Spectral Hypergraph Sparsifiers of Nearly Linear Size*

Michael Kapralov†
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
michael.kapralov@epfl.ch

Robert Krauthgamer‡
Weizmann Institute of Science
Rehovot, Israel
robert.krauthgamer@weizmann.ac.il

Jakab Tardos§
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
jakab.tardos@epfl.ch

Yuichi Yoshida¶
National Institute of Informatics
Tokyo, Japan
yyoshida@nii.ac.jp

***Abstract*—Graph sparsification has been studied extensively over the past two decades, culminating in spectral sparsifiers of optimal size (up to constant factors). Spectral hypergraph sparsification is a natural analogue of this problem, for which optimal bounds on the sparsifier size are not known, mainly because the hypergraph Laplacian is non-linear, and thus lacks the linear-algebraic structure and tools that have been so effective for graphs.**

Our main contribution is the first algorithm for constructing ϵ-spectral sparsifiers for hypergraphs with $O^*(n)$ hyperedges, where O^* suppresses $(\epsilon^{-1}\log n)^{O(1)}$ factors. This bound is independent of the rank r (maximum cardinality of a hyperedge), and is essentially best possible due to a recent bit complexity lower bound of $\Omega(nr)$ for hypergraph sparsification.

This result is obtained by introducing two new tools. First, we give a new proof of spectral concentration bounds for sparsifiers of graphs; it avoids linear-algebraic methods, replacing e.g. the usual application of the matrix Bernstein inequality and therefore applies to the (non-linear) hypergraph setting. To achieve the result, we design a new sequence of hypergraph-dependent ϵ-nets on the unit sphere in $\mathbb{R}^n$. Second, we extend the weight-assignment technique of Chen, Khanna and Nagda [FOCS'20] to the spectral sparsification setting. Surprisingly, the number of spanning trees after the weight assignment can serve as a potential function guiding the reweighting process in the spectral setting.

***Keywords*-hypergraphs, sparsification, spectral**

I. Introduction

We study spectral sparsification of hypergraphs, where the goal is to reduce the size of a hypergraph while preserving its energy. Given a hypergraph $H = (V, E, w)$ with a weight function $w : E \to \mathbb{R}_+$ over its hyperedges, the *energy* of $x \in \mathbb{R}^V$ (called a potential vector) is defined as

$$Q_H(x) := \sum_{e\in E} w(e) \cdot \max_{u,v\in e} (x_u - x_v)^2.$$

The problem of minimizing $Q_H(x)$ over $x \in \mathbb{R}^V$ subject to certain constraints appears in many problems involving hypergraphs, including clustering [1], semi-supervised learning [2, 3, 4] and link prediction [5], from which we can see the relevance of $Q_H(x)$ in application domains. Note that when $x \in \mathbb{R}^V$ is a characteristic vector $\mathbb{1}_S \in \{0,1\}^V$ of a vertex subset $S \subset V$, the energy $Q_H(\mathbb{1}_S)$ coincides with the total weight of hyperedges cut by S, where we say that a hyperedge $e \in E$ is *cut* by S if $e \cap S \neq \emptyset$ and $e \cap (V \setminus S) \neq \emptyset$.

Since the number of hyperedges in a hypergraph of n vertices can be $\Omega(2^n)$, it is desirable to reduce the number of hyperedges in the hypergraph while (approximately) preserving the value of $Q_H(x)$ for every $x \in \mathbb{R}^V$, because this lets us speed up any algorithm involving Q_H and reduce its memory usage by running it on the smaller hypergraph instead of H itself. Soma and Yoshida [6] formalized this concept as spectral sparsification for hypergraphs – a natural generalization of the corresponding concept introduced by the celebrated work of [7] for graphs. Specifically, for $0 < \epsilon < 1$, we say that a hypergraph $\widetilde{H}$ is an *ϵ-spectral-sparsifier* of a hypergraph H if $\widetilde{H}$ is a reweighted subgraph of H such that

$$\forall x \in \mathbb{R}^V, \qquad Q_{\widetilde{H}}(x) \in (1 \pm \epsilon) Q_H(x).^{1} \qquad (1)$$

[1] $a \in (1 \pm \epsilon)b$ is a shorthand for $(1-\epsilon)b \le a \le (1+\epsilon)b$.

* A full version of this paper can be found at arXiv:2106.02353

† Supported in part by ERC Starting Grant 759471.
‡ Work partially supported by ONR Award N00014-18-1-2364, the Israel Science Foundation grant #1086/18, and a Minerva Foundation grant.
§ Supported by ERC Starting Grant 759471.
¶ Supported in part by JSPS KAKENHI Grant Number 18H05291 and 20H05965.

We note that when H is an ordinary graph, this definition matches that for graphs [7]. Soma and Yoshida [6] showed that every hypergraph H admits an ϵ-spectral-sparsifier with $\widetilde{O}(n^3/\epsilon^2)$ hyperedges,[2] and gave a polynomial-time algorithm for constructing such sparsifiers. Since then the number of hyperedges needed has been reduced to $\widetilde{O}(nr^3/\epsilon^2)$ [8], and recently to $\widetilde{O}(nr/\epsilon^{O(1)})$ [9], where r is the maximum size of a hyperedge in the input hypergraph H (called the *rank* of H).

The natural question whether every hypergraph admits a spectral sparsifier with $\widetilde{O}(n)$ hyperedges (for fixed ϵ) has proved to be challenging. On the one hand, it is well-known that a hypergraph is a strictly richer object than an ordinary graph (hyperedges cannot be "simulated" by edges, even approximately), and in all previous results and techniques, this extra complication introduced an extra factor of at least r. On the other hand, an exciting recent result [10] has achieved sparsifiers with $\widetilde{O}(n)$ hyperedges, if one is only interested in preserving the *hypergraph cut function*, i.e., satisfying (1) only for all characteristic vectors $x = \mathbb{1}_S$ where $S \subseteq V$. Nevertheless, the spectral version of this question has remained open, primarily due to the non-linearity of the hypergraph Laplacian and the lack of linear-algebraic tools that have been so effective for graphs.

We settle this question by showing that a nearly linear number of hyperedges suffices.

Theorem 1. *For every hypergraph with n vertices and every $1/n \leq \epsilon \leq 1/2$, there exists an ϵ-spectral-sparsifier with $O(n\epsilon^{-4}\log^3 n)$ hyperedges. Moreover, one can construct such a sparsifier in time $\widetilde{O}(mr + \mathrm{poly}(n))$, where m is the number of hyperedges and r is the maximum size of a hyperedge in H.*

We note that the bit complexity of our sparsifier is tight up to a polylogarithmic factor for a large range of r due to the lower bound of [9]. The proof of Theorem 1 can be found in the full version of the paper.

A. Additional Related Work

Recall that we call $\widetilde{H} = (V, \widetilde{E}, \widetilde{w})$ an *ϵ-cut sparsifier* of $H = (V, E, w)$ if every cut weight is preserved to within a factor of $1 \pm \epsilon$. This definition matches the one for ordinary graphs introduced by Benczúr and Karger [11], who showed that every graph has an ϵ-cut-sparsifier with $O(n \log n/\epsilon^2)$ edges, where n is the number of vertices. For hypergraphs, Kogan and Krauthgamer [12] gave the first construction of non-trivial cut sparsifiers, which uses $O(n(r + \log n)/\epsilon^2)$ hyperedges, where r is the maximum size of a hyperedge. They also mentioned that the results of Newman and Rabinovich [13] implicitly give an ϵ-cut sparsifier with $O(n^2/\epsilon^2)$ hyperedges. Chen, Khanna, and Nagda [10] improved this bound to $O(n \log n/\epsilon^2)$, which is almost tight because one needs $\Omega(n/\epsilon^2)$ edges even for ordinary graphs [14, 15].

Spielman and Teng [7] introduced the notion of a spectral sparsifier for ordinary graphs and showed that every graph on n vertices admits an ϵ-spectral sparsifier with $O(n \log^{O(1)} n/\epsilon^2)$ edges. This bound was later improved to $O(n/\epsilon^2)$ [16], which is tight [14, 15]. The literature on graph sparsification is too vast to cover here, including [7, 17, 16, 18, 19, 20] and many other constructions, and we refer the reader to the surveys [21, 22].

For an ordinary graph $G = (V, E, w)$, the *Laplacian* of G is the matrix $L_G = D_G - A_G$, where $D_G \in \mathbb{R}^{V \times V}$ is the diagonal (weighted) degree matrix and $A_G \in \mathbb{R}^{V \times V}$ is the adjacency matrix of G. Then, the energy Q_G, defined in (1), can be written also as

$$Q_G(x) = x^\top L_G x.$$

For a hypergraph $H = (V, E, w)$, it is known that we can define a (multi-valued) Laplacian operator $L_H : \mathbb{R}^V \to 2^{\mathbb{R}^V}$, so that

$$Q_H(x) = x^\top y$$

for every $x \in \mathbb{R}^V$ and $y \in L_H(x)$ [23, 24, 25] (hence we can write $Q_H(x)$ also as $x^\top L_H(x)$ without ambiguity). Although the Laplacian operator L_H is no longer a linear operator, its mathematical property has been actively investigated [26, 27, 28] through the theory of monotone operators and evolution equations [29, 30].

Yoshida [31] proposed a Laplacian operator for directed graphs and used it to study structures of real-world networks. The Laplacian operators for graphs, hypergraphs, and directed graphs mentioned above were later unified and generalized as Laplacian operator for submodular transformations/submodular hypergraphs [32, 25].

II. Preliminaries

In this paper, we deal with spectral sparsification of hypergraphs. For the sake of generality, we consider weighted hypergraphs denoted $H = (V, E, w)$, where V is the vertex set of size n, E is the hyperedge set of size m, and $w : E \to \mathbb{R}_+$ is the set of hyperedge weights. We will also, however, deal with ordinary graphs, that is graphs where each edge contains two vertices exactly. In order to distinguish

[2] Throughout, $\widetilde{O}(\cdot)$ suppresses a factor of $\log^{O(1)} n$.

clearly between graphs and hypergraphs, we will typically denote graphs as $G = (V, F, z)$, where V is the vertex set, F is the edge set, and $z : F \to \mathbb{R}_+$ is the set of edge weights. In general we will use f and g to denote ordinary edges, while reserving e to denote hyperedges.

For simplicity *all graphs and hypergraphs we consider in this paper will be connected.*

A. Spectral Graph Theory

Definition II.1. *The Laplacian of a weighted graph $G = (V, F, z)$ is defined as the matrix $L_G \in \mathbb{R}^{V \times V}$ such that*

$$(L_G)_{uv} = \begin{cases} d(u) & \text{if } u = v, \\ -z(u,v) & \text{if } (u,v) \in F, \\ 0 & \text{otherwise.} \end{cases}$$

Here $d(u)$ denotes the weighted degree of u, that is the sum of all weights of incident edges. Thus L_G is a positive semidefinite matrix, and its quadratic form can be written as

$$x^\top L_G x = \sum_{(u,v)\in F} z(u,v) \cdot (x_u - x_v)^2.$$

The spectral sparsifier of G is defined as a reweighted subgraph which closely approximates the quadratic form of the Laplacian on every possible vector.

Definition II.2. *Let $G = (V, F, z)$ be a weighted ordinary graph. Let $\widetilde{G} = (V, \widetilde{F}, \widetilde{z})$ be a reweighted subgraph of G, defined by $\widetilde{z} : F \to \mathbb{R}_+$, where $\widetilde{F} = \{f \in F \mid \widetilde{z}(f) > 0\}$. For $\epsilon > 0$, $\widetilde{G}$ is an ϵ-spectral sparsifier of G if for every $x \in \mathbb{R}^V$*

$$x^\top L_{\widetilde{G}} x \in (1 \pm \epsilon) \cdot x^\top L_G x.$$

The quadratic form of the graph Laplacian from Definition II.1 can be generalized to hypergraphs. Although this generalization is highly non-linear, we still refer to it as the "quadratic form" of the hypergraph.

Definition II.3. *The quadratic form (or sometimes* energy*) of a hypergraph $H = (V, E, w)$ is defined on the input vector $x \in \mathbb{R}^V$ as*

$$Q_H(x) = \sum_{e \in E} w(e) \cdot \max_{u,v \in e} (x_u - x_v)^2.$$

Consequently, we may also define the concept of spectral sparsification in hypergraphs, analogously to Definition II.2:

Definition II.4. *Let $H = (V, E, w)$ be a weighted hypergraph. Let $\widetilde{H} = (V, \widetilde{E}, \widetilde{w})$ be a reweighted subgraph of H, defined by $\widetilde{w} : E \to \mathbb{R}_+$, where $\widetilde{E} = \{e \in E \mid \widetilde{w}(e) > 0\}$. For $\epsilon > 0$, $\widetilde{H}$ is an ϵ-spectral sparsifier of H if for every $x \in \mathbb{R}^V$*

$$Q_{\widetilde{H}}(x) \in (1 \pm \epsilon) \cdot Q_H(x).$$

B. Effective Resistance

Definition II.5. *Let $G = (V, F, z)$ be a weighted ordinary graph. The effective resistance of a pair of vertices (u, v) is defined as*

$$R_G(u,v) = (\chi_u - \chi_v)^\top L_G^+ (\chi_u - \chi_v).$$

Here $\chi_u \in \mathbb{R}^V$ is the vector with all zeros, and a single 1 at the coordinate corresponding to u. L_G^+ is the Moore-Penrose pseudo-inverse of L_G, which is positive semidefinite.

We may write $R(u, v)$ in cases where G is clear from context.

We will often use the notation $R_G(f) = R_G(u, v)$ where $f = (u, v)$ is an edge. It is important to note, however, that effective resistance is a function of the vertex pair, not the edge, and does not depend directly on the weight of f.

We now state several well-known and useful facts about effective resistance.

Fact II.6. *The effective resistance of an edge (u, v) is alternatively defined as*

$$R_G(u,v) = \max_{x \in \mathbb{R}^V} \frac{(x_u - x_v)^2}{x^\top L x}.$$

Fact II.7. *Effective resistance constitutes a metric on V.*

Fact II.8. *For any weighted graph $G = (V, F, z)$ and any edge $f \in F$ we have $z(f) \cdot R_G(f) \le 1$, with equality if and only if f is a bridge.*

Fact II.9. *For any weighted graph $G = (V, F, z)$ we have*

$$\sum_{f \in F} z(f) \cdot R_G(f) = n - 1.$$

C. Chernoff Bound

Theorem II.10 (Chernoff bound, see for example [33])**.** *Let $Z_1, Z_2, \ldots, Z_k$ be independent random variables in the range $[0, a]$. Furthermore, let $\sum Z_i = Z$ and let $\mu \ge \mathbb{E}(Z)$. Then for any $\delta \in (0, 1)$,*

$$\mathbb{P}\left(|Z - \mathbb{E}(Z)| \ge \delta\mu\right) \le 2 \exp\left(-\frac{\delta^2 \mu}{3a}\right).$$

III. Technical Overview

A. Analyzing Ordinary Graphs

The sparsification of ordinary graphs is a highly studied topic, with several techniques proposed for the construction of spectral sparsifiers throughout the years [7, 17, 16, 18, 19, 20]. However, the

analysis of spectral sparsifiers always relies heavily on the linear nature of the graph Laplacian, e.g., using matrix concentration results such as matrix Bernstein [34] or the work of [35]. This presents a significant problem when attempting to generalize these techniques to the highly non-linear setting of hypergraph spectral sparsification. Indeed, all previous results lose at least a factor of r due to this obstacle. We therefore dedicate the entirety of our first technical section (Section IV) to presenting a new proof of the existence of nearly linear spectral sparsifiers for ordinary graphs. We use the algorithm from [17], which constructs a sparsifier $\widetilde{G}$ by sampling each edge with probability proportional to its effective resistance. However, our proof avoids using matrix concentration inequalities, and instead relies on a more direct chaining technique for proving the concentration of $x^\top L_{\widetilde{G}} x$ around its expectation, i.e. $x^\top L_G x$, for all x simultaneously. To our knowledge, this is the first nearly-optimal direct analysis of spectral sparsification through effective resistance sampling. It will also be the basis of our main result, as we adapt it to the hypergraph setting in the full version of the paper.

More formally, for an input graph $G = (V, F, z)$, we define $\widetilde{G}$ as the result of sampling each edge f of G independently with probability $p(f) \approx z(f) \cdot R_G(f)$, and setting its weight to $\widetilde{z}(f) = z(f)/p(f)$. Our aim is then to prove

$$x^\top L_{\widetilde{G}} x \approx x^\top L_G x \tag{2}$$

simultaneously for all $x \in \mathbb{R}^V$. For simplicity we assume that $x^\top L_G x = 1$. Equation (2) is in fact the concentration of a random variable around its expectation, and so we can use Chernoff bound to prove it for any specific x. Our plan is then to use a combination of Chernoff and union bounds to prove it for all possible x. Since x can take any value in $\mathbb{R}^V$ we must discretize it to some ϵ-net while retaining a good approximation to its quadratic form, i.e. $x^\top L_G x$.

Let us take a closer look at the application of Chernoff bound to Equation (2): $x^\top L_{\widetilde{G}} x$ is the sum of the independent random variables $\widetilde{z}(u,v) \cdot (x_u - x_v)^2$ for $(u,v) \in F$; hence, by Theorem II.10, the strength of the bound depends crucially on the upper bound a on values that each random individual random variable can possibly attain. The maximum value of $\widetilde{z}(u,v) \cdot (x_u - x_v)^2$ is attained when (u,v) is sampled in $\widetilde{G}$, in which case it is $\approx (x_u - x_v)^2 / R_G(u,v)$. Thus

$$\begin{aligned}
&\mathbb{P}\left(x^\top L_{\widetilde{G}} x \not\approx x^\top L_G x\right) \\
&\lessapprox \exp\left(-\frac{1}{\max_{(u,v)\in F}(x_u - x_v)^2 / R_G(u,v)}\right).
\end{aligned}$$

This upper bound can be as bad as $\exp(-\widetilde{O}(1))$ and is far too crude for our purposes—no sufficiently sparse rounding scheme (i.e., discretization) exists for x. We turn to the technique of chaining—the use of progressively finer and finer rounding schemes.

As seen above, the strength of our Chernoff bound depends primarily on the quantity $(x_u - x_v)^2 / R_G(u,v)$ for each edge (u,v), which we call the "power" of the edge. Therefore, it makes sense to partition the edges of G into a logarithmic number of classes based on their power, that is F_i contains edges (u,v) for which $(x_u - x_v)^2 \approx 2^{-i} \cdot R_G(u,v)$. When focusing only on the subgraphs $G(F_i)$ induced by F_i, we get the more fine-tuned Chernoff bound

$$\begin{aligned}
&\mathbb{P}\left(x^\top L_{\widetilde{G}(F_i)} x \not\approx x^\top L_{G(F_i)} x\right) \\
&\lessapprox \exp\left(-\frac{1}{\max_{(u,v)\in F_i}(x_u - x_v)^2 / R_G(u,v)}\right) \\
&\lessapprox \exp\left(-2^i\right).
\end{aligned}$$

We thus have the task of proposing a rounding scheme $\varphi_i : \mathbb{R}^V \to \mathbb{R}^V$ specially for each class F_i such that

- the image of φ_i is a finite set of size at most $\approx \exp\left(2^i\right)$,
- the rounding approximately preserves the quantity $(x_u - x_v)^2$ for $(u,v) \in F_i$.

To gain more intuition on what such a rounding scheme must look like, we draw inspiration from the idea of resistive embedding from [17]. We map the edges in F_i, as well as our potential vector x, into vectors in $\mathbb{R}^n$ in such a way that all the relevant quantities arise as norms or scalar products:

$$\begin{aligned}
(u,v) \mapsto \underline{a_{u,v}} &= \frac{L_G^{+/2}(\chi_u - \chi_v)}{\left\| L_G^{+/2}(\chi_u - \chi_v) \right\|}, \\
x \mapsto \underline{y_x} &= L_G^{1/2} x.
\end{aligned}$$

Notice that both $\underline{a_{u,v}}$ and $\underline{y_x}$ are normalized (since $x^\top L_G x = 1$). Furthermore, the crucial quantity, the power of the edge (u,v) arises as the square of a scalar product:

$$\langle \underline{a_{u,v}}, \underline{y_x} \rangle^2 = \frac{(x^\top(\chi_u - \chi_v))^2}{(\chi_u - \chi_v)^\top L_G^+ (\chi_u - \chi_v)} = \frac{(x_u - x_v)^2}{R_G(u,v)}.$$

Thus we are interested in rounding $\underline{y_x}$ in a way that preserves $\langle \underline{a_{u,v}}, \underline{y_x} \rangle^2$ up to small multiplicative error *in all cases where it was* $\approx 2^{-i}$ *to begin with.*

Thus, it suffices to guarantee an additive error of at most $\lessapprox 2^{-i}$ in our rounding scheme. This is the known problem of "compression of approximate inner products" and has been previously studied; [36] guarantees a rounding scheme whose image is of size at most $\approx \exp\left(2^i\right)$. This can be translated into a rounding scheme for $x \in \mathbb{R}^V$, with the same image-size, exactly as desired (see Lemma IV.2).

With the desired rounding scheme in hand, we can now use a combination of Chernoff and union bounds to prove that for all x simultaneously

$$x^\top L_{\widetilde{G(F_i)}} x \approx x^\top L_{G(F_i)} x.$$

Summing this over all edge-classes gives us Equation 2.

For the detailed proof, which is considerably more complicated than the above sketch, see the proof of Theorem IV.1 in Section IV.

B. Extension to Hypergraphs

To adapt the previous argument to the hypergraph setting, we use the idea of balanced weight assignments from [10]. Essentially, we construct an ordinary graph $G = (V, F, z)$ to accompany our input hypergraph $H = (V, E, w)$ by replacing each hyperedge e with a clique F_e over the vertices in e. However, unlike in some previous works on hypergraph sparsification, the clique F_e is not assigned weights uniformly, but instead the weight is carefully distributed among the edges. Intuitively, all the weight is shifted onto the most "important" edges. In the case of [10], the measure of importance was "strength", a quantity relevant to cut sparsification, while in our case it is effective resistance.

More formally, a weighting assignment z of the cliques is considered γ-balanced if for all $e \in E$

- $\sum_{f \in F_e} z(f) = w(e)$,
- and

$$\gamma \cdot \min_{g \in F_e:\ z(g) > 0} R_G(g) \geq \max_{f \in F_e} R_G(f).$$

In words, all but the zero-weight edges of F_e have approximately the same effective resistance. This allows hyperedge e to inherit this effective-resistance value as its importance when sampling hyperedges. Our task is now to prove the existence of balanced weight assignments for all hypergraphs, and then to adapt the proof of Section IV.

Finding balanced weight assignments.: In [10], balanced weight assignments are constructed through the following intuitive process: Find a pair of edges violating the second constraint, that is $f, g \in F_e$ where $z(g) \neq 0$ and f has significantly higher importance than g. Then shift weight from g to f; this alleviates the constraint violation either because the importances of g and f become more similar, or simply because the weight of g decreases to 0. We call this resolving the imbalance of f and g. [10] strings together such steps, carefully ordered and discretized, to eventually produce a balanced weight assignment of the input hypergraph.

However, their analysis relies heavily on a certain lemma about how "strength" (their measure of edge importance) behaves under weight updates. Lemma 6 of [10] states that altering the weight of an edge f, will not affect edges of significantly greater "strength" than f. This is not the case for effective resistances. It is easy to construct scenarios to the contrary; even ones in which altering the weight of edges of low resistance can increase the maximum effective resistance in the graph.

Thus the analysis of [10] does not extend to our setting. Instead we use a potential function argument to say that we make irreversible progress whenever we resolve the imbalance of two edges f and g. Our choice of potential function is surprising, and is one of the main technical contributions of this paper. We define the spanning tree potential (or ST-potential) of a connected weighted ordinary graph $G = (V, F, z)$, denoted $\Psi(G)$. For edge weights that equal 1 uniformly (that is for unweighted graphs) it is simply the logarithm of the number of distinct spanning trees in G. In weighted graphs it is generalized to

$$\Psi(G) = \log\left(\sum_{T \in \mathbb{T}} \prod_{f \in T} z(f)\right),$$

where $\mathbb{T}$ denotes the set of all spanning trees in G. Due to the relationship between spanning tree sampling and effective resistances (see for example [37]) we can prove a crucial update formula for $\Psi(G)$: if an edge f has its weight changed by $\lambda \in \mathbb{R}$, the ST-potential increases by $\log(1 + \lambda \cdot R(f))$. Since whenever we resolve the imbalance of a pair of edges, we shift weight from the edge of lower effective resistance to that of higher effective resistance, this allows us to argue that the ST-potential always increases throughout the process, which eventually terminates in a balanced weight assignment (see the full version of the paper for more details).

This proves the existence of balanced weight assignments, which suffices to show the existence of nearly linear size spectral sparsifiers for all hypergraphs. However, to improve running time (from exponential to polynomial in the input size), we introduce the novel concept of *approximate* balanced weight assignments, by slightly relaxing the defini-

tion. These are still sufficient to aid in constructing spectral sparsifiers, and are faster to construct using an alternate algorithm.

For more details on the ST-potential, as well as the construction of balanced weight assignments see the full version of the paper.

Using balanced weight assignments to construct hypergraph spectral sparsifiers.: Given a hypergraph $H = (V, E, w)$ and its balanced weight assignment $G = (V, F, z)$ we assign importance to each hyperedge proportionally to the maximum effective resistance in F_e (the clique corresponding to e). Thus we perform importance sampling, which samples each hyperedge independently with probability $p(e) \approx w(e) \cdot \max_{f \in F_e} R_G(f)$.

The broad strokes of the hypergraph proof in Section 6 of the full paper proceed very similarly to those of the proof for ordinary graphs in Section IV. However, numerous details need to be figured out in order to bridge the gap between the two settings. It is interesting to note that our rounding scheme is exactly the same as in Section IV, to the point of even being defined in terms of G, not H. (Indeed it is impossible to define such a rounding scheme directly in terms of H; Lemma IV.2 relies heavily on the linear nature of the ordinary graph Laplacian.) Nevertheless, we manage to extend the approximation guarantee of the rounding scheme from edges to hyperedges.

For the detailed analysis of hypergraph spectral sparsification through effective resistance-based importance sampling, see the full version of the paper.

C. Speed-Up

Using the above results, we can put together a polynomial time algorithm for spectral sparsification of hypergraphs. Simply run our algorithm to produce an approximate balanced weight assignment, and then use importance sampling. The bottleneck of this procedure is constructing the weight assignment, which takes time $O(m \cdot \text{poly}(n))$. (Given the weight assignment, it is trivial to implement importance sampling).

In the full version of the paper, we reduce this to the nearly optimal $\widetilde{O}(mr + \text{poly}(n))$. (Note that $O(mr)$ is the size of the input.) Our first step is the common trick of using a faster sparsification algorithm, but one which produces a larger output, to preprocess the input hypergraph. We use the algorithm of [8] which – with small modifications – can be made to run in the desired $\widetilde{O}(mr + \text{poly}(n))$ time. The resulting hypergraph has only polynomially many hyperedges (in n); however, the aspect ratio of edge weights (that is the ratio between the largest and smallest edge weights) can naturally be exponential in n.

Unfortunately, our basic algorithm for constructing apporximate balanced weight assignments scales linearly in the aspect ratio of edge weights, and so we propose another algorithm for finding a balanced weight assignment – one specifically designed for the setting when the input graph is polynomially sparse, but has exponential aspect ratio.

Suppose our input hypergraph, $H = (V, E, w)$ has edge weights in the range $[1, \exp(n)]$. We then divide hyperedges into weight categories such as $E_i = \{e \in E | w(e) \in [n^{10(i-1)}, n^{10i})\}$. We then bisect H into two hypergraphs H_1 and H_2, where H_1 contains all hyperedges in odd numbered categories, and H_2 all those in even numbered categories. This results in hyergraphs (H_1 and H_2) where hyperedges fall into extremely well-separated categories; so extremely in fact, that the weight of a hyperedge in a higher category (for example $e \in E_i \subseteq H_1$) has higher weight than all hyperedges of all lower categories *combined*, that is

$$w(e) \gg \sum_{e' \in E_{<i} \cap H_1} w(e').$$

We use this property to independently find weight assignments on H_1 and H_2. Informally, we go through the categories of hyperedges, from heaviest to lightest, resolving all instances of imbalance. We never return to a category once we moved on, and we prove that no amount of changes to the weight assignment of lower categories can disrupt the balance of a higher category, due to the huge discrepency in weights. For a more detailed and formal argument see the full version of the paper.

IV. Warm-Up: Ordinary Graphs

We begin by reproving the famous theorem of Spielman and Srivastava [17], which states that sampling edges of a graph with probability proportional to their effective resistance (and then reweighting appropriately) results in a spectral sparsifier with high probability. We prove a somewhat weaker version of the theorem, where we oversample by an $O(\epsilon^{-4} \log^3 n)$ factor, as opposed to the $\epsilon^{-2} \log n$ factor in the original. Another slight difference is that our version samples every edge independently, instead of sampling a predetermined number of edges with replacement in [17]. More recent proofs of the theorem of [17] that use the matrix Bernstein inequality as opposed to [35] also use the same distribution as ours.

Theorem IV.1 (A slightly weaker version of [17])**.** *Let $G = (V, F, z)$ be a weighted ordinary graph with n*

vertices and let $1/n \le \epsilon \le 1/2$. *For every edge* $f \in F$, *let* $p(f) = \min(1, \lambda \cdot z(f) \cdot R_G(f))$ *for a sufficiently large factor* $\lambda = \Theta(\epsilon^{-4} \log^3 n)$. *Sample each edge* $f \in F$ *independently with probability* $p(f)$, *and give it weight* $\widetilde{z}(f) = z(f)/p(f)$ *if sampled. The resulting graph,* $\widetilde{G} = (V, \widetilde{F}, \widetilde{z})$ *is an* ϵ*-spectral sparsifier of* G *with probability at least* $1 - O(\log n/n)$.

The original proof of this theorem used a concentration bound for matrices [35] (later simplified to use the matrix Bernstein inequality) to prove that $x^\top L_{\widetilde{G}} x$ is close to its expectation simultaneously for all $x \in \mathbb{R}^n$, as required by Definition II.2. This type of argument is difficult to adapt to hypergraph sparsification, because the extension of quadratic forms to hypergraphs is highly non-linear. We thus present an alternative proof that uses more primitive techniques to bypass the reliance on linear algebra.

Proof of Theorem IV.1: By Definition II.2, we must prove that for every $x \in \mathbb{R}^V$,

$$x^\top L_{\widetilde{G}} x \in (1 \pm \epsilon) \cdot x^\top L_G x. \tag{3}$$

We may assume without loss of generality that $x^\top L_G x = 1$. We denote the set of vectors x where this is satisfied as $S^G \subseteq \mathbb{R}^V$. Furthermore, we simplify notation by denoting L_G as L, and $L_{\widetilde{G}}$ as $\widetilde{L}$. Moreover, for any subset of edges $F' \subseteq F$, we denote the Laplacian of the subgraph of G corresponding to F' by $L_{F'}$, and similarly for the subgraph of $\widetilde{G}$ by $\widetilde{L}_{F'}$.

It is clear from the construction of $\widetilde{G}$ that

$$\mathbb{E}\left(x^\top \widetilde{L} x\right) = x^\top L x.$$

Therefore, we are in effect trying to prove the concentration of a random variable around its expectation in Equation (3). Indeed, for any specific x, Equation (3) holds with high probability by Chernoff bound (Theorem II.10). (One can consider $x^\top L x$ as the sum of independent random variables of the form $\widetilde{z}(u,v) \cdot (x_u - x_v)^2$.)

In order to prove the concentration for all $x \in S^G$ simultaneously, we employ a net argument, where we "round" x to some vector from a finite set and apply a union bound on the rounded vectors. However, our rounding scheme is progressive and has $O(\log n)$ "levels" with increasingly finer resolution. Each x will then determine a partition of the edges into levels, and we will prove concentration for each rounded vector and each level (subset of edges), and then apply a union bound over all these choices.

The existence of these rounding functions is guaranteed by the following lemma, which we will prove in Section IV-A.

Lemma IV.2. *Let* $G = (V, F, z)$ *be a connected weighted graph. Then for every* $i \in \mathbb{N}$ *there exists a rounding function*

$$\varphi_i : S^G \to \mathbb{R}^V$$

such that for all $x \in S^G$, *denoting* $x^{(i)} := \varphi_i(x)$, *we have:*

1) *The image of* φ_i *is a finite set of cardinality* $|\varphi_i(S^G)| \le \exp\left(800C \log n \cdot 2^i/\epsilon^2\right)$, *where* $C > 0$ *is the absolute constant from Theorem IV.5.*
2) *For every edge* $f = (u,v) \in F$ *such that* $\max\left((x_u - x_v)^2, (x_u^{(i)} - x_v^{(i)})^2\right) \ge 2^{-i} \cdot R_G(f)$,

$$(x_u - x_v)^2 \in \left(1 \pm \frac{\epsilon}{7}\right) \cdot (x_u^{(i)} - x_v^{(i)})^2.$$

The second guarantee of Lemma IV.2 can be expressed in terms of the Laplacian of a single edge, resulting in the following corollary.

Corollary IV.3. *For a rounding function* φ *satisfying the guarantees of Lemma IV.2, and an edge* $f = (u,v) \in F$ *such that* $\max\left((x_u - x_v)^2, (x_u^{(i)} - x_v^{(i)})^2\right) \ge 2^{-i} \cdot R_G(f)$,

$$x^\top L_{\{f\}} x \in \left(1 \pm \frac{\epsilon}{7}\right) \cdot x^{(i)\top} L_{\{f\}} x^{(i)}.$$

Let us take a sequence of the rounding functions φ_i guaranteed by Lemma IV.2 for $i = 1, \ldots, I := \log_2(7n/\epsilon) \le 3 \log n$. For each $x \in S^G$, it yields a sequence of rounded vectors $x^{(i)} = \varphi_i(x)$ for $i = 1, \ldots, I$. Furthermore, we use $x^{(i)}$ to define the subset of edges $F_i' \subseteq F$ as

$$\left\{ f = (u,v) \in F \;\middle|\; \left(x_u^{(i)} - x_v^{(i)}\right)^2 \ge 2^{-i} \cdot R_G(f) \right\}.$$

That is, the second guarantee of Lemma IV.2 holds for φ_i on edges in F_i'. Finally, we use $\{F_i'\}_i$ to partition F as follows. Let the base case be $F_0 = F_0' := \{f \in F \mid p(f) = 1\}$, where we recall that $p(f) = \min(1, \lambda \cdot z(f) \cdot R_G(f))$. For each $i \in [I]$, let $F_i := F_i' \setminus \bigcup_{j=0}^{i-1} F_j'$, and finally let $F_{I+1} = F \setminus \bigcup_{i=0}^{I} F_i'$.

Thus we have partitioned F in such a way that the second guarantee of Lemma IV.2 applies to edges in F_i, with respect to φ_i. Furthermore, F_i are defined in terms of $x^{(i)}$ (and $x^{(j)}$ for $j < i$) instead of x, so that the number of possible sets F_i is finite, and bounded thanks to the first guarantee of Lemma IV.2.

We establish the following claim for later use.

Claim IV.4. *For all* $i \in [I]$ *and* $f = (u,v) \in F_i$, *we have*

$$(x_u^{(i)} - x_v^{(i)})^2 \le 3 \cdot 2^{-i} \cdot R_G(f).$$

Proof: The second guarantee of Lemma IV.2 for φ_i applies to f, and thus $(x_u^{(i)} - x_v^{(i)})^2 \leq (x_u - x_v)^2 \cdot (1-\epsilon/7)^{-1}$.

Consider first the case $i = 1$. By Fact II.6 and since $x \in S^G$, we have $(x_u - x_v)^2 \leq R_G(f) \cdot x^\top L x = R_G(f)$, and we indeed get $(x_u^{(i)} - x_v^{(i)})^2 \leq R_G(f) \cdot (1-\epsilon/7)^{-1} \leq 3 \cdot 2^{-1} \cdot R_G(f)$.

Now consider $i > 1$, and suppose towards contradiction that $(x_u^{(i)} - x_v^{(i)})^2 > 3 \cdot 2^{-i} \cdot R_G(f)$. Notice that the second guarantee of Lemma IV.2 also applies to f for φ_{i-1}, and thus $(x_u^{(i-1)} - x_v^{(i-1)})^2 \geq (x_u^{(i)} - x_v^{(i)})^2 \cdot (1+\epsilon/7)^{-1} \cdot (1-\epsilon/7) \geq 2^{-i+1} \cdot R_G(f)$. This implies that $f \in F'_{i-1}$, which contradicts the assumption $f \in F_i = F'_i \setminus \bigcup_{j=0}^{i-1} F'_j$. ∎

We will consider each group of edges F_i separately, and prove that $x^\top \widetilde{L}_{F_i} x$ concentrates around its expectation, $x^\top L_{F_i} x$. More precisely, we will first prove concentration for every specific $(x^{(i)}, F_i)$, and then extend the concentration to all possibilities simultaneously via union bound. This is well-defined because each F_i depends on $x^{(1)}, \ldots, x^{(i)}$ but not directly on x.

Edges in F_0.: By definition, every edge $f \in F_0$ has $p(f) = 1$, and thus $x^\top \widetilde{L}_{F_0} x$ is completely deterministic and equal to $x^\top L_{F_0} x$.

Edges in F_i for $i \in [I]$.: Note that F_i is designed so that, by Corollary IV.3, for every edge $f \in F_i$ we have $\left|x^\top L_{\{f\}} x - x^{(i)\top} L_{\{f\}} x^{(i)}\right| \leq \epsilon/7 \cdot x^{(i)\top} L_{\{f\}} x^{(i)}$, and since $\widetilde{L}_{\{f\}}$ is a multiple of $L_{\{f\}}$ similarly have $\left|x^\top \widetilde{L}_{\{f\}} x - x^{(i)\top} \widetilde{L}_{\{f\}} x^{(i)}\right| \leq \epsilon/7 \cdot x^{(i)\top} \widetilde{L}_{\{f\}} x^{(i)}$. Informally, this allows us to prove concentration only for vectors $x^{(i)}$ instead of all x, and thus we next aim to bound the error

$$\left|x^{(i)\top} L_{F_i} x^{(i)} - x^{(i)\top} \widetilde{L}_{F_i} x^{(i)}\right|$$

for each $i \in [I]$ with high probability. It will then remain to bound the error introduced on the remaining edges (the ones in F_{I+1}).

Fix $i \in [I]$ and notice that over all possible vectors $x \in S^G$, there are only finitely many possible values for $(x^{(i)}, F_i)$. Therefore, we can focus on a single value of $x^{(i)}$ and F_i, and then use a union bound over all settings.

Let us therefore fix also $x^{(i)}$ and F_i. We will use Chernoff bounds to prove that with high probability, over the randomness of sampling edges to $\widetilde{G}$,

$$\left|x^{(i)\top} L_{F_i} x^{(i)} - x^{(i)\top} \widetilde{L}_{F_i} x^{(i)}\right| \leq \frac{\epsilon}{7I}. \tag{4}$$

Indeed, note that

$$x^{(i)\top} \widetilde{L}_{F_i} x^{(i)} = \sum_{f=(u,v) \in F_i} \widetilde{z}(f) \cdot (x_u^{(i)} - x_v^{(i)})^2,$$

where $\widetilde{z}(f)$ are independent random variables with expectation $\mathbb{E}(\widetilde{z}(f)) = z(f)$. Therefore, we can apply the Chernoff bound from Theorem II.10 with $\{Z_i\}_i$ being $\widetilde{z}(f) \cdot (x_u - x_v)^2$ for each $f = (u,v) \in F_i$, and their sum being $Z = x^{(i)\top} \widetilde{L}_{F_i} x^{(i)}$ with $\mathbb{E}(Z) = x^{(i)\top} L_{F_i} x^{(i)}$. We need to set a as an upper bound on $\widetilde{z}(f) \cdot (x_u^{(i)} - x_v^{(i)})^2$. Observe that $\widetilde{z}(f)$ is maximal when f is sampled, in which case it equals $z(f)/p(f)$ where $p(f) = \lambda \cdot z(f) \cdot R_G(f)$, since $f \notin F_0$. We thus get, using Claim IV.4, that for all $f = (u,v) \in F_i$,

$$\begin{aligned}\frac{z(f) \cdot (x_u^{(i)} - x_v^{(i)})^2}{\lambda \cdot z(f) \cdot R_G(f)} &= \frac{1}{\lambda} \cdot \frac{(x_u^{(i)} - x_v^{(i)})^2}{R_G(f)} \\ &\leq \frac{3 \cdot 2^{-i}}{\lambda} =: a.\end{aligned}$$

We let $\delta := \epsilon/(14I)$, we can bound

$$\begin{aligned}x^{(i)\top} L_{F_i} x^{(i)} &\leq \left(1 + \frac{\epsilon}{7}\right) \cdot x^\top L_{F_i} x \leq \left(1 + \frac{\epsilon}{7}\right) \cdot x^\top L x \\ &= 1 + \frac{\epsilon}{7} \leq 2 =: \mu.\end{aligned}$$

(This is true for an arbitrary preimage $x \in \varphi_i^{-1}(x^{(i)})$.)

Finally, Theorem II.10 implies

$$\begin{aligned}&\mathbb{P}\left(\left|x^{(i)\top} L_{F_i} x^{(i)} - x^{(i)\top} \widetilde{L}_{F_i} x^{(i)}\right| \geq \frac{\epsilon}{7I}\right) \\ \leq &2 \exp\left(-\frac{\delta^2 \mu}{3 \cdot a}\right) \\ = &2 \exp\left(-\frac{\frac{\epsilon^2}{196 I^2} \cdot 2}{9 \cdot 2^{-i}/\lambda}\right) \\ \leq &2 \exp\left(-\frac{\epsilon^2 \cdot 2^i \cdot \lambda}{10000 \log^2(n)}\right) \\ = &2 \exp\left(-\frac{2000 C \log n \cdot 2^i}{\epsilon^2}\right),\end{aligned}$$

where the last step by setting $\lambda = 2 \cdot 10^7 \cdot C \log^3 n/\epsilon^4$, where $C > 0$ is the absolute constant from Theorem IV.5.

We can now use a union bound to bound the probability that Equation (4) holds simultaneously for all values of $(x^{(i)}, F_i)$. F_i depends only on F'_j for $j \in [i]$, which in turn depend on $x^{(j)}$ for the same values of j. By the first guarantee of Lemma IV.2, the number of possible vectors $x^{(j)}$ is at most $\exp\left(800 C \log n \cdot 2^j/\epsilon^2\right)$, where $C > 0$ is the absolute constant from Theorem IV.5. Therefore, the number of possible pairs $(x^{(i)}, F_i)$ is at most

$$\begin{aligned}&\prod_{j=1}^{i} \exp\left(\frac{800 C \log n \cdot 2^j}{\epsilon^2}\right) \\ = &\exp\left(\sum_{j=1}^{i} \frac{800 C \log n \cdot 2^j}{\epsilon^2}\right)\end{aligned}$$

$$\leq \exp\left(\frac{1600C\log n\cdot 2^i}{\epsilon^2}\right).$$

Finally, the probability that Equation (4) *does not* hold simultaneously for all pairs $(x^{(i)}, F_i)$ is at most

$$\exp\left(\frac{1600C\log n\cdot 2^i}{\epsilon^2}\right)\cdot 2\exp\left(-\frac{2000C\log n\cdot 2^i}{\epsilon^2}\right)$$
$$= 2\exp\left(-\frac{400C\log n\cdot 2^i}{\epsilon^2}\right) \leq \frac{1}{n}.$$

Edges in F_{I+1}.: First we show that for any $x \in S^G$ and any edge $f = (u,v) \in F_{I+1}$ we have that $(x_u - x_v)^2 \leq \epsilon \cdot R_G(f)/(6n)$. Suppose for contradiction that this is not the case. Then the second guarantee of Lemma IV.2 applies and $(x_u^{(I)} - x_v^{(I)})^2 \geq (x_u - x_v)^2 \cdot (1-\epsilon/7) \geq \epsilon \cdot R_G(e)/(6n) \cdot (1-\epsilon/7) \geq \epsilon \cdot R_G(e)/(7n)$. Therefore $f \in F'_I$, which contradicts the assumption $f \in F_{I+1}$. (Here we used that I was defined to be $\log_2(7n/\epsilon)$.)

Next, we would like to bound $\left|x^\top \widetilde{L}_{F_{I+1}} x - x^\top L_{F_{I+1}} x\right|$ by showing that both terms are small. First,

$$\begin{aligned} x^\top L_{F_{I+1}} x &= \sum_{f=(u,v)\in F_{I+1}} z(f)\cdot(x_u - x_v)^2 \\ &\leq \sum_{f\in F_{I+1}} z(f)\cdot\epsilon\cdot\frac{R_G(f)}{6n} \\ &\leq \frac{\epsilon}{6n}\cdot\sum_{f\in F} z(f)\cdot R_G(f) \leq \frac{\epsilon}{6}, \end{aligned}$$

where the last inequality uses Fact II.9. Second, we start similarly,

$$\begin{aligned} x^\top \widetilde{L}_{F_{I+1}} x &= \sum_{f=(u,v)\in F_{I+1}} \widetilde{z}(f)\cdot(x_u - x_v)^2 \\ &\leq \sum_{f\in F_{I+1}} \widetilde{z}(f)\cdot\epsilon\cdot\frac{R_G(f)}{6n} \\ &\leq \frac{\epsilon}{6n}\cdot\sum_{f\in F} \widetilde{z}(f)\cdot R_G(f), \end{aligned}$$

and ideally we would like to show that $\sum \widetilde{z}(f) \cdot R_G(f) \leq 2n$. This is not always true, but it is a random event, independent of the choice of x, and can be shown to hold with high probability using our Chernoff bound from Theorem II.10. Indeed, $\widetilde{z}(f)\cdot R_G(f)$ are independent random variables with maximum value when f is sampled, in which case $\widetilde{z}(f) = z(f)/p(f)$, and thus

$$\begin{aligned} \widetilde{z}(f)\cdot R_G(f) &\leq \frac{z(f)\cdot R_G(f)}{p(f)} \\ &= \frac{z(f)\cdot R_G(f)}{\min(1, \lambda\cdot z(f)\cdot R_G(f))} \\ &= \max\left(z(f)\cdot R_G(f), 1/\lambda\right) \\ &=: a \leq 1, \end{aligned}$$

where the last inequality uses Fact II.8. We apply Theorem II.10 by setting $\delta := 1$ and $\mu := n$ (which we may do by Fact II.9), and obtain

$$\mathbb{P}\left(\sum_{f\in F}\widetilde{z}(f)\cdot R_G(f) \geq 2n\right) \leq 2\exp\left(-\frac{n}{3}\right).$$

Therefore, with probability at least $1 - O(1/n)$,

$$\left|x^\top \widetilde{L}_{F_{I+1}} x - x^\top L_{F_{I+1}} x\right| \leq \frac{\epsilon}{2}. \tag{5}$$

Putting everything together.: By the above derivations, Equation (4) holds for all i and all $(x^{(i)}, F_i)$ simultaneously, as well as Equation (5) holds with probability at least $1 - O(\log n/n)$. Assuming henceforth that this high probability event occurs, we shall deduce that Equation (3) holds for all $x \in S^G$. Indeed, by the triangle inequality and Equation (5),

$$\begin{aligned} \left|x^\top \widetilde{L} x - x^\top L x\right| &\leq \sum_{i=0}^{I+1}\left|x^\top \widetilde{L}_{F_i} x - x^\top L_{F_i} x\right| \\ &\leq 0 + \sum_{i=1}^{I}\left|x^\top \widetilde{L}_{F_i} x - x^\top L_{F_i} x\right| + \frac{\epsilon}{2}. \end{aligned}$$

Now for each $i \in [I]$, we can approximate terms involving x by $x^{(i)}$ and vice versa, formalized by the aforementioned fact that for every $(u,v) \in F_i$ we have $|(x_u - x_v)^2 - (x_u^{(i)} - x_v^{(i)})^2| \leq \epsilon/7 \cdot (x_u^{(i)} - x_v^{(i)})^2$ (see the second condition of Lemma IV.2 and the definition of $F'_i \subseteq F_i$), and get

$$\begin{aligned} &\left|x^\top \widetilde{L}_{F_i} x - x^\top L_{F_i} x\right| \\ \leq &\left|x^\top \widetilde{L}_{F_i} x - x^{(i)\top} \widetilde{L}_{F_i} x^{(i)}\right| + \\ &\left|x^{(i)\top} \widetilde{L}_{F_i} x^{(i)} - x^{(i)\top} L_{F_i} x^{(i)}\right| + \\ &\left|x^{(i)\top} L_{F_i} x^{(i)} - x^\top L_{F_i} x\right| \\ \leq &\frac{\epsilon}{7}\cdot x^{(i)\top} \widetilde{L}_{F_i} x^{(i)} + \\ &\left|x^{(i)\top} \widetilde{L}_{F_i} x^{(i)} - x^{(i)\top} L_{F_i} x^{(i)}\right| + \\ &\frac{\epsilon}{7}\cdot x^{(i)\top} L_{F_i} x^{(i)} \end{aligned}$$

now we use the triangle inequality,

$$\begin{aligned} \leq &\frac{\epsilon}{7}\cdot x^{(i)\top} L_{F_i} x^{(i)} + \\ &\left(1 + \frac{\epsilon}{7}\right)\cdot\left|x^{(i)\top} \widetilde{L}_{F_i} x^{(i)} - x^{(i)\top} L_{F_i} x^{(i)}\right| + \\ &\frac{\epsilon}{7}\cdot x^{(i)\top} L_{F_i} x^{(i)} \end{aligned}$$

and now we crucially use Equation (4),

$$\leq \left(1+\frac{\epsilon}{7}\right)\cdot\frac{\epsilon}{7I}+\frac{2\epsilon}{7}\cdot\left(1-\frac{\epsilon}{7}\right)^{-1}\cdot x^\top L_{F_i}x$$
$$\leq \frac{\epsilon}{6I}+\frac{2\epsilon}{6}\cdot x^\top L_{F_i}x.$$

Substituting this into our previous bound, we obtain

$$\begin{aligned}\left|x^\top\widetilde{L}x-x^\top\widetilde{L}x\right| &\leq \sum_{i=1}^{I}\left(\frac{\epsilon}{6I}+\frac{\epsilon}{3}\cdot x^\top L_{F_i}x\right)+\frac{\epsilon}{2}\\ &\leq \frac{\epsilon}{6}+\frac{\epsilon}{3}\cdot x^\top Lx+\frac{\epsilon}{2}\\ &= \epsilon\cdot x^\top Lx,\end{aligned}$$

where the last equality uses $x^\top Lx=1$. This completes the proof of Theorem IV.1. □

A. Proof of Lemma IV.2

To prove Lemma IV.2, we use the following Theorem:

Theorem IV.5 (Theorem VI.1 of [36]). *Let $a_1,\ldots,a_m\in\mathbb{R}^n$ be vectors of norm at most 1 and let $\eta\in(0,1)$. Then, over all vectors $y\in\mathbb{R}^n$ of norm at most 1, the number of possible values of the "rounded vector"*

$$\left(\left\lfloor\frac{\langle a_1,y\rangle}{\eta}\right\rfloor,\left\lfloor\frac{\langle a_2,y\rangle}{\eta}\right\rfloor,\ldots,\left\lfloor\frac{\langle a_k,y\rangle}{\eta}\right\rfloor\right)$$

is at most $\exp\left(\frac{C\log m}{\eta^2}\right)$ for some absolute constant $C>0$.

Remark IV.6. *In fact, the original theorem (Theorem 6.1 in [36]) is stated with stronger requirements on m and η, and a stronger consequence. However, we can easily get the weaker upper bound of $\exp(O(\log m/\eta^2))$ stated in Theorem IV.5 of this paper, by setting the variables appropriately: $\varepsilon:=\eta$, $n:=\max(m,1/\eta^2)$, and $k:=n$, where the left hand side always represents their variable names and the right hand side ours.*

Proof of Lemma IV.2: We use the idea of resistive embedding introduced in []. Note that $L=L_G$ is a positive semidefinite matrix, and we denote by $L^{+/2}$ the square root of its Moore-Penrose pseudo-inverse. For each (unordered) vertex pair (u,v), let $b_{u,v}\in\mathbb{R}^V$ be the vector with all zero coordinates, except for the coordinates associated with u and v, which are 1 and -1 (ordered arbitrarily). With each vertex pair (u,v), we associate the vector

$$a_{u,v}=\frac{L^{+/2}b_{u,v}}{\|L^{+/2}b_{u,v}\|_2}.$$

Furthermore, we associate with each $x\in S^G$ the vector $y_x=L^{1/2}x$.

We can then apply Theorem IV.5 to $\{a_{u,v}\mid(u,v)\in\binom{V}{2}\}$ and all possible y_x, setting $\eta=\epsilon\cdot 2^{-i/2}/20$. Indeed, $a_{u,v}$ is normalized by definition, and also y_x is normalized because $x\in S^G$ and thus

$$\|y_x\|_2^2=x^\top L^{1/2}L^{1/2}x=x^\top Lx=1.$$

For each possible value of the rounded vector

$$\left(\left\lfloor\frac{\langle a_{u,v},y_x\rangle}{\eta}\right\rfloor\right)_{(u,v)\in\binom{V}{2}}$$

choose a representative $x\in S^G$, and let φ_i map each $x\in S^G$ to its representative (i.e., with the same rounded vector). Then by Theorem IV.5, the image of φ_i is of size $|\varphi_i(S^G)|\leq\exp\left(800C\log n\cdot 2^i/\epsilon^2\right)$, as claimed. Recall that we denote $\varphi_i(x)$ by $x^{(i)}$; then

$$\left(\left\lfloor\frac{\langle a_{u,v},y_x\rangle}{\eta}\right\rfloor\right)_{(u,v)\in\binom{V}{2}} = \left(\left\lfloor\frac{\langle a_{u,v},y_{x^{(i)}}\rangle}{\eta}\right\rfloor\right)_{(u,v)\in\binom{V}{2}}.$$

It follows that for all $f=(u,v)\in F$ and all $x\in S^G$,

$$|\langle a_{u,v},y_x\rangle-\langle a_{u,v},y_{x^{(i)}}\rangle|\leq\eta. \tag{6}$$

Furthermore, $b_{u,v}$ is perpendicular to the null-space of L (which is spanned by the all-ones vector because G is connected), thus $L^{1/2}L^{+/2}b_{u,v}=b_{u,v}$ and

$$\begin{aligned}\langle a_{u,v},y_x\rangle^2 &= \frac{\left(x^\top L^{1/2}L^{+/2}b_{u,v}\right)^2}{b_{u,v}^\top L^{+/2}L^{+/2}b_{u,v}}=\frac{\left(x^\top b_{u,v}\right)^2}{b_{u,v}^\top L^{+}b_{u,v}}\\ &= \frac{(x_u-x_v)^2}{R_G(u,v)}.\end{aligned} \tag{7}$$

To prove the second guarantee of Lemma IV.2, let $f=(u,v)\in F$ and $x\in S^G$ and consider first the case $\left(x_u^{(i)}-x_v^{(i)}\right)^2\geq 2^{-i}\cdot R_G(f)$, which by (7) is equivalent to $\langle a_{u,v},y_{x^{(i)}}\rangle^2\geq 2^{-i}$. This means that the absolute error bound η in Equation (6) implies a relative error bound, namely,

$$\begin{aligned}|\langle a_{u,v},y_x\rangle-\langle a_{u,v},y_{x^{(i)}}\rangle|\leq\eta &= \frac{\epsilon\cdot 2^{-i/2}}{20}\\ &\leq \frac{\epsilon}{20}\cdot\left|\langle a_{u,v},y_{x^{(i)}}\rangle\right|.\end{aligned}$$

The other case $(x_u-x_v)^2\geq 2^{-i}\cdot R_G(f)$ is similar up to constants; by (7), this case is equivalent to $\langle a_{u,v},y_x\rangle^2\geq 2^{-i}$, and thus

$$\begin{aligned}&|\langle a_{u,v},y_x\rangle-\langle a_{u,v},y_{x^{(i)}}\rangle|\\ \leq&\eta=\frac{\epsilon\cdot 2^{-i/2}}{20}\\ \leq&\frac{\epsilon}{20}\cdot\left|\langle a_{u,v},y_x\rangle\right|,\end{aligned}$$

which implies

$$\begin{aligned}
&|\langle a_{u,v}, y_x\rangle - \langle a_{u,v}, y_{x^{(i)}}\rangle| \\
\leq &\frac{\epsilon}{20}\cdot\left(1-\frac{\epsilon}{20}\right)^{-1}|\langle a_{u,v}, y_x\rangle| \\
\leq &\frac{\epsilon}{16}\cdot|\langle a_{u,v}, y_{x^{(i)}}\rangle|\,.
\end{aligned}$$

Now in both cases,

$$\begin{aligned}
&\left|\langle a_{u,v}, y_x\rangle^2 - \langle a_{u,v}, y_{x^{(i)}}\rangle^2\right| \\
=&|\langle a_{u,v}, y_x\rangle - \langle a_{u,v}, y_{x^{(i)}}\rangle|\cdot|\langle a_{u,v}, y_x\rangle + \langle a_{u,v}, y_{x^{(i)}}\rangle| \\
\leq &\frac{\epsilon}{16}\cdot|\langle a_{u,v}, y_{x^{(i)}}\rangle|\cdot\left(2+\frac{\epsilon}{16}\right)\cdot|\langle a_{u,v}, y_{x^{(i)}}\rangle| \\
\leq &\frac{\epsilon}{7}\cdot\langle a_{u,v}, y_{x^{(i)}}\rangle^2.
\end{aligned}$$

Using (7) and scaling by $R_G(u,v)$, we can write this as $(x_u - x_v)^2 \in (1\pm\epsilon/7)\cdot(x_u^{(i)} - x_v^{(i)})^2$, which completes the proof of Lemma IV.2. □

References

[1] Y. Takai, A. Miyauchi, M. Ikeda, and Y. Yoshida, "Hypergraph clustering based on PageRank," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 2020, pp. 1970–1978.

[2] M. Hein, S. Setzer, L. Jost, and S. S. Rangapuram, "The total variation on hypergraphs - learning on hypergraphs revisited," in *Advances in Neural Information Processing Systems 26 (NIPS)*, 2013, pp. 2427–2435.

[3] N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, and P. P. Talukdar, "HyperGCN: A new method for training graph convolutional networks on hypergraphs," in *Advances in Neural Information Processing Systems 32*, 2019, pp. 1509–1520.

[4] C. Zhang, S. Hu, Z. G. Tang, and T.-H. H. Chan, "Re-revisiting learning on hypergraphs: Confidence interval, subgradient method, and extension to multiclass," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 506–518, 2020.

[5] N. Yadati, V. Nitin, M. Nimishakavi, P. Yadav, A. Louis, and P. Talukdar, "NHP: Neural hypergraph link prediction," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM)*, 2020, pp. 1705–1714.

[6] T. Soma and Y. Yoshida, "Spectral sparsification of hypergraphs," in *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, jan 2019, pp. 2570–2581.

[7] D. A. Spielman and S.-H. Teng, "Spectral sparsification of graphs," *SIAM Journal on Computing*, vol. 40, no. 4, pp. 981–1025, 2011.

[8] N. Bansal, O. Svensson, and L. Trevisan, "New notions and constructions of sparsification for graphs and hypergraphs," in *Proceedings of the IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, 2019, pp. 910–928.

[9] M. Kapralov, R. Krauthgamer, J. Tardos, and Y. Yoshida, "Towards tight bounds for spectral sparsification of hypergraphs," in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2021. ACM, 2021, pp. 598–611.

[10] Y. Chen, S. Khanna, and A. Nagda, "Near-linear size hypergraph cut sparsifiers," in *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, 2020, pp. 61–72.

[11] A. A. Benczúr and D. R. Karger, "Randomized approximation schemes for cuts and flows in capacitated graphs," *SIAM Journal on Computing*, vol. 44, no. 2, pp. 290–319, 2015.

[12] D. Kogan and R. Krauthgamer, "Sketching cuts in graphs and hypergraphs," in *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS)*, 2015, pp. 367–376.

[13] I. Newman and Y. Rabinovich, "On multiplicative λ-approximations and some geometric applications," *SIAM Journal on Computing*, vol. 42, no. 3, pp. 855–883, 2013.

[14] A. Andoni, J. Chen, R. Krauthgamer, B. Qin, D. P. Woodruff, and Q. Zhang, "On sketching quadratic forms," in *Proceedings of the 2016 Conference on Innovations in Theoretical Computer Science (ITCS)*, 2016, pp. 311–319.

[15] C. Carlson, A. Kolla, N. Srivastava, and L. Trevisan, "Optimal lower bounds for sketching graph cuts," in *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2019, pp. 2565–2569.

[16] J. D. Batson, D. A. Spielman, and N. Srivastava, "Twice-Ramanujan sparsifiers," *SIAM Journal on Computing*, vol. 41, no. 6, pp. 1704–1721, 2012. [Online]. Available: https://doi.org/10.1137/090772873

[17] D. A. Spielman and N. Srivastava, "Graph sparsification by effective resistances," *SIAM Journal on Computing*, vol. 40, no. 6, pp. 1913–1926, 2011.

[18] Z. A. Zhu, Z. Liao, and L. Orecchia, "Spectral sparsification and regret minimization beyond

matrix multiplicative updates," in *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015*. ACM, 2015, pp. 237–245. [Online]. Available: https://doi.org/10.1145/2746539.2746610

[19] Y. T. Lee and H. Sun, "Constructing linear-sized spectral sparsification in almost-linear time," in *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015*. IEEE Computer Society, 2015, pp. 250–269. [Online]. Available: https://doi.org/10.1109/FOCS.2015.24

[20] Y. T. Lee and H. Sun, "An sdp-based algorithm for linear-sized spectral sparsification," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017*. ACM, 2017, pp. 678–687. [Online]. Available: https://doi.org/10.1145/3055399.3055477

[21] N. K. Vishnoi, "Lx = b," *Foundations and Trends® in Theoretical Computer Science*, vol. 8, no. 1–2, pp. 1–141, 2013.

[22] S.-H. Teng, "Scalable algorithms for data and network analysis," *Foundations and Trends® in Theoretical Computer Science*, vol. 12, no. 1-2, pp. 1–274, 2016.

[23] A. Louis, "Hypergraph Markov operators, eigenvalues and approximation algorithms," in *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing (STOC)*, 2015, pp. 713–722.

[24] T.-H. H. Chan, A. Louis, Z. G. Tang, and C. Zhang, "Spectral properties of hypergraph Laplacian and approximation algorithms," *Journal of the ACM*, vol. 65, no. 3, pp. 1–48, 2018.

[25] Y. Yoshida, "Cheeger inequalities for submodular transformations," in *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2019, pp. 2582–2601.

[26] M. Ikeda, A. Miyauchi, Y. Takai, and Y. Yoshida, "Finding Cheeger cuts in hypergraphs via heat equation," 2019.

[27] K. Fujii, T. Soma, and Y. Yoshida, "Polynomial-time algorithms for submodular Laplacian systems," 2018.

[28] M. Ikeda, Y. Kitabeppu, and Y. Takai, "Coarse Ricci curvature of hypergraphs and its generalization," 2021.

[29] Y. Komura, "Nonlinear semi-groups in Hilbert space," *Journal of the Mathematical Society of Japan*, vol. 19, no. 4, pp. 493–507, 1967.

[30] I. Miyadera, *Nonlinear Semigroups*. American Mathematical Soc., 1992, vol. 109.

[31] Y. Yoshida, "Nonlinear Laplacian for digraphs and its applications to network analysis," in *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM)*, 2016, pp. 483–492.

[32] P. Li and O. Milenkovic, "Submodular hypergraphs: p-Laplacians, Cheeger inequalities and spectral clustering," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, vol. 80, 2018, pp. 3020–3029.

[33] N. Alon and J. H. Spencer, *The Probabilistic Method, Third Edition*, ser. Wiley-Interscience series in discrete mathematics and optimization. Wiley, 2008.

[34] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, pp. 389–434, 2011.

[35] M. Rudelson and R. Vershynin, "Sampling from large matrices: An approach through geometric functional analysis," *Journal of the ACM*, vol. 54, no. 4, p. 21, 2007.

[36] N. Alon and B. Klartag, "Optimal compression of approximate inner products and dimension reduction," in *Proceedings of the IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, 2017, pp. 639–650.

[37] L. Lovász, "Random walks on graphs: A survey," in *Combinatorics, Paul ErdHos is eighty. Vol. 2*, D. Miklós, V. T. Sós, and T. Szönyi, Eds. János Bolyai Mathematical Society, 1993, pp. 353–397. [Online]. Available: http://bolyai.math.elte.hu/~lovasz/erdos.pdf