon(1 + (1 - 2\epsilon) + \epsilon) + (q - 1)k\epsilon/3}\right) \\ &\leq \exp(-\Omega(q\sqrt{n})), \end{aligned}$$

where the second inequality is by Lemma 3.7 with $\eta = q - 1$ and $\tau = qk/\epsilon$ and the last inequality is by the assumption that $q \geq 2$. \square

So, for example, the probability that the algorithm does not terminate in $2k/\epsilon$ iterations is at most $\exp(-\Omega(\sqrt{n}))$ and the probability that it does not terminate in $k\sqrt{n}/\epsilon$ iterations is at most $\exp(-\Omega(n))$.

3.1.2 Bounding the Linear Constraints

For an arbitrary non-negative linear constraint $c \in \mathbb{R}_+^m$, the goal in this subsection is to prove that $c(S_t) \approx \langle c, x \rangle$ with high probability for any t , where we recall that $c(S_t) := \sum_{i \in S_t} c_i$ is the “cost” of the solution at time t . We first bound the expected change of the cost in an iteration.

Lemma 3.9. *Suppose $\lambda_{\min}(Z_{t-1}) < 1$. Then*

$$\frac{1}{k} \left(\langle c, x \rangle - c(S_{t-1}) \right) \leq \mathbb{E}[c_{j_t} - c_{i_t} \mid S_{t-1}] \leq \frac{1}{k} \left(\langle c, x \rangle - c(S_{t-1}) + \frac{14nc_\infty}{\epsilon} \right).$$

Proof. We first bound the conditional expectation of c_{j_t} . By the probability distribution of j_t ,

$$\begin{aligned} \mathbb{E}[c_{j_t} \mid S_{t-1}] &= \frac{1}{k} \sum_{j \in [m] \setminus S_{t-1}} c_j x_j (1 + 2\alpha \langle v_j v_j^T, A_t^{1/2} \rangle) \\ &= \frac{1}{k} \left(\langle c, x \rangle - \sum_{j \in S_{t-1}} c_j x_j + 2\alpha \sum_{j \in [m] \setminus S_{t-1}} c_j x_j \langle v_j v_j^T, A_t^{1/2} \rangle \right). \end{aligned}$$

Note that

$$0 \leq 2\alpha \sum_{j \in [m] \setminus S_{t-1}} c_j x_j \langle v_j v_j^T, A_t^{1/2} \rangle \leq 2\alpha c_\infty \sum_{j=1}^m x_j \langle v_j v_j^T, A_t^{1/2} \rangle = 2\alpha c_\infty \text{tr}(A_t^{1/2}) \leq \frac{2nc_\infty}{\epsilon},$$

where the equality holds as $\sum_{j=1}^m x_j v_j v_j^T = I_n$ and the last inequality is by (3.1) and $\alpha = \sqrt{n}/\epsilon$. Therefore,

$$\frac{1}{k} \left(\langle c, x \rangle - \sum_{i \in S_{t-1}} c_i x_i \right) \leq \mathbb{E}[c_{j_t} \mid S_{t-1}] \leq \frac{1}{k} \left(\langle c, x \rangle - \sum_{i \in S_{t-1}} c_i x_i + \frac{2nc_\infty}{\epsilon} \right). \quad (3.5)$$

Next we bound the expectation of c_{i_t} . By the probability distribution of i_t ,

$$\begin{aligned} \mathbb{E}[c_{i_t} \mid S_{t-1}] &= \frac{1}{k} \sum_{i \in S'_{t-1}} c_i (1 - x_i) (1 - 2\alpha \langle v_i v_i^T, A_t^{1/2} \rangle) \\ &= \frac{1}{k} \left(\sum_{i \in S'_{t-1}} c_i (1 - x_i) - 2\alpha \sum_{i \in S'_{t-1}} c_i (1 - x_i) \langle v_i v_i^T, A_t^{1/2} \rangle \right) \\ &= \frac{1}{k} \left(c(S_{t-1}) - \sum_{i \in S_{t-1}} c_i x_i - \sum_{i \in S_{t-1} \setminus S'_{t-1}} c_i (1 - x_i) - 2\alpha \sum_{i \in S'_{t-1}} c_i (1 - x_i) \langle v_i v_i^T, A_t^{1/2} \rangle \right). \end{aligned}$$

We would like to bound the last two terms of the right hand side. Recall that $S'_{t-1} := \{i \in S_{t-1} \mid 4\alpha \langle v_i v_i^T, A_t^{1/2} \rangle < 1\}$. This implies that

$$|S_{t-1} \setminus S'_{t-1}| \leq \sum_{i \in S_{t-1} \setminus S'_{t-1}} 4\alpha \langle v_i v_i^T, A_t^{1/2} \rangle \leq 4\alpha \sum_{i \in S_{t-1}} \langle v_i v_i^T, A_t^{1/2} \rangle \leq 4(n + \alpha\sqrt{n} \cdot \lambda_{\min}(Z_{t-1})) \leq \frac{8n}{\epsilon},$$

where the second last inequality uses Lemma 2.5 and the last inequality is by $\alpha = \sqrt{n}/\epsilon$ and the assumption that $\lambda_{\min}(Z_{t-1}) \leq 1$. Since $x \in [0, 1]^m$ and $c \geq 0$, it follows that the second last term is

$$0 \leq \sum_{i \in S_{t-1} \setminus S'_{t-1}} c_i(1 - x_i) \leq c_\infty \cdot |S_{t-1} \setminus S'_{t-1}| \leq \frac{8nc_\infty}{\epsilon}.$$

Similarly, for the last term,

$$0 \leq 2\alpha \sum_{i \in S'_{t-1}} c_i(1 - x_i) \langle v_i v_i^T, A_t^{1/2} \rangle \leq 2c_\infty \cdot \alpha \sum_{i \in S_{t-1}} \langle v_i v_i^T, A_t^{1/2} \rangle \leq 2c_\infty(n + \alpha\sqrt{n} \cdot \lambda_{\min}(Z_{t-1})) \leq \frac{4nc_\infty}{\epsilon}.$$

Plugging back these upper and lower bounds for the last two terms, we obtain

$$\frac{1}{k} \left(c(S_{t-1}) - \sum_{i \in S_{t-1}} c_i x_i - \frac{12nc_\infty}{\epsilon} \right) \leq \mathbb{E}[c_{it} \mid S_{t-1}] \leq \frac{1}{k} \left(c(S_{t-1}) - \sum_{i \in S_{t-1}} c_i x_i \right). \quad (3.6)$$

The lemma follows by combining the bounds for the expectations of c_{it} and c_{jt} in (3.5) and (3.6). \square

To bound the difference between $c(S_t)$ and $\langle c, x \rangle$, we consider the following sequences of random variables where

$$Y_t := \frac{c(S_t) - \langle c, x \rangle}{c_\infty} \text{ for } t \geq 0 \quad \text{and} \quad X_t := Y_t - Y_{t-1} = \frac{c_{jt} - c_{it}}{c_\infty} \text{ for } t \geq 1. \quad (3.7)$$

Note that Lemma 3.9 shows that the sequence $\{Y_t\}_t$ has the “self-adjusting” property that if Y_t is (more) positive then $E[Y_{t+1}] - Y_t$ is (more) negative and vice versa, so intuitively Y_t cannot be too far away from zero. The sequence $\{Y_t\}_t$ is not a martingale, and so we cannot apply Freedman’s inequality to prove concentration. Instead, we will use Theorem 2.7 to prove that the absolute value of Y_t is small with high probability. To apply Theorem 2.7, we need to bound the conditional second moment of X_t and the moment generating function of the initial solution S_0 .

Lemma 3.10. *Suppose $\lambda_{\min}(Z_{t-1}) < 1$. Then*

$$\mathbb{E}[(c_{jt} - c_{it})^2 \mid S_{t-1}] \leq \frac{c_\infty}{k} \cdot \left(\langle c, x \rangle + c(S_{t-1}) + \frac{2nc_\infty}{\epsilon} \right).$$

Proof. Since $c_i \geq 0$ for all $1 \leq i \leq m$,

$$\begin{aligned} \mathbb{E}[(c_{jt} - c_{it})^2 \mid S_{t-1}] &\leq \max_{i_t, j_t} |c_{jt} - c_{it}| \cdot \mathbb{E}[|c_{jt} - c_{it}| \mid S_{t-1}] \\ &\leq c_\infty \cdot \mathbb{E}[c_{jt} + c_{it} \mid S_{t-1}] \\ &\leq \frac{c_\infty}{k} \left(\langle c, x \rangle + c(S_{t-1}) + \frac{2nc_\infty}{\epsilon} \right), \end{aligned}$$

where the last inequality is by (3.5) and (3.6). \square

We use the fact that the initial solution S_0 is generated randomly to bound its moment generating function.

Lemma 3.11. For $a \in [-1, 1]$,

$$\mathbb{E} [e^{aY_0}] \leq e^{a^2 \langle c, x \rangle / c_\infty}.$$

Proof. Let χ_i be the indicator variable where $\chi_i = 1$ if $i \in S_0$ and $\chi_i = 0$ otherwise. Since the algorithm constructs S_0 by sampling each vector independently with probability x_i , it follows that

$$\mathbb{E} [e^{ac(S_0)/c_\infty}] = \mathbb{E} [e^{a \sum_{i=1}^m \chi_i c_i / c_\infty}] = \prod_{i=1}^m \mathbb{E} [e^{a \chi_i c_i / c_\infty}] = \prod_{i=1}^m (1 - x_i + x_i e^{ac_i/c_\infty}).$$

Note that $ac_i/c_\infty \leq 1$ as $a \in [-1, 1]$ and $c_i/c_\infty \leq 1$, and thus $e^{ac_i/c_\infty} \leq 1 + ac_i/c_\infty + a^2 c_i^2 / c_\infty^2$ as $e^p \leq 1 + p + p^2$ for $p \leq 1$. Therefore,

$$\mathbb{E} [e^{ac(S_0)/c_\infty}] \leq \prod_{i=1}^m \left(1 + \frac{ac_i x_i}{c_\infty} + \frac{a^2 c_i^2 x_i}{c_\infty^2} \right) \leq \exp \left(\sum_{i=1}^m \left(\frac{ac_i x_i}{c_\infty} + \frac{a^2 c_i^2 x_i}{c_\infty^2} \right) \right) = \exp \left(\frac{(a + a^2) \langle c, x \rangle}{c_\infty} \right),$$

where the second inequality uses $1 + p \leq e^p$ for $p \in \mathbb{R}$ and $c_i \leq c_\infty$ for $1 \leq i \leq m$. The claim follows as $Y_0 = (c(S_0) - \langle c, x \rangle) / c_\infty$. \square

We are ready to apply Theorem 2.7 to bound the cost.

Theorem 3.12. Suppose the iterative randomized swapping algorithm terminates at the τ -th iteration. Let $c \in \mathbb{R}_+^m$. For any $\delta_1 \in [0, 1]$,

$$\Pr \left[c(S_\tau) \leq (1 + \delta_1) \langle c, x \rangle + \frac{15nc_\infty}{\epsilon} \right] \geq 1 - \exp \left[-\Omega \left(\frac{\delta_1 n}{\epsilon} \right) \right].$$

Also, for any $\delta_2 \in [0, 1]$ and $\delta_3 > 0$,

$$\Pr \left[c(S_\tau) \geq (1 - \delta_2) \langle c, x \rangle - \delta_3 nc_\infty \right] \geq 1 - \exp \left(-\Omega \left(\min\{\delta_2 \delta_3, \epsilon \delta_3^2\} \cdot n \right) \right).$$

Proof. As the algorithm terminates the first time when the minimum eigenvalue of the solution is at least $1 - 2\epsilon$, we can assume that $\lambda_{\min}(Z_t) < 1 - 2\epsilon < 1$ for $0 \leq t < \tau$. We will apply Theorem 2.7 on the sequences $\{X_t\}_t$ and $\{Y_t\}_t$ as defined in (3.7). Firstly, note that $|X_t| \leq 1$ by definition for all $t \geq 1$. Secondly, as $\mathbb{E}[X_t | Y_0, \dots, Y_{t-1}] = \mathbb{E}[(c_{j_t} - c_{i_t})/c_\infty | S_{t-1}]$ and $Y_{t-1} = (c(S_{t-1}) - \langle c, x \rangle) / c_\infty$, Lemma 3.9 implies that

$$\mathbb{E}[X_t | Y_0, \dots, Y_{t-1}] \leq \frac{1}{kc_\infty} \left(\langle c, x \rangle - c(S_{t-1}) + \frac{14nc_\infty}{\epsilon} \right) = -\frac{Y_{t-1}}{k} + \frac{14n}{k\epsilon},$$

and

$$\mathbb{E}[X_t | Y_0, \dots, Y_{t-1}] \geq \frac{1}{kc_\infty} \left(\langle c, x \rangle - c(S_{t-1}) \right) = -\frac{Y_{t-1}}{k}.$$

Thirdly, since $\mathbb{E}[X_t^2 | Y_0, \dots, Y_{t-1}] = \mathbb{E}[(c_{j_t} - c_{i_t})^2 / c_\infty^2 | S_{t-1}]$, Lemma 3.10 implies that

$$\mathbb{E}[X_t^2 | Y_0, \dots, Y_{t-1}] \leq \frac{1}{kc_\infty} \left(\langle c, x \rangle + c(S_{t-1}) + \frac{2nc_\infty}{\epsilon} \right) = \frac{Y_{t-1}}{k} + \frac{2}{kc_\infty} \left(\langle c, x \rangle + \frac{nc_\infty}{\epsilon} \right).$$

Finally, Lemma 3.11 states that $\mathbb{E}[e^{aY_0}] \leq \exp(a^2\langle c, x \rangle / c_\infty)$ for $a \in [-1, 1]$. By setting

$$\gamma = \frac{1}{k}, \quad \beta_u = \frac{14n}{k\epsilon}, \quad \beta_l = 0, \quad \sigma = \frac{2}{kc_\infty} \left(\langle c, x \rangle + \frac{nc_\infty}{\epsilon} \right),$$

we can check that all the conditions of Theorem 2.7 are satisfied. Applying Theorem 2.7 with $\eta = \delta_1 \langle c, x \rangle / c_\infty + n/\epsilon$ for $\delta_1 \in [0, 1]$,

$$\begin{aligned} \Pr \left[c(S_t) \geq (1 + \delta_1) \langle c, x \rangle + \frac{15nc_\infty}{\epsilon} \right] &= \Pr \left[Y_t \geq \frac{\beta_u}{\gamma} + \eta \right] \\ &\leq \exp \left[-\frac{\eta^2}{4(\sigma + \beta_u)/\gamma + 2\eta} \right] \\ &= \exp \left[-\frac{c_\infty \eta^2}{8\langle c, x \rangle + 64nc_\infty/\epsilon + 2\eta c_\infty} \right] \\ &\leq \exp \left[-\Omega \left(\frac{\delta_1 n}{\epsilon} \right) \right], \end{aligned}$$

where the last inequality is because $\eta c_\infty = O(\langle c, x \rangle + nc_\infty/\epsilon)$ and thus the denominator is $\Theta(\langle c, x \rangle + nc_\infty/\epsilon)$, and the numerator is $\eta^2 c_\infty = \eta(\delta_1 \langle c, x \rangle + nc_\infty/\epsilon) \geq (n/\epsilon)\delta_1(\langle c, x \rangle + nc_\infty/\epsilon)$.

Similarly, for the cost lower bound, we apply Theorem 2.7 with $\eta = \delta_2 \langle c, x \rangle / c_\infty + \delta_3 n$ for $\delta_2 \in [0, 1]$ and $\delta_3 > 0$ to obtain

$$\begin{aligned} \Pr [c(S_t) \leq (1 - \delta_2) \langle c, x \rangle - \delta_3 nc_\infty] &= \Pr \left[Y_t \leq -\frac{\beta_l}{\gamma} - \eta \right] \\ &\leq \exp \left[-\frac{\eta^2}{4\sigma/\gamma + \eta} \right] \\ &= \exp \left[-\frac{\eta^2 c_\infty}{8(\langle c, x \rangle + nc_\infty/\epsilon) + \eta c_\infty} \right] \\ &\leq \exp \left[-\Omega \left(\frac{\delta_3 n(\delta_2 \langle c, x \rangle + \delta_3 nc_\infty)}{\langle c, x \rangle + nc_\infty/\epsilon + \delta_3 nc_\infty} \right) \right] \\ &\leq \exp \left[-\Omega \left(\min\{\delta_2 \delta_3, \epsilon \delta_3^2\} \cdot n \right) \right], \end{aligned}$$

where the second last inequality is by similar calculations as in the previous case. \square

3.1.3 Exact One-Sided Spectral Rounding

Theorem 3.3 follows directly from Theorem 3.8 and Theorem 3.12. This shows that the iterative randomized swapping algorithm will return a solution S with $\sum_{i \in S} v_i v_i^T \succcurlyeq (1 - 2\epsilon)I_n$ and $c(S_t) \approx \langle c, x \rangle$ with high probability for any $c \in \mathbb{R}_+^m$.

To prove Theorem 1.8 where the goal is to return a solution S with $\sum_{i \in S} v_i v_i^T \succcurlyeq I_n$, our idea is to scale up the fractional solution x and then apply Theorem 3.3. The following is the detailed description of the algorithm.

Exact One-Sided Spectral Rounding

Input: $v_1, \dots, v_m \in \mathbb{R}^n$ and $x \in [0, 1]^m$ with $\sum_{i=1}^m x_i v_i v_i^T = I_n$, and an error parameter $\epsilon \in (0, \frac{1}{4})$.

Output: a subset $S \subseteq [m]$ such that $\sum_{i \in S} v_i v_i^T \succcurlyeq I_n$ and $c(S) \approx \langle c, x \rangle$ for any $c \in \mathbb{R}_+^m$ with high probability.

1. Define $y_i := x_i / (1 - 2\epsilon)$ and $u_i := \sqrt{1 - 2\epsilon} \cdot v_i$ for $i \in [m]$. Note that $\sum_{i=1}^m y_i u_i u_i^T = I_n$.
2. Let $S_{\text{big}} := \{i \in [m] : y_i > 1\}$, $S_{\text{small}} := \{i \in [m] : 0 \leq y_i \leq 1\}$, and $Z_{\text{big}} = \sum_{i \in S_{\text{big}}} y_i u_i u_i^T$.
3. Define $w_i := (I_n - Z_{\text{big}})^{-\frac{1}{2}} u_i$ for each $i \in S_{\text{small}}$, so that $\sum_{i \in S_{\text{small}}} y_i w_i w_i^T = I_n$ ¹.
4. Apply the iterative randomized swapping algorithm with $\{w_i \mid i \in S_{\text{small}}\}$ and $\{y_i \mid i \in S_{\text{small}}\}$ as input to obtain a solution set $S'_{\text{small}} \subseteq S_{\text{small}}$ with $\sum_{i \in S'_{\text{small}}} w_i w_i^T \succcurlyeq (1 - 2\epsilon) I_n$.
5. Return $S := S_{\text{big}} \cup S'_{\text{small}}$ as the solution.

Proof of Theorem 1.8. We first analyze the spectral lower bound. By the definitions of w_i and u_i ,

$$\sum_{i \in S'_{\text{small}}} w_i w_i^T \succcurlyeq (1 - 2\epsilon) I_n \implies \sum_{i \in S'_{\text{small}}} u_i u_i^T \succcurlyeq (1 - 2\epsilon)(I_n - Z_{\text{big}}) \implies \sum_{i \in S'_{\text{small}}} v_i v_i^T \succcurlyeq I_n - Z_{\text{big}}.$$

For the vectors in S_{big} , as $x_i \in [0, 1]$,

$$\sum_{i \in S_{\text{big}}} v_i v_i^T \succcurlyeq \sum_{i \in S_{\text{big}}} x_i v_i v_i^T = \sum_{i \in S_{\text{big}}} y_i u_i u_i^T = Z_{\text{big}}.$$

Therefore, it follows that

$$\sum_{i \in S} v_i v_i^T = \sum_{i \in S_{\text{big}} \cup S'_{\text{small}}} v_i v_i^T = \sum_{i \in S'_{\text{small}}} v_i v_i^T + \sum_{i \in S_{\text{big}}} v_i v_i^T \succcurlyeq (I_n - Z_{\text{big}}) + Z_{\text{big}} = I_n.$$

Next, we prove that $c(S) \approx \langle c, x \rangle$ with high probability for any vector $c \in \mathbb{R}_+^m$. Let $\langle c, x \rangle_{\text{small}} := \sum_{i \in S_{\text{small}}} c_i x_i$ and $\langle c, x \rangle_{\text{big}} := \sum_{i \in S_{\text{big}}} c_i x_i$. For the vectors in S_{big} , as $y_i > 1$ for $i \in S_{\text{big}}$ and $y_i = x_i / (1 - 2\epsilon)$ for all $i \in [m]$, it follows that

$$\langle c, x \rangle_{\text{big}} \leq c(S_{\text{big}}) \leq \langle c, y \rangle_{\text{big}} = \frac{\langle c, x \rangle_{\text{big}}}{1 - 2\epsilon}.$$

For the vectors in S_{small} , by Theorem 3.3 with $\delta_1 = \epsilon$, the returned set S'_{small} in Step 4 satisfies the cost upper bound

$$c(S'_{\text{small}}) \leq (1 + \epsilon) \langle c, y \rangle_{\text{small}} + \frac{15nc_{\infty}}{\epsilon} = \frac{(1 + \epsilon) \langle c, x \rangle_{\text{small}}}{1 - 2\epsilon} + \frac{15nc_{\infty}}{\epsilon}$$

¹If $I_n - B$ is singular, we first project the vectors to the orthogonal complement of the nullspace before applying the transformation. We can add dummy coordinates to keep the vectors to have the same dimension n for simplicity of the analysis.

with probability at least $1 - \exp(-\Omega(n))$, which implies that for $\epsilon \in (0, \frac{1}{4})$,

$$c(S) = c(S_{\text{big}}) + c(S'_{\text{small}}) \leq \frac{1+\epsilon}{1-2\epsilon} (\langle c, x \rangle_{\text{big}} + \langle c, x \rangle_{\text{small}}) + \frac{15nc_{\infty}}{\epsilon} \leq (1+6\epsilon) \langle c, x \rangle + \frac{15nc_{\infty}}{\epsilon}.$$

Similarly, by Theorem 3.3 with $\delta_2 = \epsilon$ and $\delta_3 = \delta$ for some $\delta > 0$, the returned set S'_{small} in Step 4 satisfies the cost lower bound

$$c(S'_{\text{small}}) \geq (1-\epsilon) \langle c, y \rangle_{\text{small}} - \delta nc_{\infty} = \frac{1-\epsilon}{1-2\epsilon} \langle c, x \rangle_{\text{small}} - \delta nc_{\infty} \geq \langle c, x \rangle_{\text{small}} - \delta nc_{\infty}$$

with probability at least $1 - \exp(-\Omega(\min\{\epsilon\delta, \epsilon\delta^2\} \cdot n))$, which implies that

$$c(S) = c(S_{\text{big}}) + c(S'_{\text{small}}) \geq \langle c, x \rangle_{\text{big}} + \langle c, x \rangle_{\text{small}} - \delta nc_{\infty} = \langle c, x \rangle - \delta nc_{\infty}.$$

□

3.1.4 Proof of the Concentration Inequality for Self-Adjusting Process (Theorem 2.7)

The proof is by computing the moment generating function of Y_t and applying Markov's inequality, which is standard in concentration inequalities. In the following, we write the conditional expectation as $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot | Y_0, \dots, Y_{t-1}]$ for simplicity.

Upper Tail: We start with the proof for the upper tail. For any $a \in [0, 1]$, the conditional moment generating function of X_t with any given Y_0, \dots, Y_{t-1} is

$$\begin{aligned} \mathbb{E}_t [e^{aX_t}] &= \mathbb{E}_t \left[\sum_{l=0}^{\infty} \frac{a^l X_t^l}{l!} \right] \leq \mathbb{E}_t \left[1 + aX_t + X_t^2 \sum_{l=2}^{\infty} \frac{a^l}{l!} \right] \\ &= 1 + a\mathbb{E}_t[X_t] + \mathbb{E}_t[X_t^2] \cdot (e^a - 1 - a) \\ &\leq 1 + a\mathbb{E}_t[X_t] + a^2\mathbb{E}_t[X_t^2] \\ &\leq 1 - a\gamma Y_{t-1} + a\beta_u + a^2\gamma Y_{t-1} + a^2\sigma \\ &\leq \exp(a^2\sigma + a\beta_u - \gamma(1-a)aY_{t-1}), \end{aligned}$$

where the first inequality is by the bounded difference property that $|X_t| \leq 1$ always, the second inequality is because $e^p \leq 1 + p + p^2$ for $p \leq 1$, the third inequality is by the self-adjusting property and the bounded variance property and $a \geq 0$, and the last inequality uses $1 + p \leq e^p$ for $p \in \mathbb{R}$. Then we can bound the moment generating function of Y_t as

$$\begin{aligned} \mathbb{E}_{Y_0, \dots, Y_t} [e^{aY_t}] &= \mathbb{E}_{Y_0, \dots, Y_{t-1}} [e^{aY_{t-1}} \cdot \mathbb{E}_t [e^{aX_t}]] \\ &\leq \mathbb{E}_{Y_0, \dots, Y_{t-1}} [\exp(a^2\sigma + a\beta_u + (1-\gamma(1-a))aY_{t-1})] \\ &\leq \exp(a^2\sigma + a\beta_u) \cdot \mathbb{E}_{Y_0, \dots, Y_{t-1}} [\exp(a(1-(1-a)\gamma)Y_{t-1})] \\ &= \exp(a^2\sigma + a\beta_u) \cdot \mathbb{E}_{Y_0, \dots, Y_{t-1}} [\exp(f(a) \cdot Y_{t-1})], \end{aligned}$$

where we define $f(a) := a(1 - (1-a)\gamma)$. Note that $f(a) \in [0, a]$ by the assumptions $\gamma \in (0, \frac{1}{2}]$ and $a \in [0, 1]$. Define the sequence $a_{(0)} = a$ and $a_{(i)} = f(a_{(i-1)})$ for $i \geq 1$. Apply the same argument

inductively, it follows that

$$\mathbb{E}_{Y_0, \dots, Y_t} [e^{aY_t}] \leq \exp \left[\sum_{i=0}^{t-1} (a_{(i)}^2 \sigma + a_{(i)} \beta_u) \right] \cdot \mathbb{E}_{Y_0} [e^{a_{(t)} Y_0}] \leq \exp \left[\sum_{i=0}^{t-1} (a_{(i)}^2 \sigma + a_{(i)} \beta_u) + \frac{a_{(t)}^2 \sigma}{\gamma} \right],$$

where the last inequality follows from the initial concentration property of Y_0 . To bound the moment generating function, we use the following claim whose proof follows from the definition of the sequence $\{a_{(i)}\}_i$.

Claim 3.13. *The sequence $\{a_{(i)}\}_{i \geq 0}$ is decreasing and dominated by the geometric sequence $\{ar^i\}_{i \geq 0}$ with common ratio $r := 1 - (1 - a)\gamma$. The sequence $\{a_{(i)}^2\}_i$ is also decreasing and dominated by the geometric sequence $\{a^2 r^{2i}\}_{i \geq 0}$ with common ratio r^2 . Furthermore, $r^2 < r < 1$ when $a \in [0, 1)$.*

Using Claim 3.13, when $a \in [0, 1)$, we can upper bound the moment generating function by

$$\begin{aligned} \mathbb{E}_{Y_0, \dots, Y_t} [e^{aY_t}] &\leq \exp \left[(a^2 \sigma + a \beta_u) \sum_{i=0}^{t-1} r^i + \frac{a^2 \sigma r^t}{\gamma} \right] \\ &= \exp \left[(a^2 \sigma + a \beta_u) \cdot \frac{1 - r^t}{1 - r} + \frac{a^2 \sigma r^t}{\gamma} \right] \\ &= \exp \left[\frac{a^2 \sigma + a \beta_u}{(1 - a)\gamma} \cdot (1 - r^t) + \frac{a^2 \sigma r^t}{\gamma} \right] \\ &\leq \exp \left[\frac{a^2 \sigma + a \beta_u}{(1 - a)\gamma} \right], \end{aligned}$$

where the last inequality uses $a \in [0, 1)$. By Markov inequality, for any $a \in [0, 1)$ and any $\eta > 0$,

$$\begin{aligned} \Pr \left[Y_t \geq \frac{\beta_u}{\gamma} + \eta \right] &= \Pr \left[e^{aY_t} \geq e^{a \left(\frac{\beta_u}{\gamma} + \eta \right)} \right] \leq \mathbb{E}_{Y_0, \dots, Y_t} [e^{aY_t}] / e^{a \left(\frac{\beta_u}{\gamma} + \eta \right)} \\ &\leq \exp \left[\frac{a^2 \sigma + a \beta_u}{(1 - a)\gamma} - a \left(\frac{\beta_u}{\gamma} + \eta \right) \right] \\ &= \exp \left[\frac{a^2 (\sigma + \beta_u)}{(1 - a)\gamma} - a\eta \right]. \end{aligned}$$

To prove the best upper bound, we optimize over a and set

$$a = 1 - \sqrt{\frac{(\sigma + \beta_u)/\gamma}{(\sigma + \beta_u)/\gamma + \eta}} = 1 - \sqrt{\frac{\nu}{\nu + \eta}},$$

where we use $\nu := (\sigma + \beta_u)/\gamma$ as a shorthand. Notice that $a \in [0, 1)$ as $\sigma, \gamma, \eta > 0$ and $\beta_u \geq 0$, so the above probability bound applies. Putting this choice of a back into the exponent on the right

hand side, the exponent is

$$\begin{aligned}
\frac{a^2(\sigma + \beta_u)}{(1-a)\gamma} - a\eta &= \frac{(1 - \sqrt{\nu/(\nu + \eta)})^2 \cdot \nu}{\sqrt{\nu/(\nu + \eta)}} - \left(1 - \sqrt{\frac{\nu}{\nu + \eta}}\right) \cdot \eta \\
&= \left(1 + \frac{\nu}{\nu + \eta} - 2\sqrt{\frac{\nu}{\nu + \eta}}\right) \cdot \sqrt{\nu(\nu + \eta)} - \left(1 - \sqrt{\frac{\nu}{\nu + \eta}}\right) \cdot \eta \\
&= \sqrt{\nu(\nu + \eta)} + \nu\sqrt{\frac{\nu}{\nu + \eta}} - 2\nu - \eta + \eta\sqrt{\frac{\nu}{\nu + \eta}} \\
&= -(2\nu + \eta) + 2\sqrt{\nu(\nu + \eta)} \\
&= -(2\nu + \eta) + \sqrt{(2\nu + \eta)^2 - \eta^2} \\
&= -(2\nu + \eta) + (2\nu + \eta)\sqrt{1 - \frac{\eta^2}{(2\nu + \eta)^2}} \\
&\leq -\frac{\eta^2/2}{2\nu + \eta},
\end{aligned}$$

where we used $\sqrt{1-p} \leq 1 - p/2$ for $p \in [0, 1]$ in the last inequality. Therefore, we conclude that

$$\Pr(Y_t \geq \eta) \leq \exp\left(\frac{a^2(\sigma + \beta_u)}{(1-a)\gamma} - a\eta\right) \leq \exp\left(-\frac{\eta^2/2}{2(\sigma + \beta_u)/\gamma + \eta}\right),$$

which completes the proof for the upper tail.

Lower Tail: The proof for the lower tail is quite similar to that for the upper tail. The main difference is that we work with the moment generating function $\mathbb{E}[e^{-aY_t}]$, instead of $\mathbb{E}[e^{aY_t}]$. For any $a \in [0, 1]$, the conditional moment generating function of $-X_t$ is

$$\begin{aligned}
\mathbb{E}_t[e^{-aX_t}] &= \mathbb{E}_t\left[\sum_{l=0}^{\infty} \frac{(-a)^l X_t^l}{l!}\right] \leq \mathbb{E}_t\left[1 - aX_t + X_t^2 \sum_{l=2}^{\infty} \frac{a^l}{l!}\right] \\
&= 1 - a\mathbb{E}_t[X_t] + \mathbb{E}_t[X_t^2] \cdot (e^a - 1 - a) \\
&\leq 1 - a\mathbb{E}_t[X_t] + a^2\mathbb{E}_t[X_t^2] \\
&\leq 1 + a\gamma Y_{t-1} + a\beta_l + a^2\gamma Y_{t-1} + a^2\sigma \\
&\leq \exp(a^2\sigma + a\beta_l + \gamma(1+a)aY_{t-1}),
\end{aligned}$$

where the first inequality is by the bounded difference property $|X_t| \leq 1$ and $a \geq 0$, the second inequality is because $e^p \leq 1 + p + p^2$ for $p \leq 1$, the third inequality is by the self-adjusting property and the bounded variance property and $a \geq 0$, and the last inequality is by $1 + p \leq e^p$ for $p \in \mathbb{R}$. Then we can bound the moment generating function of Y_t as

$$\begin{aligned}
\mathbb{E}_{Y_0, \dots, Y_t}[e^{-aY_t}] &= \mathbb{E}_{Y_0, \dots, Y_{t-1}}[e^{-aY_{t-1}} \cdot \mathbb{E}_t[e^{-aX_t}]] \\
&\leq \mathbb{E}_{Y_0, \dots, Y_{t-1}}[\exp(a^2\sigma + a\beta_l - a(1 - \gamma(1+a))Y_{t-1})] \\
&\leq \exp(a^2\sigma + a\beta_l) \cdot \mathbb{E}_{Y_0, \dots, Y_{t-1}}[\exp(-a(1 - (1+a)\gamma)Y_{t-1})] \\
&= \exp(a^2\sigma + a\beta_l) \cdot \mathbb{E}_{Y_0, \dots, Y_{t-1}}[\exp(-g(a) \cdot Y_{t-1})],
\end{aligned}$$

where we define $g(a) := a(1 - (1 + a)\gamma)$. Note that $g(a) \in [0, a]$ by the assumptions $\gamma \in (0, \frac{1}{2}]$ and $a \in [0, 1]$. Define the sequence $a_{(0)} = a$ and $a_{(i)} = g(a_{(i-1)})$ for $i \geq 1$. Apply the same argument inductively, it follows that

$$\mathbb{E}_{Y_0, \dots, Y_t} [e^{-aY_t}] \leq \exp \left[\sum_{i=0}^{t-1} (a_{(i)}^2 \sigma + a_{(i)} \beta_l) \right] \cdot \mathbb{E}_{Y_0} [e^{-a_{(t)} Y_0}] \leq \exp \left[\sum_{i=0}^{t-1} (a_{(i)}^2 \sigma + a_{(i)} \beta_l) + \frac{a_{(t)}^2 \sigma}{\gamma} \right],$$

where the last inequality follows from the initial concentration property of Y_0 . To bound the moment generating function, we use the following claim whose proof follows from the definition of the sequence $\{a_{(i)}\}_i$.

Claim 3.14. *The sequence $\{a_{(i)}\}_{i \geq 0}$ is decreasing and dominated by the geometric sequence $\{ar^i\}_{i \geq 0}$ with common ratio $r := 1 - \gamma$. The sequence $\{a_{(i)}^2\}_i$ is also decreasing and dominated by the geometric sequence $\{a^2 r^{2i}\}_{i \geq 0}$ with common ratio r^2 . Furthermore, $r^2 < r < 1$ when $a \in [0, 1]$.*

Using Claim 3.14, when $a \in [0, 1]$, we can upper bound the moment generating function by

$$\begin{aligned} \mathbb{E}_{Y_0, \dots, Y_t} [e^{-aY_t}] &\leq \exp \left[(a^2 \sigma + a \beta_l) \sum_{i=0}^{t-1} r^i + \frac{a^2 \sigma r^t}{\gamma} \right] \\ &= \exp \left[(a^2 \sigma + a \beta_l) \cdot \frac{1 - r^t}{1 - r} + \frac{a^2 \sigma r^t}{\gamma} \right] \\ &= \exp \left[\frac{a^2 \sigma + a \beta_l}{\gamma} \cdot (1 - r^t) + \frac{a^2 \sigma r^t}{\gamma} \right] \\ &\leq \exp \left[\frac{a^2 \sigma + a \beta_l}{\gamma} \right]. \end{aligned}$$

By Markov inequality, for any $a \in [0, 1]$ and any $\eta > 0$,

$$\begin{aligned} \Pr \left[Y_t \leq -\frac{\beta_l}{\gamma} - \eta \right] &= \Pr \left[e^{-aY_t} \geq e^{a \left(\frac{\beta_l}{\gamma} + \eta \right)} \right] \leq \mathbb{E}_{Y_0, \dots, Y_t} [e^{-aY_t}] / e^{a \left(\frac{\beta_l}{\gamma} + \eta \right)} \\ &\leq \exp \left[\frac{a^2 \sigma + a \beta_l}{\gamma} - a \left(\frac{\beta_l}{\gamma} + \eta \right) \right] \\ &= \exp \left[\frac{a^2 \sigma}{\gamma} - a \eta \right]. \end{aligned}$$

When $\eta \leq 2\sigma/\gamma$, we set $a = (\eta\gamma)/(2\sigma) \in [0, 1]$, so the above probability bound applies and gives

$$\Pr \left[Y_t \leq -\frac{\beta_l}{\gamma} - \tau \right] \leq \exp \left[-\frac{\eta^2 \gamma}{4\sigma} \right] \leq \exp \left[-\frac{\eta^2}{4\sigma/\gamma + \eta} \right].$$

When $\eta > 2\sigma/\gamma$, we simply set $a = 1$, and the above probability bound gives

$$\Pr \left[Y_t \leq -\frac{\beta_l}{\gamma} - \eta \right] \leq \exp \left[\frac{\sigma}{\gamma} - \eta \right] \leq \exp \left[-\frac{\tau^2}{4\sigma/\gamma + \eta} \right],$$

where the last inequality holds by the assumption that $\eta > 2\sigma/\gamma$. This finishes the proof for the lower tail and thus the proof of Theorem 2.7.

3.2 Two-Sided Spectral Rounding

In this section, we show that the two-sided spectral rounding result in Theorem 1.6 can be extended to incorporate one non-negative linear constraint that is given as part of the input.

There is a standard reduction used in [65] to construct spectral sparsifiers that satisfy additional linear constraints. Suppose Corollary 1.7 were to work for rank two matrices, then we can simply incorporate the linear constraint to the input matrices as $A_i := \begin{pmatrix} v_i v_i^T & 0 \\ 0 & c_i / \langle c, x \rangle \end{pmatrix}$ so that $\sum_{i=1}^m x_i A_i = I_{n+1}$, and any $z \in \{0, 1\}^m$ so that $\sum_{i=1}^m z_i A_i \approx I_{n+1}$ would have $\langle c, z \rangle \approx \langle c, x \rangle$. But the rank one assumption is crucial in the proof of Theorem 1.6 and it is an open problem to generalize it to work with higher rank matrices.

Our idea is to use the following signing trick, suggested to us by Akshay Ramachandran, to essentially carry out the same reduction using only rank one matrices. We state the results in a more general form, where $\sum_{i=1}^m x_i v_i v_i^T$ is not necessarily equal to the identity matrix, so that we can also apply them to additive spectral sparsifiers in Section 4.4.

Lemma 3.15. *Let $v_1, \dots, v_m \in \mathbb{R}^n$, $x \in [0, 1]^m$, and $c \in \mathbb{R}_+^m$. Suppose $\|\sum_{i=1}^m x_i v_i v_i^T\|_{\text{op}} \leq \lambda$ and $\|v_i\| \leq l$ for $1 \leq i \leq m$. Then there exists a signing $s_1, \dots, s_m \in \{\pm 1\}$ such that if we let $u_i := \begin{pmatrix} v_i \\ s_i \sqrt{c_i \lambda / \langle c, x \rangle} \end{pmatrix} \in \mathbb{R}^{n+1}$ then $\|\sum_{i=1}^m x_i u_i u_i^T\|_{\text{op}} \leq \lambda + l\sqrt{\lambda}$.*

Proof. By the definition of u_i ,

$$\begin{aligned} \sum_{i=1}^m x_i u_i u_i^T &= \begin{pmatrix} \sum_{i=1}^m x_i v_i v_i^T & \sum_{i=1}^m s_i x_i \sqrt{\frac{c_i \lambda}{\langle c, x \rangle}} v_i \\ \sum_{i=1}^m s_i x_i \sqrt{\frac{c_i \lambda}{\langle c, x \rangle}} v_i^T & \sum_{i=1}^m \frac{c_i x_i \lambda}{\langle c, x \rangle} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^m x_i v_i v_i^T & 0 \\ 0 & \lambda \end{pmatrix} + \begin{pmatrix} 0 & \sum_{i=1}^m s_i x_i \sqrt{\frac{c_i \lambda}{\langle c, x \rangle}} v_i \\ \sum_{i=1}^m s_i x_i \sqrt{\frac{c_i \lambda}{\langle c, x \rangle}} v_i^T & 0 \end{pmatrix} \end{aligned}$$

The operator norm of the second matrix is bounded by $\left\| \sum_{i=1}^m s_i x_i \sqrt{c_i \lambda / \langle c, x \rangle} v_i \right\|$. It follows from triangle inequality that $\|\sum_{i=1}^m x_i u_i u_i^T\|_{\text{op}} \leq \lambda + \left\| \sum_{i=1}^m s_i x_i \sqrt{c_i \lambda / \langle c, x \rangle} v_i \right\|$. We show that there is a signing $s_1, \dots, s_m \in \{\pm 1\}$ such that $\left\| \sum_{i=1}^m s_i x_i \sqrt{c_i \lambda / \langle c, x \rangle} v_i \right\| \leq l\sqrt{\lambda}$ and this will complete the proof. Take a uniform random signing and consider

$$\begin{aligned} \mathbb{E}_{s \in \{\pm 1\}^m} \left\| \sum_{i=1}^m s_i x_i \sqrt{\frac{c_i \lambda}{\langle c, x \rangle}} v_i \right\|^2 &= \sum_{i=1}^m \mathbb{E}_s \left[s_i^2 x_i^2 \|v_i\|^2 \frac{\lambda c_i}{\langle c, x \rangle} \right] + \sum_{i \neq j} \mathbb{E}_s \left[s_i s_j x_i x_j \langle v_i, v_j \rangle \frac{\lambda \sqrt{c_i c_j}}{\langle c, x \rangle} \right] \\ &= \sum_{i=1}^m x_i^2 \|v_i\|^2 \frac{\lambda c_i}{\langle c, x \rangle} \leq l^2 \sum_{i=1}^m \frac{\lambda c_i x_i}{\langle c, x \rangle} = l^2 \lambda, \end{aligned}$$

where the second line uses that $s_i^2 = 1$, $\mathbb{E}[s_i s_j] = \mathbb{E}[s_i] \cdot \mathbb{E}[s_j] = 0$, and $x_i \in [0, 1]$, $\|v_i\| \leq l$ in the inequality. This implies that there exists such a signing. \square

We apply the signing in Lemma 3.15 to incorporate one non-negative linear constraint into the two-sided spectral rounding result of Kyng, Luh and Song [45].

Theorem 3.16. *Let $v_1, \dots, v_m \in \mathbb{R}^n$, $x \in [0, 1]^m$, and $c \in \mathbb{R}_+^m$. Suppose $\|\sum_{i=1}^m x_i v_i v_i^T\|_{\text{op}} \leq \lambda$ and $\|v_i\| \leq l$ for $1 \leq i \leq m$. Suppose further that $c_\infty \leq l^2 \langle c, x \rangle / \lambda$ and $l \leq \sqrt{\lambda}$. Then there exists $z \in \{0, 1\}^m$ such that*

$$\left\| \sum_{i=1}^m x_i v_i v_i^T - \sum_{i=1}^m z_i v_i v_i^T \right\|_{\text{op}} \leq 8l\sqrt{\lambda} \quad \text{and} \quad |\langle c, x \rangle - \langle c, z \rangle| \leq \frac{8l}{\sqrt{\lambda}} \langle c, x \rangle$$

Proof. Let $u_i = \left(\frac{v_i}{s_i \sqrt{c_i \lambda / \langle c, x \rangle}} \right)$ for $1 \leq i \leq m$, where s_1, \dots, s_m is the signing given in Lemma 3.15.

By the assumption that $c_\infty \leq l^2 \langle c, x \rangle / \lambda$, it follows that $\|u_i\|^2 = \|v_i\|^2 + c_i \lambda / \langle c, x \rangle \leq 2l^2$. Let ξ_i be a zero-one random variable with probability x_i being one. Applying Theorem 1.6 on u_1, \dots, u_m and ξ_1, \dots, ξ_m , there exists $z \in \{0, 1\}^m$ such that

$$\left\| \sum_{i=1}^m x_i u_i u_i^T - \sum_{i=1}^m z_i u_i u_i^T \right\|_{\text{op}} \leq 4 \left\| \sum_{i=1}^m \mathbf{Var}[\xi_i] (u_i u_i^T)^2 \right\|_{\text{op}}^{1/2} \leq 4 \left\| \sum_{i=1}^m x_i \|u_i\|^2 u_i u_i^T \right\|_{\text{op}}^{1/2} \leq 4\sqrt{2l^2(\lambda + l\sqrt{\lambda})},$$

where we use that $\mathbf{Var}[\xi_i] = x_i(1 - x_i) \leq x_i$, $\|u_i\|^2 \leq 2l^2$ and $\|\sum_{i=1}^m x_i u_i u_i^T\|_{\text{op}} \leq \lambda + l\sqrt{\lambda}$ by Lemma 3.15. By looking at the top left $n \times n$ block, this implies that $\|\sum_{i=1}^m x_i v_i v_i^T - \sum_{i=1}^m z_i v_i v_i^T\|_{\text{op}} \leq 4\sqrt{2l^2(\lambda + l\sqrt{\lambda})} \leq 8l\sqrt{\lambda}$ where we use the assumption that $l \leq \sqrt{\lambda}$. By looking at the bottom right entry, we have

$$\left| \sum_{i=1}^m \frac{x_i c_i \lambda}{\langle c, x \rangle} - \sum_{i=1}^m \frac{z_i c_i \lambda}{\langle c, x \rangle} \right| \leq 4\sqrt{2l^2(\lambda + l\sqrt{\lambda})} \leq 8l\sqrt{\lambda} \quad \implies \quad |\langle c, x \rangle - \langle c, z \rangle| \leq \frac{8l}{\sqrt{\lambda}} \langle c, x \rangle.$$

\square

This proves Theorem 1.9 that incorporates one non-negative linear constraint into Corollary 1.7, by plugging $\lambda = 1$ and $l = \epsilon$ into Theorem 3.16.

3.3 Tight Examples

We provide two examples showing the tightness of Theorem 1.8.

First, consider the following simple example, which shows the nc_∞ additive error term is necessary.

Example 3.17. There are $m = 2n$ vectors $v_{11}, v_{12}, \dots, v_{n1}, v_{n2} \in \mathbb{R}^n$, a vector $x \in [0, 1]^m$, a vector $c \in \mathbb{R}_+^m$, and a parameter ϵ . They are defined as follows

$$\begin{aligned} x_{i1} &= 1, & v_{i1} &= \sqrt{1 - \epsilon} \cdot e_i, & c_{i1} &= 0 & \text{and} \\ x_{i2} &= \epsilon, & v_{i2} &= e_i, & c_{i2} &= c_\infty, & \forall i \in \{1, \dots, n\}. \end{aligned}$$

Note that $\langle c, x \rangle = \epsilon n c_\infty$ and $\sum_{i=1}^n \sum_{j=1,2} x_{ij} v_{ij} v_{ij}^T = I_n$.

Claim 3.18. For any constant $\alpha > 1$, any $z \in \{0, 1\}^m$ satisfying the spectral lower bound in Example 3.17 must have $\langle c, z \rangle \geq \alpha \langle c, x \rangle + \Omega(n c_\infty)$.

Proof. Note that the only vector $z \in \{0, 1\}^m$ that satisfies the spectral lower bound exactly is $z = \vec{1}_m$. This implies that $\langle c, z \rangle - \alpha \langle c, x \rangle = n c_\infty - \alpha \epsilon n c_\infty = (1 - \alpha \epsilon) n c_\infty$. For any $\alpha > 1$, there exists ϵ such that $\langle c, z \rangle - \alpha \langle c, x \rangle$ is at least say $n c_\infty / 2$. \square

Next, we modify an integrality gap example in [59] to show that, even if $c = \vec{1}$ and we are allowing integral-solution instead of zero-one solution, the additive error $O(n c_\infty / \epsilon)$ in Theorem 1.8 is best possible.

Example 3.19. The example contains $m = \binom{n}{2}$ vectors $v_1, \dots, v_m \in \mathbb{R}^{n-1}$, a vector $x \in [0, 1]^m$ and a vector $c = \vec{1}_m$. Let $\Pi \in \mathbb{R}^{(n-1) \times n}$ be the orthogonal projection onto the $(n-1)$ -dimensional subspace orthogonal to the all-one vector. Given some parameter k , we define

$$v_{ij} = \sqrt{\frac{n-1}{2k}} \cdot \Pi(\chi_i - \chi_j) \quad \text{and} \quad x_{ij} = \frac{2k}{n(n-1)}, \quad \forall 1 \leq i < j \leq n.$$

Note that $\langle c, x \rangle = k$ and $\sum_{i < j} x_{ij} v_{ij} v_{ij}^T = I_{n-1}$ and x has the smallest $\|x\|_1$ among all vectors satisfying $\sum x_{ij} v_{ij} v_{ij}^T \succeq I_{n-1}$.

We will use the following result from [59].

Theorem 3.20 (Theorem C.2 in [59]). Let $G = (V, E)$ be a graph with average degree $d_{\text{avg}} = 2m/n$, and let L_G be its unnormalized Laplacian matrix. Then, as long as d_{avg} is large enough, and n is large enough with respect to d_{avg} ,

$$\lambda_2(L_G) \leq d_{\text{avg}} - \rho \sqrt{d_{\text{avg}}},$$

where $\lambda_2(L_G)$ is the second smallest eigenvalue of L_G , and $\rho > 0$ is an absolute constant. Furthermore, the upper bound for $\lambda_2(L_G)$ still holds for graphs with parallel edges.

Using the above theorem, we can prove the following lemma.

Lemma 3.21. Let $\{v_{ij}\}, c, x$ be defined as in Example 3.19. For any $z \in \mathbb{Z}_+^m$, if $\sum_{1 \leq i < j \leq n} z_{ij} v_{ij} v_{ij}^T \succeq I_{n-1}$, then we have

$$\langle c, z \rangle \geq k + \Omega(\sqrt{kn} + n).$$

Proof. Given any $z \in \mathbb{Z}_+^m$, let G_z be the multi-graph corresponding to z with Laplacian matrix

$$L_z = \sum_{1 \leq i < j \leq n} z_{ij} (\chi_i - \chi_j)(\chi_i - \chi_j)^T = \frac{2k}{n-1} \Pi \left(\sum_{1 \leq i < j \leq n} z_{ij} v_{ij}^T v_{ij}^T \right) \Pi \succcurlyeq \frac{2k}{n-1} \left(I_n - \frac{1}{n} J_n \right),$$

where the last inequality holds by the assumption on z . Therefore, $\lambda_2(L_z) \geq 2k/(n-1)$.

On the other hand, since the average degree of G_z is $d_{\text{avg}} = 2\|z\|_1/n$, we apply Theorem 3.20 with properly chosen n , for some constant ρ we have

$$\lambda_2(L_z) \leq d_{\text{avg}} - \rho\sqrt{d_{\text{avg}}} \implies \lambda_2(L_z) \leq \frac{2\|z\|_1}{n} - \rho\sqrt{\frac{2\|z\|_1}{n}}.$$

Combining with $\lambda_2(L_z) \geq 2k/(n-1)$, we have

$$\frac{2k}{n-1} \leq \frac{2\|z\|_1}{n} - \rho\sqrt{\frac{2\|z\|_1}{n}} \implies 2k \leq 2\|z\|_1 - \rho\sqrt{2n\|z\|_1}.$$

For the quadratic inequality $2y^2 - \rho\sqrt{2n}y - 2k \geq 0$, we know that the nonnegative solution for y should satisfy

$$y \geq \frac{\rho\sqrt{2n} + \sqrt{2\rho^2n + 16k}}{4}.$$

Therefore, letting $y = \sqrt{\|z\|_1}$, we have

$$\begin{aligned} \langle c, z \rangle = \|z\|_1 &\geq (\rho\sqrt{2n} + \sqrt{2\rho^2n + 16k})^2/16 \\ &= \rho^2n/4 + k + \rho\sqrt{4\rho^2n + 32kn}/8 \\ &\geq k + \rho\sqrt{2kn}/2 + \rho^2n/4 \\ &\geq k + \Omega(\sqrt{kn} + n). \end{aligned}$$

□

Suppose we set the parameter $k = qn$ for $q > 16$ in Example 3.19. If we apply Theorem 1.8 to the vectors $v_1, \dots, v_m \in \mathbb{R}^{n-1}$ and $x \in [0, 1]^m$ defined in Example 3.19 with $\epsilon = \sqrt{n/k} < 1/4$, then there exists a $z \in \{0, 1\}^m$ such that

$$\sum_{1 \leq i < j \leq n} z_{ij} v_{ij} v_{ij}^T \succcurlyeq I_{n-1} \quad \text{and} \quad \langle c, z \rangle \leq (1 + 6\epsilon)\langle c, x \rangle + \frac{15nc_\infty}{\epsilon} = k + O(\sqrt{kn}),$$

where the last equality uses $\langle c, x \rangle = k$. Note that if the additive error term $O(\epsilon\langle c, x \rangle + nc_\infty/\epsilon)$ has a better dependency on ϵ , then we can set ϵ accordingly such that the cost upper bound will contradict with the lower bound in Lemma 3.21. For example, if Theorem 1.8 were improved to $\langle c, z \rangle \leq (1 + 6\epsilon)\langle c, x \rangle + 15nc_\infty/\sqrt{\epsilon}$, then we could set $\epsilon = (n/k)^{2/3}$ which would imply that $\langle c, z \rangle \leq k + O(k^{1/3}n^{2/3})$, contradicting with the lower bound $\langle c, z \rangle \geq k + \Omega(\sqrt{kn})$ when k is large enough. This shows Theorem 1.8 is tight up to a constant factor in the additive error term nc_∞/ϵ .

4 Applications

In this section, we will show that the spectral rounding results in Section 3 have many applications including survivable network design (Section 4.1), experimental design (Section 4.2), network design with spectral properties (Section 4.3) and unweighted spectral sparsification (Section 4.4).

4.1 General Survivable Network Design

We will show that the spectral rounding results provide a new approach to design algorithms for the survivable network design problem. The main advantage of this approach is that it significantly extends the scope of useful properties that can be incorporated into survivable network design.

The organization of this subsection is as follows. We begin by writing a large convex program that incorporates many useful constraints into survivable network design in Section 4.1.1, and explain how the spectral rounding results can be used to find a solution for this general survivable network design problem in Section 4.1.2. Then we will see the implications of Theorem 1.8 to network design in Section 4.1.3 and of Theorem 1.9 to network design in Section 4.1.4. Finally, we discuss how these new results make some progress towards Bansal's question [10] of designing an approximation algorithm for survivable network design with concentration property in Section 4.1.5.

4.1.1 Convex Programming Relaxation

We can write a convex programming relaxation for the general network design problem incorporating all these constraints as discussed in Section 1.4. In the following, the input graph is $G = (V, E)$ with $|V| = n$ and $|E| = m$. The fractional solution is $x \in \mathbb{R}^m$ where the intended solution is to set $x_e = 1$ if we choose edge e and $x_e = 0$ otherwise. We first present the convex program and then explain the constraints below.

$$\begin{array}{llll}
\min_x \langle c, x \rangle & & & \\
x(\delta(S)) \geq f(S) & \forall S \subseteq V & & \text{(connectivity constraints)} \\
x(\delta(v)) \leq d_v & \forall v \in V & & \text{(degree constraints)} \\
Ax \leq a & A \in \mathbb{R}_+^{p \times m}, a \in \mathbb{R}_+^p & & \text{(linear packing constraints)} \\
Bx \geq b & B \in \mathbb{R}_+^{q \times m}, b \in \mathbb{R}_+^q & & \text{(linear covering constraints)} \\
\text{Reff}_x(u, v) \leq r_{uv} & \forall u, v \in V & & \text{(effective resistance constraints)} \\
L_x \succcurlyeq M & M \succcurlyeq 0 & & \text{(spectral constraints)} \\
\lambda_2(L_x) \geq \lambda & & & \text{(algebraic connectivity constraint)} \\
0 \leq x_e \leq 1 & \forall e \in E & & \text{(capacity constraints)}
\end{array} \tag{CP}$$

Let us explain the constraints one by one. For the connectivity constraints, we have a connectivity requirement $f_{u,v}$ that there are at least $f_{u,v}$ edge-disjoint paths between every pair u, v of vertices. For each subset $S \subseteq V$, we let $f(S) := \max_{u,v: u \in S, v \notin S} f_{u,v}$ and write a constraint that at least $f(S)$ edges in $\delta(S)$ should be chosen, where $x(\delta(S))$ denotes $\sum_{e \in \delta(S)} x_e$. By Menger's theorem, if an

integral solution satisfies all these constraints, then all the connectivity requirements are satisfied. For the degree constraints, each vertex has a degree upper bound d_v and we write a constraint that at most d_v edges in $\delta(v)$ can be chosen, where $x(\delta(v)) := \sum_{e \in \delta(v)} x_e$. For the linear packing and covering constraints, all the entries in A, B, a, b are nonnegative, and we assume that A, B have at most a polynomial number of rows in n, m . For effective resistance constraints, we have an upper bound $r_{u,v}$ on the effective resistance between every pair $u, v \in V$. As in Section 2.2, we write $\text{Reff}_x(u, v) = b_{st}^T L_x^\dagger b_{st}$ as the effective resistance between u and v in the fractional solution x where each edge e has conductance x_e . In the spectral and the algebraic connectivity constraints, we write $L_x := \sum_{e \in E} x_e L_e$ as the Laplacian matrix of the fractional solution x where L_e is the Laplacian matrix of an edge as defined in Section 2.2. In the spectral constraint, we require that $L_x \succcurlyeq M$ for a positive semidefinite matrix M . One could have polynomially many constraints of this form (just as linear packing and covering constraints), but we only write one for simplicity. In the algebraic connectivity constraint, we require the second smallest eigenvalue of the Laplacian matrix of the solution is at least λ , which is related to the graph expansion of the fractional solution as described in Section 2.2.

This convex program can be solved by the ellipsoid method in polynomial time in n and m . There are exponentially many connectivity constraints but we can use a max-flow min-cut algorithm as a polynomial time separation oracle for these constraints (see e.g. [43]). Other linear constraints can easily be checked efficiently, as we assume there are only polynomially many of them. Next we consider the non-linear constraints. For the effective resistance constraints, it is known [38] that $\text{Reff}_x(u, v)$ is a convex function in x . For the algebraic connectivity constraint, it is known [37] that λ_2 is a concave function in x . For the spectral constraint, the feasible set is a positive semidefinite cone and is convex in x . So the feasible set for these non-linear constraints form a convex set. Also, these non-linear constraints can all be checked in polynomial time using standard numerical computations. Therefore, we can use the ellipsoid algorithm to find an ϵ -approximate solution to this convex program in polynomial time in n and m with dependency on ϵ being $\log(1/\epsilon)$.

4.1.2 Spectral Rounding

Suppose we are given an optimal solution x to the convex programming relaxation (CP). To design approximation algorithms, the task is to round this fractional solution x into an integral solution z so that z satisfies all the constraints and $\langle c, z \rangle$ is close to $\langle c, x \rangle$. There are many different types of constraints and it seems difficult to handle them simultaneously. In the spectral approach, the main observation is that if we can find an integral solution z such that $\sum_{e \in E} z_e L_e \approx \sum_{e \in E} x_e L_e$ and $\langle c, x \rangle \approx \langle c, z \rangle$, then all the constraints can be (approximately) satisfied simultaneously. We state this observation in the following lemma.

Lemma 4.1. *Let $x \in \mathbb{R}_+^m$ be a feasible solution x to (CP). For $\epsilon \in [0, \frac{1}{2}]$, any $z \in \mathbb{Z}_+^m$ satisfies*

$$\sum_{e \in E} z_e L_e \succcurlyeq (1 - \epsilon) \sum_{e \in E} x_e L_e \implies \begin{cases} z(\delta(S)) \geq (1 - \epsilon)f(S) \text{ for all } S \subseteq V \\ \text{Reff}_z(u, v) \leq (1 + 2\epsilon)r_{u,v} \text{ for all } u, v \in V \\ L_z \succcurlyeq (1 - \epsilon)M, \\ \lambda_2(L_z) \geq (1 - \epsilon)\lambda. \end{cases}$$

For $\epsilon \in [0, 1]$, any $z \in \mathbb{Z}_+^m$ satisfies

$$\sum_{e \in E} z_e L_e \preceq (1 + \epsilon) \sum_{e \in E} x_e L_e \implies z(\delta(v)) \leq (1 + \epsilon) d_v \text{ for all } v \in V.$$

Proof. Let $L_x := \sum_{e \in E} x_e L_e$ and $L_z := \sum_{e \in E} z_e L_e$. We start with the connectivity constraints. For any $S \subseteq V$, let $\chi_S \in \mathbb{R}^n$ be the characteristic vector of S with $\chi_S(i) = 1$ if $i \in S$ and zero otherwise. It is well-known that

$$\chi_S^T L_z \chi_S = \chi_S^T \left(\sum_{e \in E} z_e L_e \right) \chi_S = \sum_{e \in E} z_e \chi_S^T L_e \chi_S = \sum_{e \in \delta(S)} z_e = z(\delta(S))$$

and similarly $\chi_S^T L_x \chi_S = x(\delta(S))$. So, if $L_z \succcurlyeq (1 - \epsilon) L_x$, then for all $S \subseteq V$ we have

$$z(\delta(S)) = \chi_S^T L_z \chi_S \geq (1 - \epsilon) \chi_S^T L_x \chi_S = (1 - \epsilon) x(\delta(S)) \geq (1 - \epsilon) f(S).$$

For the effective resistance constraints, since $L_z \succcurlyeq (1 - \epsilon) L_x$, it implies that $L_z^\dagger \preceq (1 - \epsilon)^{-1} L_x \preceq (1 + 2\epsilon) L_x$ for $\epsilon \in [0, \frac{1}{2}]$, and thus

$$\text{Reff}_z(u, v) = b_{uv}^T L_z^\dagger b_{uv} \leq (1 + 2\epsilon) b_{uv}^T L_x^\dagger b_{uv} = (1 + 2\epsilon) \text{Reff}_x(u, v) \leq (1 + 2\epsilon) r_{u,v}.$$

The statements about the spectral lower bound and the algebraic connectivity constraint follows directly from the assumption that $L_z \succcurlyeq (1 - \epsilon) L_x$. Finally, for the degree constraints, suppose we are given $L_z \preceq (1 + \epsilon) L_x$, then it follows that

$$z(\delta(v)) = \chi_v^T L_z \chi_v \leq (1 + \epsilon) \chi_v^T L_x \chi_v = (1 + \epsilon) x(\delta(v)) \leq (1 + \epsilon) d_v.$$

□

Lemma 4.1 says that if z satisfies the spectral lower bound $L_z \succcurlyeq L_x$, then the solution z will simultaneously satisfy all connectivity constraints, effective resistance constraints, spectral constraints, and the algebraic connectivity constraint exactly. Moreover, if z also satisfies the spectral upper bound approximately, then the solution z will approximately satisfy all degree constraints as well.

4.1.3 Applications of One-Sided Spectral Rounding

We apply Theorem 1.8 to design approximation algorithms for network design problems that significantly extend the scope of existing techniques.

$$\text{cp} := \min_x \langle c, x \rangle$$

$x(\delta(S)) \geq f(S)$	$\forall S \subseteq V$	(connectivity constraints)	
$Ax \leq a$	$A \in \mathbb{R}_+^{p \times m}, a \in \mathbb{R}_+^p$	(linear packing constraints)	
$Bx \geq b$	$B \in \mathbb{R}_+^{q \times m}, b \in \mathbb{R}_+^q$	(linear covering constraints)	
$\text{Reff}_x(u, v) \leq r_{uv}$	$\forall u, v \in V$	(effective resistance constraints)	(CP1)
$L_x \succcurlyeq M$	$M \succcurlyeq 0$	(spectral constraint)	
$\lambda_2(L_x) \geq \lambda$		(algebraic connectivity constraint)	
$0 \leq x_e \leq 1$	$\forall e \in E$	(capacity constraints)	

In network design, a zero-one solution corresponds to a subset of edges where each edge is used at most once (satisfying the capacity constraints). The following theorem is a consequence of Theorem 1.8.

Theorem 4.2. *Suppose we are given an optimal solution x to the convex program (CP1). For any $\epsilon \in (0, \frac{1}{4})$, there is a polynomial time randomized algorithm to return a zero-one solution $z \in \{0, 1\}^m$ to (CP1) satisfying all the constraints exactly with probability at least $1 - \exp(-\Omega(n))$ except for the linear constraints. The solution z has objective value*

$$\langle c, z \rangle \leq (1 + 6\epsilon)\text{cp} + \frac{15nc_\infty}{\epsilon}$$

with probability at least $1 - \exp(-\Omega(n))$, and satisfies

$$\langle A_i, z \rangle \leq (1 + 6\epsilon)a_i + \frac{15n \|A_i\|_\infty}{\epsilon}$$

where A_i is the i -th row of A , with probability at least $1 - \exp(-\Omega(n))$ for each linear packing constraint, and satisfies

$$\langle B_j, z \rangle \geq b_j - \delta n \|B_j\|_\infty,$$

where B_j is the j -th row of B , with probability at least $1 - \exp(-\min\{\epsilon\delta, \epsilon\delta^2\} \cdot \Omega(n))$ for any $\delta > 0$ for each linear covering constraint.

Proof. We apply the following standard transformation to reduce to the one-sided spectral rounding problem. We assume without loss of generality that the graph G_x formed by the support of the fractional solution x is connected, and so L_x has rank $n - 1$. Let $\Pi = I_n - \frac{1}{n}J_n$ be the orthogonal projection onto the $n - 1$ dimensional subspace orthogonal to the all-one vector, where J_n is the $n \times n$ all-one matrix. For each edge $e \in E$, we define a vector $v_e := L_{G_x}^{\dagger/2} \Pi b_e$ which is contained in the $n - 1$ dimensional subspace orthogonal to the all-one vector. Then

$$\sum_{e \in E} x_e v_e v_e^T = L_x^{\dagger/2} \Pi \left(\sum_{e \in E} x_e b_e b_e^T \right) \Pi L_x^{\dagger/2} = L_x^{\dagger/2} \Pi L_x \Pi L_x^{\dagger/2} = I_{n-1}.$$

For any $\epsilon \in (0, \frac{1}{4})$, we apply Theorem 1.8 to $x \in [0, 1]^m$, $\{v_e\}_{e \in E}$ to find a zero-one solution $z \in \{0, 1\}^m$ such that $\sum_{e \in E} z_e v_e v_e^T \succcurlyeq I_{n-1}$ with probability at least $1 - \exp(-\Omega(n))$, which implies $\sum_{e \in E} z_e b_e b_e^T \succcurlyeq L_{G_x}$, thus the zero-one solution z satisfies all the constraints in (CP1) except for the linear constraints by Lemma 4.1.

Theorem 1.8 also guarantees that with probability at least $1 - \exp(-\Omega(n))$ the objective value of z is at most

$$\langle c, z \rangle \leq (1 + 6\epsilon)\langle c, x \rangle + \frac{15nc_\infty}{\epsilon}.$$

The guarantees for the linear packing constraints follow the same way as for the objective function, and the guarantees for the linear covering constraints follow from the lower bound part of Theorem 1.8. \square

We demonstrate the use of Theorem 4.2 in some concrete settings. The first example shows that Theorem 4.2 provides a spectral alternative to Jain’s iterative rounding algorithm to achieve $O(1)$ -approximation for a fairly general subclass of the survivable network design problem.

Example 4.3. *Theorem 4.2 is a constant factor approximation algorithm as long as $nc_\infty = O(\text{cp})$. Suppose that in our network design problem the average degree is at least d_{avg} and the costs on edges are positive integers with $c_\infty = O(d_{\text{avg}})$ (e.g. in the minimum k -edge-connected subgraph problem every vertex has degree at least k and $1 \leq c_e \leq O(k)$ for $e \in E$, or the solution requires a connected subgraph and $1 \leq c_e \leq O(1)$ for $e \in E$, etc). Then $\text{cp} \geq \Omega(d_{\text{avg}}n) \geq \Omega(c_\infty n)$ and Theorem 4.2 provides a constant factor approximation algorithm.*

The additive error term nc_∞ is the reason that we could not achieve constant factor approximation in general, but this term is unavoidable in the one-sided spectral rounding setting when we need to satisfy the spectral lower bound exactly. See Section 3.3 for examples showing the limitations. Heuristically, we can compute cp and if $nc_\infty = O(\text{cp})$ then we know Theorem 4.2 will provide good approximate solutions.

The second example shows that Theorem 4.2 returns good approximate solution to survivable network design while incorporating many other constraints simultaneously.

Example 4.4. *Suppose the connectivity requirement is to find a k -edge-connected subgraph, or more generally $f_{u,v} \geq k$ for all $u, v \in V$. Assume the cost c_e of each edge e is at least one. Then $\text{cp} \geq \Omega(kn)$.*

When the cost function satisfies $c_\infty = O(k)$, then Theorem 4.2 implies that there is a polynomial time randomized algorithm to return a simple k -edge-connected subgraph satisfying all the constraints in (CP1) except for the linear constraints (with some non-trivial guarantees), and the cost of the subgraph is at most a constant factor of the optimal value.

When the cost function satisfies $c_\infty = O(1)$, then Theorem 4.2 implies that there is a polynomial time randomized algorithm to return a k -edge-connected subgraph satisfying all the constraints in (CP1) except for the linear constraints, and the cost of the subgraph is at most $1 + O(1/\sqrt{k})$ factor of the optimal value by setting $\epsilon = \Theta(1/\sqrt{k})$.

The third example shows when the linear packing and covering constraints can be satisfied up to a multiplicative constant factor. See also Section 4.1.5 for a related question asked by Bansal [10].

Example 4.5. *For linear covering constraints, suppose they are of the form $\sum_{e \in F} x_e \geq b_j$ for some subset $F \subseteq E$ where $b_j \geq n$, then the returned solution z will almost satisfy this constraint as $\sum_{e \in F} z_e \geq b_j - \delta n \|B_j\|_\infty \geq (1 - \delta)b_j$ for some $\delta > 0$. So, these unweighted covering constraints with large right hand side can be incorporated into survivable network design, even though they can be unstructured. By a similar argument, any unweighted packing constraints with large right hand side will be only violated by at most a multiplicative constant factor with high probability. It was not known that Jain’s iterative rounding can be adapted to incorporate these linear covering and packing constraints.*

We will present more applications of Theorem 4.2 in Section 4.3, where they can be used to design approximation algorithms for network design problems with spectral requirements. These problems

were studied in the literature before but not much is known about approximation algorithms with performance guarantees.

4.1.4 Applications of Two-Sided Spectral Rounding

If we can achieve two-sided spectral rounding in network design, then we can also approximately satisfy the degree constraints by Lemma 4.1. However, to apply Theorem 1.9, we need to satisfy the assumption that the vector lengths are small. It is known that the vector lengths in the spectral rounding setting corresponds to the effective resistance of the edges in the fractional solution x . In the following, we describe when two-sided spectral rounding can be applied, and discuss what are the implications for network design.

$$\begin{aligned}
\text{cp} &:= \min_x \langle c, x \rangle \\
x(\delta(S)) &\geq f(S) & \forall S \subseteq V & \text{(connectivity constraints)} \\
x(\delta(v)) &\leq d_v & \forall v \in V & \text{(degree constraints)} \\
\text{Reff}_x(u, v) &\leq r_{uv} & \forall u, v \in V & \text{(effective resistance constraints)} \\
L_x &\succcurlyeq M & M \succcurlyeq 0 & \text{(spectral lower bound)} \\
\lambda_2(L_x) &\geq \lambda & & \text{(algebraic connectivity constraint)} \\
0 &\leq x_e \leq 1 & \forall e \in E & \text{(capacity constraints)}
\end{aligned} \tag{CP2}$$

Theorem 4.6. *Suppose we are given an optimal solution x to the convex program (CP2). For any $\epsilon \in [0, 1]$, if $\text{Reff}_x(u, v) \leq \epsilon^2$ for every $uv \in E$ and $c_\infty \leq \epsilon^2 \langle c, x \rangle$, then there exists a zero-one solution $z \in \{0, 1\}^m$*

$$(1 - O(\epsilon))L_x \preceq L_z \preceq (1 + O(\epsilon))L_x \quad \text{and} \quad (1 - O(\epsilon))\langle c, x \rangle \leq \langle c, z \rangle \leq (1 + O(\epsilon))\langle c, x \rangle$$

This implies that all the constraints of (CP2) will be approximately satisfied by z (e.g. $z(\delta(S)) \geq (1 - O(\epsilon))f(S)$ for all $S \subseteq V$ and $z(\delta(v)) \leq (1 + O(\epsilon))d_v$ for all $v \in V$) and the objective value of z is at most $(1 + O(\epsilon))\text{cp}$.

Proof. We apply the same standard transformation as in Theorem 4.2 to reduce to the two-sided spectral rounding problem. Let $\Pi = I_n - \frac{1}{n}J_n$ as defined in Theorem 4.2. For each edge e , we define a vector $v_e := L_x^{\dagger/2} \Pi b_e$ which is contained in the $n - 1$ dimensional subspace orthogonal to the all-one vector. Then $\sum_{e \in E} x_e v_e v_e^T = I_{n-1}$ as in Theorem 4.2. Using the assumption that $\text{Reff}_x(i, j) \leq \epsilon^2$ for every edge $ij \in E$, it follows that

$$\|v_{ij}\|^2 = b_{ij}^T L_x^\dagger b_{ij} = \text{Reff}_x(i, j) \leq \epsilon^2 \text{ for all } ij \in E,$$

and thus the assumption in Theorem 1.9 is satisfied. We can then apply Theorem 1.9 on $\{v_e\}_e$ and c to conclude that there exists $z \in \{0, 1\}^m$ such that

$$(1 - O(\epsilon))I_{n-1} \preceq \sum_{e \in E} z_e v_e v_e^T \preceq (1 + O(\epsilon))I_{n-1} \quad \text{and} \quad \langle c, z \rangle \leq (1 + O(\epsilon))\langle c, x \rangle.$$

By the definition of $v_e = L_x^{\dagger/2} \Pi b_e$, this implies that

$$(1 - O(\epsilon))L_x \preceq L_z = \sum_{e \in E} z_e b_e b_e^T \preceq (1 + O(\epsilon))L_x.$$

By Lemma 4.1, the zero-one solution z satisfies all the constraints of (CP2) approximately. \square

In the following, we compare Theorem 4.6 to Theorem 4.2.

1. **Approximation guarantees:** When Theorem 4.6 applies, it can handle degree constraints as well and basically preserves all properties of the fractional solution (e.g. upper bound and lower bound on every cut). It also gives strong approximation guarantee for the objective value, getting arbitrarily close to the optimal value. However, the constraints are only approximately satisfied, while in Theorem 4.2 they are exactly satisfied. Theorem 4.6 can only handle one linear constraint, which is used for the objective function, while Theorem 4.2 can handle many linear constraints simultaneously with an additive error term.
2. **Assumptions:** Theorem 4.2 apply without any assumptions, but Theorem 4.6 only applies when $\text{Reff}_x(u, v) \leq \epsilon^2$ for all $uv \in E$ and $c_\infty \leq \epsilon^2 \langle c, x \rangle$. The assumption about the cost is moderate, as it only requires the maximum cost of an edge is at most ϵ^2 fraction of the total cost of the solution, which should be satisfied in many applications with small ϵ . The main restriction is the first assumption about effective resistances, which may not be satisfied in network design applications, and we would like to provide some combinatorial characterizations under which the assumption will hold. Let $\text{Reff}_{\text{diam}} := \max_{u, v} \text{Reff}(u, v)$ be the effective resistance diameter of a graph; note that the maximum is taken over all pairs (not just for edges as required in Theorem 4.6). For example, it is known that [21] a d -regular graph with constant expansion has $\text{Reff}_{\text{diam}} \leq O(1/d)$. So, if the fractional solution x is close to a d -regular expander graph, then Theorem 4.6 can be applied with $\epsilon \geq 1/\sqrt{d}$. It is proved in [2] that a much milder expansion condition guarantees small effective resistance diameter. For example, in a d -regular graph G , as long as for some $0 < \delta \leq 1/2$,

$$|\delta(S)| \geq \Omega\left((d|S|)^{\frac{1}{2}+\delta}\right) \text{ for all } S \subseteq V \implies \text{Reff}_{\text{diam}} \leq O\left(\frac{1}{d^{2\delta}}\right).$$

Note that a d -regular graph with constant expansion satisfies the much stronger assumption that $|\delta(S)| \geq \Omega(d|S|)$. Informally, the above result only requires $|\delta(S)|$ to be roughly the square root of $d|S|$ to show that the graph has a small effective resistance diameter (e.g. 3-dimensional mesh). So, as long as the fractional solution x is a mild expander as defined in [2], the assumption in Theorem 4.6 will be satisfied with small ϵ . As another example, if the algebraic connectivity $\lambda_2(L_x)$ of the fractional solution is at least say $1/2\epsilon^2$, then we have $\text{Reff}_{\text{diam}} \leq \epsilon^2$ so that Theorem 4.6 can be applied. Heuristically, if one could add the constraints that $\text{Reff}_x(u, v) \leq \epsilon^2$ for $uv \in E$ so that the convex program (CP2) is still feasible without increasing the objective value too much, then one could then apply Theorem 4.6 to bound the integrality gap of the convex program.

3. **Algorithms:** There are polynomial time algorithms to return the solutions guaranteed in Theorem 4.2, while the proof of Theorem 4.6 is non-constructive. In network design, Theorem 4.2 give us approximation algorithms, while Theorem 4.6 only gives us integrality gap results for the convex programming relaxation (that there exists a zero-one solution almost satisfying all the constraints with objective value close to the optimal value).

4.1.5 Concentration Property in Survivable Network Design

Recently, Bansal [10] designed a rounding technique that achieves the guarantees by iterative rounding and randomized rounding simultaneously. Suppose there is an iterative rounding algorithm for a problem satisfying some technical assumptions. Bansal's algorithm will satisfy essentially the same guarantees of the iterative rounding algorithm, and simultaneously the following concentration property with $\beta = O(1)$ with respect to linear constraints as if the algorithm does independent randomized rounding.

Definition 4.7 (β -concentration). *Let $\beta \geq 1$. For a vector valued random variable $X = (X_1, \dots, X_m)$, where X_i are possible dependent 0-1 random variables, we say X is β -concentrated around the mean $x = (x_1, \dots, x_m)$ where $x_i = \mathbb{E}[X_i]$, if for every $a \in \mathbb{R}^n$ with $M := \max_i |a_i|$, $\langle a, X \rangle$ is well-concentrated and satisfies Bernstein's inequality up to a factor of β in the exponent, i.e.*

$$\Pr(\langle a, X \rangle - \langle a, x \rangle \geq t) \leq \exp\left(-\frac{t^2/\beta}{2(\sum_{i=1}^m a_i^2 x_i(1-x_i) + Mt/3)}\right).$$

Bansal showed various interesting application of his techniques, with x being the fractional solution to the linear programming relaxation and X being the zero-one solution output by the approximation algorithm. However, he left it as an open question whether there is an $O(1)$ -approximation algorithm for survivable network design (the guarantee achieved by Jain's iterative rounding algorithm) with $O(1)$ -concentration property.

Our iterative randomized swapping algorithms satisfy similar but weaker concentration properties. Let $x \in [0, 1]^m$ be the fractional solution to the one-sided spectral rounding problem. The algorithm in Theorem 1.8 will output a vector-valued random variable $X \in \{0, 1\}^m$ such that for any $a \in \mathbb{R}_+^n$ with $M := \max_i a_i$,

$$\mathbb{E}[\langle a, X \rangle] \leq (1+O(\epsilon))\langle a, x \rangle + O\left(\frac{nM}{\epsilon}\right) \quad \text{and} \quad \Pr(\langle a, X \rangle - \mathbb{E}[\langle a, X \rangle] \geq \eta) \leq \exp\left[-\Omega\left(\frac{\eta^2}{\sigma^2 + M\eta}\right)\right],$$

where n is the dimension of the problem (i.e. the dimension of the vectors) and $\sigma^2 = O(M(\langle a, x \rangle + nM/\epsilon))$ is a term related to the variance of the randomized swapping process. In other words, the random variable $\langle a, X \rangle$ is concentrated around the expected value $\mathbb{E}[\langle a, X \rangle]$, but the expected value $\mathbb{E}[\langle a, X \rangle]$ could deviate from $\langle a, x \rangle$ by $O(\epsilon\langle a, x \rangle + nM/\epsilon)$ and the concentration property is weaker than the one required in β -concentration, as the upper bound of σ^2 we can obtain is larger than the term $\sum_{i=1}^m a_i^2 x_i(1-x_i)$ in the β -concentration definition. We note that both Bansal's proof and our proof use Freedman's concentration inequality or its variant. Using Theorem 4.2, we made some progress towards Bansal's question.

Corollary 4.8. *Let $x \in [0, 1]^m$ be an optimal fractional solution to the survivable network design problem (i.e. (CP1) with only connectivity and capacity constraints). Suppose $nc_\infty = O(\langle c, x \rangle)$. Then there is a randomized polynomial time algorithm to return a solution $z \in \{0, 1\}^m$ to the survivable network design problem so that $\langle c, z \rangle \leq O(\langle c, x \rangle)$ with probability at least $1 - \exp(-\Omega(n))$. Furthermore, for any $a \in \mathbb{R}_+^n$ and $\delta \in (0, 1)$ it holds that $\langle a, x \rangle - \delta n \|a\|_\infty \leq \langle a, z \rangle \leq O(\langle a, x \rangle + n \|a\|_\infty)$ with probability at least $1 - O(\exp(-\Omega(\delta^2 n)))$.*

We remark that one can add linear constraints a to the convex program in our framework before we apply the rounding, so that we have some control over $\langle a, x \rangle$ of the fractional solution x and hence some control over $\langle a, z \rangle$ of the zero-one solution z . But it may not be possible to add linear constraints to the relaxation in Bansal’s setting, as adding constraints may make the underlying iterative rounding algorithm stops working (e.g. we do not know of an iterative rounding algorithm for the survivable network design problem with additional linear packing or covering constraints). See Example 4.5 for a related discussion. Our results suggest that the spectral approach is perhaps more suitable for achieving concentration property for survivable network design.

4.2 Experimental Design

In this subsection, we will apply the one-sided spectral rounding results to design approximation algorithms for weighted experimental design problems, extending the work in [4, 5] for (unweighted) experimental design problems. The presentations will mostly follow those in [4, 5].

4.2.1 Previous Work

Experimental design is classical in statistics and has found new applications in machine learning [61, 8, 5]. In the general problem, we would like to select k points from a large design pool $\{v_1, \dots, v_m\} \in \mathbb{R}^n$ to maximize the statistical efficiency regressed on the selected k design points. This can be formulated as a discrete optimization problem of choosing a subset $S \subseteq [m]$ of at most k vectors, so that its covariance matrix $\Sigma_S := \sum_{i \in S} v_i v_i^T$ has the smallest function value $f(\Sigma_S)$ for some objective function f . Some popular choices of f include

- A(verage)-optimality: $f_A(\Sigma) = \text{tr}(\Sigma^{-1})/n$,
- D(eterminant)-optimality: $f_D(\Sigma) = (\det \Sigma)^{-1/n}$,
- E(igen)-optimality: $f_E(\Sigma) = \lambda_{\max}(\Sigma^{-1})$,
- V(ariance)-optimality: $f_V(\Sigma) = \text{tr}(V \Sigma^{-1} V^T)$,
- G-optimality: $f_G(\Sigma) = \max \text{diag}(V \Sigma^{-1} V^T)$.

Many of these optimization problems are known to be NP-hard [25, 59], and we are interested in designing approximation algorithms for these problems. There are two settings in experimental design.

1. With Repetition: A vector can be chosen multiple times. This is equivalent to finding a vector $z \in \mathbb{Z}_+^m$ to minimize $f(\sum_{i=1}^m z_i v_i v_i^T)$ subject to the constraint that $\sum_{i=1}^m z_i \leq k$. This is common in statistic literature, where multiple measurements with respect to the same design point lead to different values with statistically independent noise.
2. Without Repetition: A vector can be chosen at most once. This is equivalent to finding a vector $z \in \{0, 1\}^m$ to minimize $f(\sum_{i=1}^m z_i v_i v_i^T)$ subject to the constraint that $\sum_{i=1}^m z_i \leq k$. This is more relevant in machine learning applications, as same data points often give the same result.

To design approximation algorithms for these discrete optimization problems, the approach in [4, 5] is to first solve a convex programming relaxation to obtain $x \in [0, 1]^m$ that minimizes $f(\sum_{i=1}^m x_i v_i v_i^T)$ subject to the constraint that $\sum_{i=1}^m x_i \leq k$, and then round it to $z \in \{0, 1\}^m$ with $\sum_{i=1}^m z_i \leq k$ and $f(\sum_{i=1}^m z_i v_i v_i^T) \leq \gamma f(\sum_{i=1}^m x_i v_i v_i^T)$ for some small constant $\gamma \geq 1$. Under some mild assumptions on the objective function f (which are satisfied for all the popular choices above), Allen-Zhu, Li, Singh and Wang [4, 5] showed that designing a polynomial time γ -approximation algorithm for the experimental design problem can be reduced to the following one-sided spectral rounding problem.

Problem 4.9. *Given $x \in [0, 1]^m$ with $\sum_{i=1}^m x_i \leq k$ and $\sum_{i=1}^m x_i v_i v_i^T = I_n$, find $z \in \{0, 1\}^m$ with $\sum_{i=1}^m z_i \leq k$ and $\sum_{i=1}^m z_i v_i v_i^T \succeq \frac{1}{\gamma} I_n$ in polynomial time.*

Theorem 1.5 proves that this one-sided spectral rounding problem is always solvable with $\gamma = 1 + \epsilon$ as long as $k \geq \Omega(n/\epsilon^2)$, using the regret minimization framework. This implies a $(1 + \epsilon)$ -approximation algorithm for a large class of experimental design problem as long as $k \geq \Omega(n/\epsilon^2)$, in both the with repetition and without repetition settings. The assumption that $k \geq \Omega(n/\epsilon^2)$ is shown to be necessary in achieving a $(1 + \epsilon)$ -approximation for E-optimal design [59]. For some other objective functions, it is possible to relax the assumption $k = \Omega(n/\epsilon^2)$: Singh and Xie [67] and Madan, Singh, Tantipongpipat and Xie [54] gave $(1 + \epsilon)$ -approximation algorithms for D-optimal design when $k = \Omega(n/\epsilon)$, and Nikolov, Singh and Tantipongpipat [59] gave a $(1 + \epsilon)$ -approximation algorithm for A-optimal design when $k = \Omega(n/\epsilon)$.

4.2.2 Weighted Experimental Design

We consider the generalization of the experimental design problem where different design points may have different costs. In this problem, we are given design points $\{v_1, \dots, v_m\} \in \mathbb{R}^n$ and a cost vector $c \in \mathbb{R}_+^m$ and a cost budget C , the objective is to choose a subset $S \subseteq [m]$ that minimizes $f(\sum_{i \in S} v_i v_i^T)$ subject to the constraint that $\sum_{i \in S} c_i \leq C$. The problem in the previous subsection is the special case when c is the all-one vector and $C = k$. We believe that this more general problem will be useful in applications, as it is natural that different experiments have different operation costs.

The approximate spectral rounding Theorem 3.3 imply the following one-sided spectral rounding results that satisfy the more general cost constraint (which includes the cardinality constraint as a special case).

Theorem 4.10. *Let $v_1, \dots, v_m \in \mathbb{R}^n$ and $x \in [0, 1]^m$. Let $c \in \mathbb{R}_+^m$ and $C = \langle c, x \rangle$. Suppose $\sum_{i=1}^m x_i v_i v_i^T = I_n$ and $C \geq 15nc_\infty/\epsilon^2$. For any $\epsilon \in (0, \frac{1}{2}]$, there is a randomized polynomial time algorithm that returns an integral solution $z \in \{0, 1\}^m$ such that $\langle c, z \rangle \leq C$ and $\sum_{i=1}^m z_i v_i v_i^T \succeq (1 - 4\epsilon)I_n$ with probability at least $1 - \exp(-\Omega(n))$.*

Proof. The idea is to scale down x then apply Theorem 3.3. We let $\alpha = 1 - 2\epsilon$ and set

$$y := \alpha x \quad \text{and} \quad u_i := \frac{v_i}{\sqrt{\alpha}} \quad \implies \quad \langle c, y \rangle = \alpha \langle c, x \rangle = \alpha C \quad \text{and} \quad \sum_{i=1}^m y_i u_i u_i^T = \sum_{i=1}^m x_i v_i v_i^T = I_n.$$

We apply Theorem 3.3 on u_1, \dots, u_m and y, c with $\delta_1 = \epsilon, q = \sqrt{n}$ to obtain $z \in \{0, 1\}^m$ so that

$$\sum_{i=1}^m z_i u_i u_i^T \succeq (1 - 2\epsilon)I_n \quad \implies \quad \sum_{i=1}^m z_i v_i v_i^T \succeq \alpha(1 - 2\epsilon)I_n \succeq (1 - 4\epsilon)I_n$$

and

$$\langle c, z \rangle \leq (1 + \epsilon)\langle c, y \rangle + 15nc_\infty/\epsilon \leq (1 + \epsilon)(1 - 2\epsilon)C + \epsilon C < C,$$

where we use the assumptions that $15nc_\infty/\epsilon^2 \leq C$. The failure probability is at most $\exp(-\Omega(n))$. \square

Using the same reduction in [4, 5], Theorem 4.10 implies the following approximation algorithms for weighted experimental design, including the weighted version of A/D/E/V/G-design.

Theorem 4.11. *Suppose we are given m design points that are represented by n -dimensional vectors $v_1, \dots, v_m \in \mathbb{R}^n$, a cost vector $c \in \mathbb{R}_+^m$ and a cost budget $C \in \mathbb{R}_+$. Assuming that the objective function f satisfies the monotonicity, reciprocal sub-linearity and the polynomial time approximability conditions as described in [4, 5] (which hold for A/D/E/V/G-design), we have the following approximation results for weighted experimental design.*

For any fixed $\epsilon \leq \frac{1}{5}$, if $C \geq 15nc_\infty/\epsilon^2$, then there exists a polynomial time randomized algorithm that returns an integral vector $z \in \{0, 1\}^m$ such that with probability at least $1 - \exp(-\Omega(n))$ it holds that

$$f\left(\sum_{i=1}^m z_i v_i v_i^T\right) \leq (1 + O(\epsilon)) \cdot \min_{y \in [0, 1]^m : \langle c, y \rangle \leq C} f\left(\sum_{i=1}^m y_i v_i v_i^T\right) \quad \text{and} \quad \langle c, z \rangle \leq C.$$

We note that the algorithms in Theorem 4.11 can incorporate some additional linear packing and covering constraints, with the same guarantees as in Theorem 4.2.

Finally, we mention that the two-sided spectral rounding result can also be applied to weighted experimental design. Assuming all the vectors have length at most ϵ , it shows that there is a zero-one solution which achieves $(1 + O(\epsilon))$ -approximation in weighted experimental design, but it does not provide a polynomial time algorithm to find such a zero-one solution.

4.3 Spectral Network Design

There are several previous work on network design problems with spectral requirements. In this section, we will see that these problems are special cases of the general network design problem and the weighted experimental design problem in Section 4.1 and Section 4.2, and our results provide improved approximation algorithms for these problems and also generalize these problems to incorporate many additional constraints.

4.3.1 Maximizing Algebraic Connectivity

Ghosh and Boyd [37] study the problem of choosing a subgraph that maximizes the algebraic connectivity (the second smallest eigenvalue of its Laplacian matrix) subject to a cost constraint. The problem is formulated as follows:

$$\begin{aligned} \lambda_{\text{opt}} &:= \max_{x \in \mathbb{R}^{|E|}} \lambda_2 \left(\sum_{e \in E} x_e b_e b_e^T \right) \\ \text{subject to} \quad & \sum_{e \in E} c_e x_e \leq C, \\ & x_e \in \{0, 1\}, \forall e \in E, \end{aligned} \tag{4.1}$$

where c_e is the cost of edge e for $e \in E$ and C is the given cost budget. As mentioned in [37], the algebraic connectivity is a good measure on the well-connectedness of a graph, as

$$\lambda_2(L_G) \leq \min_{S \subseteq V} \frac{n|\delta(S)|}{|S||\bar{S}|} \leq 2 \min_{0 \leq |S| \leq \frac{n}{2}} \frac{|\delta(S)|}{|S|}$$

where the first inequality is proved in [29]. Thus, any graph with large λ_{opt} has no sparse cuts, which also implies that the mixing time of random walks is small.

Ghosh and Boyd show that if the constraint $x_e \in \{0, 1\}$ is relaxed to $x_e \in [0, 1]$, then the relaxation is convex and can be written as a semidefinite program. They proposed a greedy heuristic based on the Fiedler vector for the zero-one cost setting (where $c_e \in \{0, 1\}$ for all e), but they do not provide any approximation guarantee of their heuristic algorithm.

Kolla, Makarychev, Saberi and Teng [44] provide the first algorithm with non-trivial approximation guarantee in the zero-one cost setting. Using subgraph sparsification techniques, they give an algorithm that returns a solution which violates the cost constraint by a factor of at most 8 and having algebraic connectivity at least $\Omega(\lambda_{\text{opt}}^2/\Delta)$ where Δ is the maximum degree of the graph.

We observe that if we project the vectors b_e onto the rank $n - 1$ subspace orthogonal to the all-one vector, then the objective function of (4.1) is simply the reciprocal of the objective function of the E-optimal design problem described in Section 4.2. This immediately implies that the result of Allen-Zhu, Li, Singh and Wang [5] can be applied to give a $(1 + \epsilon)$ -approximation algorithm for the unweighted problem as long as $C \geq 5n/\epsilon^2$, although this connection was not made before. Theorem 4.11 implies the following approximation result for general non-negative cost function.

Theorem 4.12. Suppose $C \geq 15nc_\infty/\epsilon^2$ for some $\epsilon \in (0, \frac{1}{2}]$. There is a polynomial time randomized algorithm which returns a zero-one solution $z \in \{0, 1\}^m$ for (4.1) with probability at least $1 - \exp(-\Omega(n))$ such that

$$\lambda_2 \left(\sum_{e \in E} z_e b_e b_e^T \right) \geq (1 - O(\epsilon)) \lambda_{\text{opt}} \quad \text{and} \quad \sum_{e \in E} c_e z_e \leq C.$$

As shown in Section 4.1, the constraint $\lambda_2(\sum_{e \in E} x_e b_e b_e^T) \geq \lambda_{\text{opt}}$ can be incorporated into network design, and so Theorem 4.2 implies the following result.

Theorem 4.13. There is a polynomial time randomized algorithm which returns a zero-one solution $z \in \{0, 1\}^m$ with probability at least $1 - \exp(-\Omega(n))$ such that

$$\lambda_2 \left(\sum_{e \in E} z_e b_e b_e^T \right) \geq \lambda_{\text{opt}} \quad \text{and} \quad \sum_{e \in E} c_e z_e \leq (1 + O(\epsilon))C + O\left(\frac{nc_\infty}{\epsilon}\right).$$

Furthermore, this can be done while incorporating other constraints (e.g. connectivity constraints) as described in Theorem 4.2.

4.3.2 Minimizing Total Effective Resistance

Ghosh, Boyd and Saberi [38] study the problem of designing a network that minimizes the total effective resistance. The problem is formulated as follows.

$$\begin{aligned} R_{\text{opt}} &:= \min_{x \in \mathbb{R}^{|E|}} \frac{1}{2} \sum_{u,v \in V} \text{Reff}_x(u, v) \\ \text{subject to} \quad &\sum_{e \in E} x_e \leq k, \\ &x_e \in \{0, 1\}, \forall e \in E. \end{aligned} \tag{4.2}$$

They showed that if the constraint $x_e \in \{0, 1\}$ is relaxed to $x_e \in [0, 1]$, then the relaxation is convex and can be written as a semidefinite program. They did not provide any result for the discrete optimization version in (4.2).

Ghosh, Boyd and Saberi [38] also show that the total effective resistance is a useful measure in different problems, e.g. average commute time, power dissipation in a resistor network, elmore delay in a RC Circuit, total time constant of an averaging network, and euclidean variance. Furthermore, they established a connection between (4.2) and the A-design problem described in Section 4.2. To see this, note that the objective of (4.2) can be written as

$$\frac{1}{2} \sum_{u,v \in V} \text{Reff}_x(u, v) = \frac{1}{2} \sum_{u \neq v \in V} b_{uv}^T L_x^\dagger b_{uv} = \left\langle L_x^\dagger, \frac{1}{2} \sum_{u \neq v \in V} b_{uv} b_{uv}^T \right\rangle = \left\langle L_x^\dagger, nI_n - \mathbf{1}_n \mathbf{1}_n^T \right\rangle = n \text{tr} \left(L_x^\dagger \right),$$

where the last equality follows as L_x^\dagger is orthogonal to 1_n . Hence, minimizing total effective resistance is equivalent to minimizing $\text{tr}(L_{G_x}^\dagger) = \text{tr}(\sum_{e \in E} x_e b_e b_e^T)^\dagger$, which is the same as the A-design objective function after we project the vectors onto the subspace orthogonal to the all-one vector.

With this connection, all the recent algorithms for the A-optimal design can be applied to solve (4.2). For instances, the regret minimization algorithm in [5] gives a $(1+\epsilon)$ -approximation algorithm when $k \geq \Omega(n/\epsilon^2)$, and the proportional volume sampling in [59] achieves $(1+\epsilon)$ -approximation with weaker assumption $k \geq \Omega(n/\epsilon)$.

Theorem 4.11 implies the following approximation result for the more general weighted setting, where every edge has a cost c_e and we are given a cost budget C as in (4.1).

Theorem 4.14. *Suppose $C \geq 15nc_\infty/\epsilon^2$. There is a polynomial time randomized $(1 + O(\epsilon))$ -approximation algorithm for the weighted version of (4.2).*

As shown in Section 4.1, the effective resistance constraints can be incorporated into network design, and so Theorem 4.2 implies the following result.

Theorem 4.15. *There is a polynomial time randomized algorithm which returns a zero-one solution $z \in \{0, 1\}^m$ with probability at least $1 - \exp(-\Omega(n))$ such that*

$$\frac{1}{2} \sum_{u,v \in V} \text{Reff}_z(u,v) \leq R_{\text{opt}} \quad \text{and} \quad \sum_{e \in E} c_e x_e \leq (1 + O(\epsilon))C + O\left(\frac{nc_\infty}{\epsilon}\right).$$

Furthermore, this can be done while incorporating other constraints (e.g. connectivity constraints) as described in Theorem 4.2.

4.3.3 Network Design for Effective Resistances

In [20], together with Chan, Schild and Wong, we consider the following new problem about network design for s - t effective resistance. Given a graph $G = (V, E)$ and two vertices $s, t \in V$, find a subgraph H with at most k edges to minimize the effective resistance between s and t . The main result in [20] is a constant factor approximation algorithm for the problem. This result motivates the current paper.

Using the results in Section 4.2, we can generalize the problem by allowing the edges to have costs and considering the sum of effective resistance of multiple pairs. Using the results in Section 4.1, we can add the effective resistance constraints for multiple pairs with the objective of minimizing the cost of the solution subgraph, while incorporating other constraints as described in Theorem 4.2.

4.4 Unweighted Spectral Sparsification

We show that the spectral rounding results can also be applied to the study of unweighted spectral sparsification.

4.4.1 Previous Work

Batson, Spielman, and Srivastava [15] proved that any graph has a $(1 \pm \epsilon)$ -spectral sparsifier with only $O(n/\epsilon^2)$ edges, by carefully reweighting the edges of the original graph where different edges may have different weights. If we require all the edges to have the same weight, then there are simple examples (e.g. barbell graphs) showing that linear-sized spectral sparsification is not always possible. In a recent paper [14], Bansal, Svensson and Trevisan ask whether there is a non-trivial notion of unweighted spectral sparsification with which linear-sized spectral sparsification is always possible. They study a notion suggested by Oveis Gharan.

Definition 4.16 (Additive Unweighted Spectral Sparsifier). *Given a graph $G = (V, E)$ with n vertices, m edges and maximum degree d , a subgraph $\tilde{G} = (V, F)$ with \tilde{m} edges is an additive spectral sparsifier with error $\epsilon \in [0, 1]$ if*

$$-\epsilon d I \preceq \frac{m}{\tilde{m}} L_{\tilde{G}} - L_G \preceq \epsilon d I.$$

Bansal, Svensson and Trevisan [14] prove that sparse additive unweighted spectral sparsification is always possible, and they provide both deterministic and randomized algorithms for constructing these sparsifiers.

Theorem 4.17 (Randomized Construction [14]). *Given a graph $G = (V, E)$ with n vertices, m edges, maximum degree d , and $\epsilon \in (0, 1)$, there is a polynomial time randomized algorithm that finds a subset of edges $F \subseteq E$ with size $\tilde{m} = |F| = O(n \log(1/\epsilon)^3 / \epsilon^2)$ such that $\tilde{G} = (V, F)$ satisfies*

$$-\epsilon d I \preceq \frac{m}{\tilde{m}} L_{\tilde{G}} - L_G \preceq \epsilon d I.$$

Theorem 4.18 (Deterministic Construction [14]). *Given a graph $G = (V, E)$ with n vertices, m edges, maximum degree d , and $\epsilon \in (0, 1)$, there is a polynomial time deterministic algorithm that finds a multi-set F of edges with size $\tilde{m} = |F| = O(n/\epsilon^2)$ such that $\tilde{G} = (V, F)$ satisfies*

$$2 \frac{m}{\tilde{m}} D_{\tilde{G}} - 2 D_G - \epsilon d I \preceq \frac{m}{\tilde{m}} L_{\tilde{G}} - L_G \preceq \epsilon d I,$$

where D_G is the diagonal degree matrix of G .

The proof of Theorem 4.17 is by Lovász local lemma and the converse of expander mixing lemma by Bilu and Linial. The proof of Theorem 4.18 is by the regret minimization framework of Allen-Zhu, Liao and Orecchia [3].

Note that Theorem 4.18 has a slightly weaker spectral lower bound guarantee than Theorem 4.17. Also, Theorem 4.18 can only return a multi-set solution where some edges can be used more than once, and so the sparsifier is integer weighted rather than unweighted where every edge has the same weight.

4.4.2 Nonconstructive Spectral Rounding and Unweighted Spectral Sparsification

We show that the existence of a linear-sized additive unweighted spectral sparsifier follows from the two-sided rounding result in Theorem 3.16. The idea is to view the original graph as a fractional solution where every edge e has $x_e = \tilde{m}/m$, and then use Theorem 3.16 to round this fractional solution to a zero-one solution while preserving the spectral properties of the original graph. The additional linear constraint in Theorem 3.16 allows us to bound the number of edges in the sparsifier.

Theorem 4.19. *Suppose we are given a graph $G = (V, E)$ with n vertices, m edges, and maximum degree d . Let $\tilde{m} = n/\epsilon^2$. For any $\epsilon \in (0, 1]$, there exists a subset of edges $F \subseteq E$ with $|F| \leq 8n/\epsilon^2$ such that*

$$-8\sqrt{2}\epsilon d I_n \preceq L_G - \frac{m}{\tilde{m}} \sum_{e \in F} b_e b_e^T \preceq 8\sqrt{2}\epsilon d I_n.$$

Proof. The plan is to apply Theorem 3.16 with $v_e := b_e$, $x_e := \tilde{m}/m$ and $c := \vec{1}_m$. We will first define the parameters λ and l and check that the assumptions $l \leq \sqrt{\lambda}$ and $c_\infty \leq l^2 \langle c, x \rangle / \lambda$ in Theorem 3.16 are satisfied. Note that

$$\left\| \sum_{e \in E} x_e v_e v_e^T \right\|_{\text{op}} = \frac{\tilde{m}}{m} \|L_G\|_{\text{op}} \leq \frac{2d\tilde{m}}{m} \quad \text{and} \quad \|v_e\| = \sqrt{2} \text{ for all } e \in E.$$

So we define $\lambda := 2d\tilde{m}/m$ and $l := \sqrt{2}$. We check that $\lambda = 2d\tilde{m}/m = 2dn/(\epsilon^2 m) \geq 2/\epsilon^2 \geq 2 = l^2$, and $l^2 \langle c, x \rangle / \lambda = 2\tilde{m}/(2d\tilde{m}/m) = m/d \geq 1 = c_\infty$. Therefore, we can apply Theorem 3.16 to conclude that there exists a subset of edges $F \subseteq E$ (corresponding to the zero-one solution z) such that

$$\left\| \sum_{e \in E} x_e v_e v_e^T - \sum_{e \in F} v_e v_e^T \right\|_{\text{op}} \leq 16\sqrt{\frac{d\tilde{m}}{m}} \quad \text{and} \quad \left| \sum_{e \in E} x_e c_e - \sum_{e \in F} c_e \right| \leq 8\sqrt{\frac{m}{d\tilde{m}}} \cdot \langle c, x \rangle.$$

Plugging in $x_e = \tilde{m}/m$ and $c = \vec{1}$ and $\tilde{m} = n/\epsilon^2$, the first statement implies that

$$\left\| L_G - \frac{m}{\tilde{m}} \sum_{e \in F} v_e v_e^T \right\|_{\text{op}} = \left\| \sum_{e \in E} v_e v_e^T - \frac{m}{\tilde{m}} \sum_{e \in F} v_e v_e^T \right\|_{\text{op}} \leq 16\sqrt{\frac{dm}{\tilde{m}}} = 16\sqrt{\frac{\epsilon^2 dm}{n}} \leq 8\sqrt{2}\epsilon d,$$

where the last inequality uses $m \leq dn/2$ as the maximum degree is d . Finally, the second statement implies that

$$|\tilde{m} - |F|| \leq 8\sqrt{\frac{\epsilon^2 m}{dn}} \cdot \tilde{m} \leq 4\sqrt{2}\epsilon \tilde{m} \implies |F| \leq (1 + 4\sqrt{2}\epsilon)\tilde{m} \leq \frac{8n}{\epsilon^2}.$$

□

Note that Theorem 4.19 improves Theorem 4.17 slightly by removing a factor of $\log^3(1/\epsilon)$ in the number of edges of the sparsifier. This confirms the existence of unweighted additive spectral

sparsifiers with $O(n/\epsilon^2)$ edges, which was not known before. More generally, we can use the same proof with a cost function c with $c_\infty \leq \|c\|_1/d$ to obtain a sparsifier with $\tilde{m} = n/\epsilon^2$ and

$$\left\| L_G - \frac{m}{\tilde{m}} \sum_{e \in F} v_e v_e^T \right\|_{\text{op}} \leq 8\sqrt{2}\epsilon d \quad \text{and} \quad (1 - 4\sqrt{2}\epsilon) \sum_{e \in E} c_e \leq \frac{m}{\tilde{m}} \sum_{e \in F} c_e \leq (1 + 4\sqrt{2}\epsilon) \sum_{e \in E} c_e.$$

We remark that the same reduction in [14] can be used to replace dI by $(D_G + d_{\text{avg}})I$ where D_G is the diagonal degree matrix of G and d_{avg} is the average degree in G .

The main disadvantage of Theorem 4.19 is that it does not provide a polynomial time algorithm to find such a sparsifier. It is a major open problem to make the method of interlacing polynomials used in [56, 57, 45] constructive.

4.4.3 Constructive Spectral Rounding and Unweighted Spectral Sparsification

For the deterministic algorithm, using similar techniques in [4, 5] which proves Lemma 2.5, we can strengthen Theorem 4.18 by returning a subgraph with no parallel edges.

Theorem 4.20. *Given a graph $G = (V, E)$ with n vertices, m edges, maximum degree d , and $\epsilon \in (0, 1)$, there is a polynomial time deterministic algorithm that finds a subset F of edges with size $\tilde{m} = |F| = O(n/\epsilon^2)$ such that $\tilde{G} = (V, F)$ satisfies*

$$2\frac{m}{\tilde{m}}D_{\tilde{G}} - 2D_G - O(\epsilon)dI_n \preceq \frac{m}{\tilde{m}}L_{\tilde{G}} - L_G \preceq O(\epsilon)dI_n.$$

The algorithm is a slight modification of the algorithm in [14], which is a greedy algorithm based on the regret minimization framework. The feedback matrices are of the following form

$$F_0 = 0 \quad \text{and} \quad F_t = \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix} - 2dI_{2n} \quad \text{for some } e \in E \text{ and } t \geq 1,$$

where L_G is the Laplacian matrix of the original graph, $L_G^+ := D_G + A_G$ is the signless-Laplacian of the original graph, and L_e and L_e^+ are the Laplacian and signless-Laplacian matrix of a single edge e . Note that we always have $F_t \preceq 0$, as $L_G \preceq 2dI_n$ and $L_G^+ \preceq 2dI_n$ for a graph G of maximum degree d .

Greedy Additive Spectral Sparsification

Input: An error parameter $\epsilon \in (0, 1)$, and a graph $G = (V, E)$ with n vertices, $m \geq 2n/\epsilon^2$ edges and maximum degree d .

Output: A subgraph \tilde{G} of G with $\tilde{m} = O(n/\epsilon^2)$ edges satisfying

$$2\frac{m}{\tilde{m}}D_{\tilde{G}} - 2D_G - \epsilon dI_n \preceq \frac{m}{\tilde{m}}L_{\tilde{G}} - L_G \preceq \epsilon dI_n.$$

1. Initialization: Set $S_0 := \emptyset$, $F_0 := 0$, $\tau := n/\epsilon^2$, and $\alpha = q\epsilon/\sqrt{dm}$ for some small $q > 0$.

2. For $t = 1$ to τ do

(a) Compute the action matrix $A_t = (\alpha \sum_{j=0}^{t-1} F_j + l_t I)^{-2}$, where $l_t \in \mathbb{R}$ is the unique value such that $A_t \succ 0$ and $\text{tr}(A_t) = 1$.

(b) Select an edge $e_t \in E \setminus S_{t-1}$ such that

$$\left\langle A_t, \begin{pmatrix} L_G - mL_{e_t} & \\ & L_G^+ - mL_{e_t}^+ \end{pmatrix} \right\rangle \geq -\frac{2\sqrt{n}}{\alpha m} = -O(\epsilon)d.$$

(c) Set

$$F_t := \begin{pmatrix} L_G - mL_{e_t} & \\ & L_G^+ - mL_{e_t}^+ \end{pmatrix} - 2dI_{2n} \quad \text{and} \quad S_t := S_{t-1} \cup \{e_t\}.$$

3. Return $\tilde{G} = (V, S_\tau)$ as the solution.

Note that we can assume $m \geq 2n/\epsilon^2 = 2\tau$, as otherwise we can simply return $\tilde{G} = G$ as our solution. The only difference with the algorithm in [14] is in Step 2(b), where we insist on choosing an edge $e_t \in E \setminus S_{t-1}$ to guarantee that the returned solution is a simple subgraph. If there is no such restriction, then a simple averaging argument in [14] shows that there is an edge $e \in E$ with the inner product in Step 2(b) being non-negative. With this restriction, we will use the closed-form of the action matrix and Lemma 2.5 to show that there is still an edge with the inner product in Step 2(b) being not too small. The following lemma is the new ingredient for the proof of Theorem 4.20.

Lemma 4.21. *For each $1 \leq t \leq \tau$, there always exists an edge $e \in E \setminus S_{t-1}$ such that*

$$\left\langle A_t, \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix} \right\rangle \geq -\frac{2\sqrt{n}}{\alpha m} \geq -O(\epsilon)d.$$

Proof. The sum of the inner product over all edges in $E \setminus S_{t-1}$ is

$$\begin{aligned} & \sum_{e \in E \setminus S_{t-1}} \left\langle A_t, \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix} \right\rangle \\ &= \sum_{e \in E} \left\langle A_t, \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix} \right\rangle - \sum_{e \in S_{t-1}} \left\langle A_t, \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix} \right\rangle \\ &= \left\langle A_t, \begin{pmatrix} mL_G - m \sum_{e \in E} L_e & \\ & mL_G^+ - m \sum_{e \in E} L_e^+ \end{pmatrix} \right\rangle - \left\langle A_t, \sum_{e \in S_{t-1}} \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix} \right\rangle \\ &= - \left\langle A_t, \sum_{e \in S_{t-1}} \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix} \right\rangle, \end{aligned}$$

where the last equality follows from $\sum_{e \in E} L_e = L_G$ and $\sum_{e \in E} L_e^+ = L_G^+$. Let

$$Z_{t-1} := \sum_{e \in S_{t-1}} \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix},$$

and let the eigenvalues of Z_{t-1} be $\lambda_1, \dots, \lambda_{2n}$. Note that $\lambda_{\min}(Z_{t-1}) \leq 0$ as $\text{tr}(L_G) = \text{tr}(L_G^+) = 2m$ which implies that $\text{tr}(Z_{t-1}) = 0$.

Observe that $A_t = (l_t I_{2n} + \alpha Z_{t-1} - 2\alpha(t-1)d I_{2n})^{-2}$ and so A_t and Z_{t-1} have the same eigenbasis, and the i -th eigenvalue of A_t is $(l_t + \alpha\lambda_i - 2\alpha(t-1)d)^{-2}$. It follows that

$$\begin{aligned} -\langle A_t, Z_{t-1} \rangle &= \sum_{i=1}^{2n} \frac{-\lambda_i}{(l_t + \alpha\lambda_i - 2\alpha(t-1)d)^2} \\ &= \sum_{i=1}^{2n} \frac{l_t/\alpha - 2(t-1)d}{(l_t + \alpha\lambda_i - 2\alpha(t-1)d)^2} - \frac{1}{\alpha} \sum_{i=1}^{2n} \frac{l_t + \alpha\lambda_i - 2\alpha(t-1)d}{(l_t + \alpha\lambda_i - 2\alpha(t-1)d)^2} \\ &= \frac{l_t}{\alpha} - 2(t-1)d - \frac{\text{tr}(A_t^{1/2})}{\alpha} \\ &\geq -\lambda_{\min}(Z_{t-1}) - \frac{\text{tr}(A_t^{1/2})}{\alpha} \\ &\geq -\frac{\sqrt{n}}{\alpha}, \end{aligned}$$

where the third equality is because $\text{tr}(A_t) = 1$ and $(l_t + \alpha\lambda_i - 2\alpha(t-1)d)^{-1}$ is the i -th eigenvalue of $A_t^{1/2}$, the first inequality is by $A_t \succ 0$ which implies that $l_t > 2\alpha(t-1)d - \lambda_{\min}(Z_{t-1})$, and the last inequality is by $\lambda_{\min}(Z_{t-1}) \leq 0$ and $\text{tr}(A_t^{1/2}) \leq \sqrt{n}$ from (3.1).

Since $|E \setminus S_{t-1}| = m - t + 1$, an averaging argument shows that there exists an edge $e \in E \setminus S_{t-1}$ such that

$$\left\langle A_t, \begin{pmatrix} L_G - mL_e & \\ & L_G^+ - mL_e^+ \end{pmatrix} \right\rangle \geq -\frac{\sqrt{n}}{\alpha(m-t+1)} \geq -\frac{2\sqrt{n}}{\alpha m},$$

where the last inequality is because $m-t+1 \geq m-\tau+1 \geq m/2$ by our assumption $\tau = n/\epsilon^2 \leq m/2$. Finally, when $\alpha = q\epsilon/\sqrt{dm}$ for some constant $q > 0$,

$$\frac{2\sqrt{n}}{\alpha m} = O\left(\frac{1}{\epsilon} \sqrt{\frac{dn}{m}}\right) \leq O(\sqrt{d}) \leq O(\epsilon d),$$

where the first inequality is by our assumption $\tau = n/\epsilon^2 \leq m/2$, and the second inequality follows from $dn \geq m \geq 2n/\epsilon^2$ which implies $\epsilon \geq \sqrt{2/d}$. \square

Given Lemma 4.21, the rest of the proof is almost the same as that in [14], but we include the proofs for completeness. The following lemma bounds the width term, which is essentially the same as Claim 1 in [14] with minor modification.

Lemma 4.22. *If $\alpha = q\epsilon/\sqrt{dm}$ for a sufficiently small constant $q > 0$, then*

$$\alpha \left\| A_t^{\frac{1}{4}} F_t A_t^{\frac{1}{4}} \right\|_{\text{op}} \leq \min \left\{ \frac{1}{4}, \epsilon \right\}.$$

Proof. Since the feedback matrices F_t have a block diagonal structure, by the closed-form solution of the action matrix in (2.1), A_t also has the same block diagonal structure

$$A_t = \begin{pmatrix} B_t & \\ & C_t \end{pmatrix}, \quad \text{where } 0 \preceq B_t, C_t \preceq I_n.$$

Therefore,

$$\left\| A_t^{\frac{1}{4}} F_t A_t^{\frac{1}{4}} \right\|_{\text{op}} = \max \left\{ \left\| B_t^{\frac{1}{4}} (L_G - mL_{e_t} - 2dI_n) B_t^{\frac{1}{4}} \right\|_{\text{op}}, \left\| C_t^{\frac{1}{4}} (L_G^+ - mL_{e_t}^+ - 2dI_n) C_t^{\frac{1}{4}} \right\|_{\text{op}} \right\}.$$

We will just bound the first term, as the second term can be bounded the same way. By triangle inequality and the facts that $0 \preceq B_t \preceq I_n$ and $0 \preceq L_G \preceq 2dI$, it follows that

$$\left\| B_t^{\frac{1}{4}} (L_G - mL_{e_t} - 2dI_n) B_t^{\frac{1}{4}} \right\|_{\text{op}} \leq m \left\| B_t^{\frac{1}{4}} L_{e_t} B_t^{\frac{1}{4}} \right\|_{\text{op}} + 2d,$$

By the choice of edge e_t as guaranteed by Lemma 4.21,

$$-\frac{2\sqrt{n}}{\alpha m} \leq \left\langle A_t, \begin{pmatrix} L_G - mL_{e_t} & \\ & L_G^+ - mL_{e_t}^+ \end{pmatrix} \right\rangle = \left\langle A_t, \begin{pmatrix} L_G & \\ & L_G^+ \end{pmatrix} \right\rangle - m\langle L_{e_t}, B_t \rangle - m\langle L_{e_t}^+, C_t \rangle.$$

Since $\text{tr}(A_t) = 1$, $L_G, L_G^+ \preceq 2dI_n$ and $\langle L_{e_t}^+, C_t \rangle \geq 0$, the above inequality implies that

$$m\langle L_{e_t}, B_t \rangle \leq 2d + \frac{2\sqrt{n}}{\alpha m}.$$

Let $B_t = \sum_{i=1}^n \lambda_i y_i y_i^T$ be the eigendecomposition of B_t , and let $w = \sqrt{m} \cdot b_{e_t}$ so that $ww^T = mL_{e_t}$ and $\|w\|_2 = \sqrt{2m}$. Then

$$\begin{aligned} m \left\| B_t^{\frac{1}{4}} L_{e_t} B_t^{\frac{1}{4}} \right\|_{\text{op}} &= w^T B_t^{\frac{1}{2}} w = \sum_{i=1}^n \sqrt{\lambda_i} \cdot \langle w, y_i \rangle^2 \\ &\leq \sqrt{\sum_{i=1}^n \langle w, y_i \rangle^2} \cdot \sqrt{\sum_{i=1}^n \lambda_i \cdot \langle w, y_i \rangle^2} \\ &= \|w\|_2 \cdot \sqrt{w^T B_t w} \\ &\leq \sqrt{2m} \cdot \sqrt{2d + \frac{2\sqrt{n}}{\alpha m}} \\ &= 2\sqrt{dm + \frac{\sqrt{n}}{\alpha}}, \end{aligned}$$

where the first inequality is by Cauchy-Schwartz, and the second inequality follows from $w^T B_t w = m\langle B_t, L_{e_t} \rangle$ and the above upper bound on $m\langle B_t, L_{e_t} \rangle$. The same arguments gives the same upper bound on $m\|C_t^{1/4} L_{e_t}^+ C_t^{1/4}\|$. Therefore, for $\alpha = q\epsilon/\sqrt{dm}$,

$$\begin{aligned} \alpha \left\| A_t^{\frac{1}{4}} F_t A_t^{\frac{1}{4}} \right\|_{\text{op}} &\leq 2\alpha d + 2\sqrt{\alpha^2 dm + \alpha\sqrt{n}} \\ &= 2q\epsilon\sqrt{\frac{d}{m}} + 2\sqrt{q^2\epsilon^2 + q\epsilon\sqrt{\frac{n}{dm}}} \\ &\leq O(\sqrt{q}\epsilon), \end{aligned}$$

where the last inequality follows from $d/m \leq n/m \leq \epsilon^2/2$ by the assumption that $m \geq 2n/\epsilon^2$. Then the lemma follows when q is sufficiently small. \square

We are ready to prove Theorem 4.20, whose proof is essentially the same as in [14].

Proof of Theorem 4.20. By Lemma 4.22, when $\alpha = q\epsilon/\sqrt{dm}$ for a small enough constant $q > 0$, then $\alpha\|A_t^{1/4} F_t A_t^{1/4}\| \leq \frac{1}{4}$ for any t which also implies that $\alpha A_t^{1/4} F_t A_t^{1/4} \succcurlyeq -\frac{1}{4}I$ for any t . Therefore, we can apply Theorem 2.3 and get

$$R_\tau = \sum_{t=1}^{\tau} \langle A_t, F_t \rangle - \lambda_{\min} \left(\sum_{t=1}^{\tau} F_t \right) \leq O(\alpha) \sum_{t=1}^{\tau} \langle A_t, |F_t| \rangle \cdot \left\| A_t^{\frac{1}{4}} F_t A_t^{\frac{1}{4}} \right\|_{\text{op}} + \frac{2\sqrt{n}}{\alpha}. \quad (4.3)$$

By Lemma 4.21 and Lemma 4.22 and the fact that $F_t \preccurlyeq 0$, it holds that

$$\langle A_t, F_t \rangle \geq -(2 + O(\epsilon))d, \quad \langle A_t, |F_t| \rangle \leq (2 + O(\epsilon))d \quad \text{and} \quad \alpha\|A_t^{\frac{1}{4}} F_t A_t^{\frac{1}{4}}\| \leq \min \left\{ \frac{1}{4}, \epsilon \right\}.$$

Together with $\alpha = \Theta(\epsilon/\sqrt{dm}) \geq \Omega(\epsilon/d\sqrt{n}) = \Omega(\sqrt{n}/(\epsilon d\tau))$, the regret minimization bound (4.3) implies that

$$-(2 + O(\epsilon))\tau d - \lambda_{\min} \left(\left(\tau L_G - m \sum_{t=1}^{\tau} L_{e_t} \quad \tau L_G^+ - m \sum_{t=1}^{\tau} L_{e_t}^+ \right) - 2\tau d I \right) \leq O(\epsilon)\tau d.$$

Let $\tilde{m} := \tau$ and $L_{\tilde{G}} = \sum_{t=1}^{\tau} L_{e_t}$. From the first block, we have

$$\tau L_G - \sum_{t=1}^{\tau} m L_{e_t} \succcurlyeq -O(\epsilon)\tau d I_n \quad \implies \quad \frac{m}{\tilde{m}} L_{\tilde{G}} - L_G \preccurlyeq O(\epsilon) d I_n.$$

From the second block, we have

$$\tau L_G^+ - \sum_{t=1}^{\tau} m L_{e_t}^+ \succcurlyeq -O(\epsilon)\tau d I_n \quad \implies \quad \frac{m}{\tilde{m}} L_{\tilde{G}} - L_G \succcurlyeq 2\frac{m}{\tilde{m}} D_{\tilde{G}} - 2D_G - O(\epsilon) d I_n,$$

where we used that $L_G^+ = 2D_G - L_G$. \square

Concluding Remarks

We propose a spectral approach to design approximation algorithms for network design problems. We show that the techniques developed in spectral graph theory and discrepancy theory can be used to significantly extend the scope of network design problems that can be solved. We believe that this connection will bring new techniques and stronger results for network design, and will also introduce new formulations and interesting questions to spectral graph theory and discrepancy theory. It also gives extra motivation to design a constructive algorithm for the method of interlacing polynomials, as this will lead to very strong approximation algorithms for network design. We leave it as an open question to improve the spectral approach to fully recover Jain’s result.

Acknowledgement

We thank Akshay Ramachandran for many useful discussions, for bringing [45] to our attention, and for his suggestion of the signing trick in Lemma 3.15 to proving Theorem 1.9. We also thank Shayan Oveis Gharan for comments that improve the presentation of the paper.

References

- [1] Ajit Agrawal, Philip Klein, and R. Ravi. *When trees collide: An approximation algorithm for the generalized steiner problem on networks*. SIAM Journal on Computing, 24(3):440–456, 1995. 1, 2
- [2] Vedat Levi Alev, Nima Anari, Lap Chi Lau, and Shayan Oveis Gharan. *Graph Clustering using Effective Resistance*. In Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS), 94:41, 2018. 41
- [3] Zeyuan Allen-Zhu, Zhenyu Liao, Lorenzo Orecchia. *Spectral sparsification and regret minimization beyond matrix multiplicative updates*. In Proceedings of the forty-seventh annual ACM Symposium on Theory of Computing (STOC), 237–245, 2015. 1, 5, 7, 13, 19, 49
- [4] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. *Near-optimal design of experiments via regret minimization*. In Proceedings of the 34th International Conference on Machine Learning (ICML), 70:126–135, 2017. 14, 43, 44, 45, 51
- [5] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. *Near-optimal discrete optimization for experimental design: A regret minimization approach*. Mathematical Programming, Jan 10:1–40. Springer, 2020. 5, 7, 8, 9, 13, 14, 15, 16, 19, 20, 43, 44, 45, 46, 48, 51
- [6] Nima Anari and Shayan Oveis Gharan. *Effective-resistance-reducing flows, spectrally thin trees, and asymmetric TSP*. In Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS), pages 20–39. IEEE, 2015. 1, 3, 5
- [7] David G. Anderson, Ming Gu, and Christopher Melgaard. *An efficient algorithm for unweighted spectral graph sparsification*. arXiv preprint arXiv:1410.4273, 2014. 5

- [8] Dana Angluin. *Queries and concept learning*. Machine learning, 2(4):319–342, 1988. 9, 43
- [9] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. *An $O(\log n / \log \log n)$ -approximation algorithm for the asymmetric traveling salesman problem*. In Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms (SODA), pages 379–389. SIAM, 2010. 1
- [10] Nikhil Bansal. *On a generalization of iterated and randomized rounding*. In Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC), pages 1125–1135, 2019. 1, 4, 9, 35, 39, 42
- [11] Nikhil Bansal, Daniel Dadush, Shashwat Garg, and Shachar Lovett. *The Gram–Schmidt Walk: A Cure for the Banaszczyk Blues*. Theory OF Computing, 15(21):1–27, 2019. 15
- [12] Nikhil Bansal and Shashwat Garg. *Algorithmic discrepancy beyond partial coloring*. In Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC), pages 914–926, 2017. 15
- [13] Nikhil Bansal, Rohit Khandekar, Jochen Könnemann, Viswanath Nagarajan, and Britta Peis. *On generalizations of network design problems with degree bounds*. Mathematical Programming, 141(1-2):479–506. Springer, 2013. 2
- [14] Nikhil Bansal, Ola Svensson, and Luca Trevisan. *New Notions and Constructions of Sparsification for Graphs and Hypergraphs*. In Proceedings of the 60th Annual Symposium on Foundations of Computer Science (FOCS), pages 910–928, 2019. 5, 7, 10, 11, 15, 49, 51, 52, 53, 55
- [15] Joshua Batson, Daniel A. Spielman, and Nikhil Srivastava. *Twice-ramanujan sparsifiers*. SIAM Journal on Computing, 41(6):1704–1721, 2012. 1, 5, 9, 49
- [16] András A. Benczúr and David R. Karger. *Approximating st Minimum Cuts in $\tilde{O}(n^2)$ Time*. In Proceedings of the twenty-eighth annual ACM symposium on Theory of computing (STOC), pages 47–55. ACM, 1996. 1
- [17] Vittorio Bilo, Vineet Goyal, Ramamoorthi Ravi, and Mohit Singh. *On the crossing spanning tree problem*. In Proceedings of 7th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX), pages 51-60, 2004. 2
- [18] Stephen Boyd, Persi Diaconis, and Lin Xiao. *Fastest mixing markov chain on a graph*. SIAM review, 46(4):667–689, 2004. 3
- [19] Jaroslaw Byrka, Fabrizio Grandoni, Thomas Rothvoss, and Laura Sanita. *Steiner tree approximation via iterative randomized rounding*. Journal of the ACM, 60(1):6, 2013. 1, 2
- [20] Pak Hay Chan, Lap Chi Lau, Aaron Schild, Sam Chiu-wai Wong, and Hong Zhou. *Network design for s-t effective resistance*. arXiv preprint arXiv:1904.03219, 2019. 2, 10, 48

- [21] Ashok K. Chandra, Prabhakar Raghavan, Walter L. Ruzzo, Roman Smolensky, and Prasoon Tiwari. *The electrical resistance of a graph captures its commute and cover times*. Computational Complexity, 6(4):312–340, 1996. 2, 41
- [22] Joseph Cheriyan and László A. Végh. *Approximating minimum-cost k -node connected subgraphs via independence-free graphs*. SIAM Journal on Computing, 43(4):1342–1362. SIAM, 2014. 2
- [23] Joseph Cheriyan, Santosh Vempala, and Adrian Vetta. *Network design via iterative rounding of setpair relaxations*. Combinatorica, 26(3):255–275, 2006. 1, 2
- [24] Paul Christiano, Jonathan A. Kelner, Aleksander Madry, Daniel A. Spielman, and Shang-Hua Teng. *Electrical flows, laplacian systems, and faster approximation of maximum flow in undirected graphs*. In Proceedings of the forty-third annual ACM symposium on Theory of computing (STOC), pages 273–282. ACM, 2011. 1, 2
- [25] Ali Çivril and Malik Magdon-Ismail. *On selecting a maximum volume sub-matrix of a matrix and related problems*. Theoretical Computer Science, 410(47-49):4801–4811, Elsevier, 2009. 43
- [26] Jian Ding, James R. Lee, and Yuval Peres. *Cover times, blanket times, and majorizing measures*. In Proceedings of the forty-third annual ACM symposium on Theory of computing (STOC), pages 61–70. ACM, 2011. 2
- [27] Yevgeniy Dodis and Sanjeev Khanna. *Design networks with bounded pairwise distance*. In Proceedings of the thirty-first annual ACM symposium on Theory of computing (STOC), pages 750–759. ACM, 1999. 2
- [28] Alina Ene and Ali Vakilian. *Improved approximation algorithms for degree-bounded network design problems with node connectivity requirements*. In Proceedings of the forty-sixth annual ACM symposium on Theory of computing (STOC), pages 754–763. ACM, 2014. 1, 2
- [29] Shaun M. Fallat, Steve Kirkland, Sukanta Pati. *On graphs with algebraic connectivity equal to minimum edge density*. Linear algebra and its applications, 373:31-50, 2003. 46
- [30] Lisa Fleischer, Kamal Jain, and David P. Williamson. *An iterative rounding 2-approximation algorithm for the element connectivity problem*. In Proceedings of 42nd IEEE Symposium on Foundations of Computer Science (FOCS), pages 339–347. IEEE, 2001. 1, 2
- [31] David A. Freedman. *On tail probabilities for martingales*. Annals of Probability, 3(1):100–118, 1975. 15
- [32] Takuro Fukunaga, Zeev Nutov, and R. Ravi. *Iterative rounding approximation algorithms for degree-bounded node-connectivity network design*. SIAM Journal on Computing, 44(5):1202–1229, 2015. 1, 2
- [33] Martin Furer and Balaji Raghavachari. *Approximating the minimum-degree Steiner tree to within one of optimal*. Journal of Algorithms, 17(3):409–423, 1994. 5

- [34] Harold N. Gabow. *On the ℓ_∞ -norm of extreme points for crossing supermodular directed network lps.* In International Conference on Integer Programming and Combinatorial Optimization (IPCO), pages 392–406. Springer, 2005. 1, 2
- [35] Harold N. Gabow, Michel X. Goemans, Éva Tardos, and David P. Williamson. *Approximating the smallest k -edge connected spanning subgraph by lp -rounding.* Networks, 53(4):345–357, 2009. 1, 2, 5
- [36] Naveen Garg, Goran Konjevod, and R. Ravi. *A polylogarithmic approximation algorithm for the group Steiner tree problem.* Journal of Algorithms, 37(1):66–84, 2000. 1
- [37] Arpita Ghosh and Stephen Boyd. *Growing well-connected graphs.* In Proceedings of the 45th IEEE Conference on Decision and Control (CDC), pages 6605–6611. IEEE, 2006. 3, 10, 12, 36, 46
- [38] Arpita Ghosh, Stephen Boyd, and Amin Saberi. *Minimizing effective resistance of a graph.* SIAM review, 50(1):37–66, 2008. 3, 10, 13, 36, 47
- [39] Michel X. Goemans. *Minimum bounded degree spanning trees.* In Proceedings of the 47th annual IEEE symposium on Foundations of computer science (FOCS), pages 273–282, 2006. 2, 5
- [40] Michel X. Goemans, Andrew V. Goldberg, Serge A. Plotkin, David B. Shmoys, Eva Tardos, and David P. Williamson. *Improved Approximation Algorithms for Network Design Problems.* In Proceedings of the 50th annual ACM-SIAM symposium on Discrete algorithms (SODA), pages 223–232. SIAM, 1994. 1, 2
- [41] Michel X. Goemans and David P. Williamson. *A general approximation technique for constrained forest problems.* SIAM Journal on Computing, 24(2):296–317, 1995. 1, 2
- [42] Fabrizio Grandoni, Bundit Laekhanukit, and Shi Li. *$O(\log^2 k / \log \log k)$ -approximation algorithm for directed Steiner tree: a tight quasi-polynomial-time algorithm.* In Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC), 253–264. ACM, 2019. 1
- [43] Kamal Jain. *A factor 2 approximation algorithm for the generalized steiner network problem.* Combinatorica, 21(1):39–60, 2001. 1, 2, 5, 36
- [44] Alexandra Kolla, Yury Makarychev, Amin Saberi, and Shang-Hua Teng. *Subgraph sparsification and nearly optimal ultrasparsifiers.* In Proceedings of the 42nd ACM symposium on Theory of computing (STOC), pages 57–66. ACM, 2010. 10, 46
- [45] Rasmus Kyng, Kyle Luh, Zhao Song. *Four deviations suffice for rank 1 matrices.* arXiv preprint arXiv:1901.06731, 2019. 5, 6, 32, 51, 56
- [46] Lap Chi Lau, Joseph Naor, Mohammad R. Salavatipour, and Mohit Singh. *Survivable network design with degree or order constraints.* SIAM Journal on Computing, 39(3):1062–1087, 2009. 1, 2

- [47] Lap Chi Lau, Ramamoorthi Ravi, and Mohit Singh. *Iterative methods in combinatorial optimization*. Cambridge University Press, 2011. [1](#)
- [48] Lap Chi Lau and Mohit Singh. *Additive approximation for bounded degree survivable network design*. SIAM Journal on Computing, 42(6):2217–2242. SIAM, 2013. [2](#)
- [49] Lap Chi Lau and Hong Zhou. *A unified algorithm for degree bounded survivable network design*. Mathematical Programming, 154(1-2):515–532, 2015. [1](#), [2](#)
- [50] Yin Tat Lee and He Sun. *An SDP-based algorithm for linear-sized spectral sparsification*. In Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC), pages 678–687. ACM, 2017. [5](#)
- [51] Yin Tat Lee and He Sun. *Constructing linear-sized spectral sparsification in almost-linear time*. SIAM Journal on Computing, 47(6):2315–2336, 2018. [5](#), [9](#)
- [52] André Linhares and Chaitanya Swamy. *Approximating min-cost chain-constrained spanning trees: a reduction from weighted to unweighted problems*. Mathematical Programming, 172(1-2):17–34. Springer, 2018 [2](#)
- [53] Anand Louis and Nisheeth K. Vishnoi. *Improved algorithm for degree bounded survivable network design problem*. In Proceedings of Scandinavian Workshop on Algorithm Theory (SWAT), pages 408–419, 2010. [2](#)
- [54] Vivek Madan, Mohit Singh, Uthaipon Tantipongpipat, and Weijun Xie. *Combinatorial Algorithms for Optimal Design*. In Proceedings of Conference on Learning Theory (COLT), pages 2210–2258, 2019. [44](#)
- [55] Aleksander Madry, Damian Straszak, and Jakub Tarnawski. *Fast generation of random spanning trees and the effective resistance metric*. In Proceedings of the 26th annual ACM-SIAM symposium on Discrete algorithms (SODA), pages 2019–2036. SIAM, 2015. [3](#)
- [56] Adam W. Marcus, Daniel A. Spielman, Nikhil Srivastava. *Interlacing families I: Bipartite Ramanujan graphs of all degrees*. Annals of Mathematics, 307–325, 2015. [5](#), [51](#)
- [57] Adam W. Marcus, Daniel A. Spielman, Nikhil Srivastava. *Interlacing families II: Mixed characteristic polynomials and the Kadison Singer problem*. Annals of Mathematics, 327–350, 2015. [5](#), [51](#)
- [58] Peter Matthews. *Covering problems for brownian motion on spheres*. The Annals of Probability, pages 189–199, 1988. [2](#)
- [59] Aleksandar Nikolov, Mohit Singh, and Uthaipon Tao Tantipongpipat. *Proportional volume sampling and approximation algorithms for A-optimal design*. In Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 1369–1386. SIAM, 2019. [10](#), [33](#), [43](#), [44](#), [48](#)
- [60] Neil Olver and Rico Zenklusen. *Chain-constrained spanning trees*. Mathematical Programming, 167(2):293–314. Springer, 2018. [2](#)

- [61] Friedrich Pukelsheim. *Optimal design of experiments*. SIAM, 2016. 9, 43
- [62] Balaji Raghavachari. *Algorithms for finding low degree structures*. Approximation algorithms for NP-hard problems, pages 266–295, 1996. 2
- [63] Ramamoorthi Ravi, Madhav V. Marathe, S. S. Ravi, Daniel J. Rosenkrantz, and Harry B. Hunt III. *Approximation Algorithms for Degree-Constrained Minimum-Cost Network-Design Problems*. Algorithmica, 31(1):58–78, 2001. 2
- [64] Aaron Schild. *An almost-linear time algorithm for uniform random spanning tree generation*. In Proceedings of the 50th annual ACM symposium on Theory of computing (STOC), pages 214–227. ACM, 2018. 1, 3
- [65] Marcel K. Silva, Nicholas J. A. Harvey, and Cristiane M. Sato. *Sparse Sums of Positive Semidefinite Matrices*. ACM Transactions on Algorithms, 12(1):9, 2016 5, 31
- [66] Mohit Singh and Lap Chi Lau. *Approximating minimum bounded degree spanning tress to within one of optimal*. Journal of the ACM, 62(1):1, 2015. 2
- [67] Mohit Singh and Weijun Xie. *Approximate Positively Correlated Distributions and Approximation Algorithms for D-optimal Design*. In Proceedings of Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 2240–2255. ACM-SIAM, 2018. 44
- [68] Daniel A. Spielman and Nikhil Srivastava. *Graph sparsification by effective resistances*. SIAM Journal on Computing, 40(6):1913–1926, 2011. 1, 2, 5
- [69] Daniel A. Spielman and Shang-Hua Teng. *Spectral sparsification of graphs*. SIAM Journal on Computing, 40(4):981–1025. SIAM, 2011. 1
- [70] Aravind Srinivasan. *Distributions on level-sets with applications to approximation algorithms*. In Proceedings 42nd IEEE Symposium on Foundations of Computer Science (FOCS), pages 588–597. IEEE, 2001. 1
- [71] Joel A. Tropp. *Freedman’s inequality for matrix martingales*. Electronic Communications in Probability, 16:262–270, 2011. 15