Spectral hypergraph sparsification via chaining

James R. Lee*

Abstract

In a hypergraph on n vertices where D is the maximum size of a hyperedge, there is a weighted hypergraph spectral ε-sparsifier with at most $O(\varepsilon^{-2}\log(D)\cdot n\log n)$ hyperedges. This improves over the bound of Kapralov, Krauthgamer, Tardos and Yoshida (2021) who achieve $O(\varepsilon^{-4}n(\log n)^3)$, as well as the bound $O(\varepsilon^{-2}D^3 n\log n)$ obtained by Bansal, Svensson, and Trevisan (2019). The same sparsification result was obtained independently by Jambulapati, Liu, and Sidford (2022).

Contents

1 Introduction

Consider a weighted hypergraph $H=(V,E,w)$ with $w\in\mathbb{R}_+^E$ and the corresponding energy: For $x\in\mathbb{R}^V$,

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

The problem of minimizing the energy Q_H over various convex bodies occurs in many applied contexts, especially in machine learning; we refer to the discussion in [KKTY21a].

*Computer Science & Engineering, University of Washington. `jrl@cs.washington.edu`

In the graph case—when all the hyperedges have cardinality 2—this corresponds to the quadratic form associated to the weighted Laplacian and carries a physical interpretation as the potential energy of a family of springs indexed by $\{u, v\} \in E$ whose respective endpoints are pinned at x_u and x_v. Let us mention the appealing analog for hypergraphs: If we stretch a rubber band around vertices pinned at locations $\{x_u : u \in e\}$, then $\max_{\{u,v\} \in \binom{e}{2}} (x_u - x_v)^2$ is proportional to its potential energy. Here the weight w_e represents the elasticity of the band.

For hypergraphs, the edge set E could have cardinality as large $2^{|V|}$, and one can ask if there is a substantially smaller hypergraph that approximates the energy for every configuration of vertices. Soma and Yoshida [SY19] formalized the following notion of spectral sparsification for hypergraphs, generalizing the well-studied notion for graphs [ST11]. Say that a weighted hypergraph $\tilde{H} = (V, \tilde{E}, \tilde{w})$ is a *spectral ε-sparsifier* for H if $\tilde{E} \subseteq E$, and

$$|Q_H(x) - Q_{\tilde{H}}(x)| \leqslant \varepsilon Q_H(x), \qquad \forall x \in \mathbb{R}^V . \tag{1.1}$$

We will use $n := |V|$ throughout. The authors [SY19] showed that one can always find a spectral ε-sparsifier $\tilde{H}$ with $|\tilde{E}| \leqslant O(n^3/\varepsilon^2)$. In [BST19], the authors established a bound of $O(\varepsilon^{-2} D^3 n \log n)$, where $D := \max\{|e| : e \in E\}$ is often called the *rank of H*, and subsequently the authors of [KKTY21b] achieved an upper bound of $nD(\varepsilon^{-1} \log n)^{O(1)}$.

Finally, in a recent and remarkable breakthrough, the authors of [KKTY21a] show that one can obtain a spectral sparsifier with at most $O(n(\log n)^3/\varepsilon^4)$ hyperedges, bypassing the polynomial dependence on the rank, and coming within $\mathrm{poly}(\varepsilon^{-1} \log n)$ factors of the optimal bound. By refining their approach via Talagrand's powerful generic chaining theory, we obtain the following improvement.

Theorem 1.1. *For any n-vertex weighted hypergraph $H = (V, E, w)$ and $\varepsilon > 0$, there is a spectral ε-sparsifier $\tilde{H} = (V, \tilde{E}, \tilde{w})$ for H with*

$$|\tilde{E}| \leqslant O\left(\frac{\log D}{\varepsilon^2} n \log n\right),$$

where $D := \max_{e \in E} |e|$.

As in many prior works, Theorem 1.1 is proved by defining a distribution on E and then sampling edges independently from this distribution. For approaches based on independent sampling, the bound of Theorem 1.1 is tight up to a constant factor for every fixed D. In particular, this generalizes the analysis of independent random sampling for graph sparsifiers [SS11] where $D = 2$.

It should be noted that for *cut sparsifiers*, the $\log D$ factor can be removed [CKN20]. This corresponds to the weaker notion where we only require that (1.1) holds for $x \in \{-1, 1\}^V$. Whether the $\log D$ factor can be removed in general remains an intriguing open question.

Our proof of Theorem 1.1 entails an algorithm for constructing the sparsifier $\tilde{H}$ whose running time is polynomial in the size of the input. But our sampling analysis can also be applied directly to the faster algorithm presented in [KKTY21a] whose running time is $|E| D\, \mathrm{poly}(\log |E|) + \mathrm{poly}(n)$.

Theorem 1.1 was proved independently and concurrently by Jambulapati, Liu, and Sidford [JLS22], via a closely related approach. While their main chaining result is somewhat less general than the one proved here (see (1.5) below), they also present a near-linear time algorithm for generating suitable sampling probabilities $\{\mu_e : e \in E\}$. This improves the running time to $|E| D\, \mathrm{poly}(\log |E|)$.

1.1 The random selector method and chaining for subgaussian processes

Suppose we have a probability distribution $\mu \in (0,1]^E$ on hyperedges in H. We sample hyperedges $\tilde{E} = \{e_1, e_2, \ldots, e_M\}$ independently according to μ, and define the random weighted hypergraph $\tilde{H} = (V, \tilde{E}, \tilde{w})$ so that

$$Q_{\tilde{H}}(x) = \frac{1}{M} \sum_{k=1}^{M} \frac{w_{e_k}}{\mu_{e_k}} Q_{e_k}(x)\,,$$

where we define

$$Q_e(x) := \max_{\{i,j\} \in \binom{e}{2}} (x_i - x_j)^2\,,$$

and the edge weights

$$\tilde{w}_e := \frac{\#\{k \in [M] : e_k = e\}}{M} \cdot \frac{w_e}{\mu_e}\,. \tag{1.2}$$

In particular, this gives, for all $x \in \mathbb{R}^n$,

$$\mathbb{E}[Q_{\tilde{H}}(x)] = \mathbb{E}\left[\frac{w_{e_1}}{\mu_{e_1}} Q_{e_k}(x)\right] = \sum_{e \in E} \mu_e \frac{w_e}{\mu_e} Q_{e_k}(x) = \sum_{e \in E} Q_{e_k}(x) = Q_H(x)\,.$$

Now in order to find a spectral ε-sparsifier, we want to choose M sufficiently large so that

$$\mathbb{E} \max_{x : Q_H(x) \leqslant 1} \left|Q_H(x) - Q_{\tilde{H}}(x)\right| \leqslant \varepsilon\,.$$

To control concentration of $Q_{\tilde{H}}(x)$ around its mean, it suffices to bound the average maximal fluctuations. Thus by a standard sort of reduction (see Section 3.1 and also [Tal14, Lem 9.1.11] for a general formulation), it suffices to prove that

$$\mathop{\mathbb{E}}_{\tilde{H}}\left[\mathop{\mathbb{E}}_{\varepsilon_1,\ldots,\varepsilon_M} \max_{x : Q_H(x) \leqslant 1} \sum_{k=1}^{M} \varepsilon_k \frac{w_{e_k}}{\mu_{e_k}} Q_{e_k}(x)\right] \leqslant O(\varepsilon M)\,, \tag{1.3}$$

where $\varepsilon_1, \ldots, \varepsilon_M \in \{-1, 1\}$ are i.i.d. random signs.

Thus our task is now to control the quantity in brackets in (1.3) for a *fixed* choice of hyperedges $e_1, \ldots, e_M$. If we define the random variable

$$V_x := \sum_{k=1}^{M} \varepsilon_k \frac{w_{e_k}}{\mu_{e_k}} Q_{e_k}(x)\,,$$

then $\{V_x : x \in \mathbb{R}^n\}$ is a subgaussian process (defined in (2.1)) with respect to the (semi)metric

$$d(x, \hat{x}) := \left(\sum_{k=1}^{M} \left(\frac{w_{e_k}}{\mu_{e_k}}\right)^2 \left|Q_{e_k}(x) - Q_{e_k}(\hat{x})\right|^2\right)^{1/2}.$$

There are well-developed tools for studying quantities like $\mathbb{E}\max\{V_x : Q_H(x) \leqslant 1\}$, but they rely on an understanding of the geometry of the space $(\mathbb{R}^n, d)$, and a correct choice of distribution μ is essential for making this geometry well-behaved.

Importance sampling. For spectral graph sparsification, one chooses the sampling probability μ_e to be proportional to the effective resistance across e [SS11]. In order to extend this to hypergraphs, the authors of [BST19] define sampling probabilities $\{\mu_e : e \in E\}$ derived from the graph $G = (V, F)$, where $F := \bigcup_{e \in E} \binom{e}{2}$ is a union of cliques on every hyperedge. They take

$$\mu_e \propto \sum_{\{u,v\} \in \binom{e}{2}} \mathsf{R}_{uv} \,,$$

where R_{uv} denotes the effective resistance between a pair of vertices u, v in G.

To remove the polynomial dependence on D, the authors of [KKTY21a] choose a *weighted graph* $G = (V, F, c)$ and define

$$\mu_e \propto w_e \max\left\{\mathsf{R}_{uv} : \{u, v\} \in \tbinom{e}{2}\right\}.$$

Now R_{uv} is the effective resistance in G, where edges $\{u, v\} \in F$ have conductance c_{uv}.

Let L_G denote the corresponding (weighted) graph Laplacian, and use L_G^+ to denote its pseudoinverse. Define $T := \{v \in \mathbb{R}^n : Q_H(L_G^{+/2} v) \leqslant 1\}$. This construction of the sampling probabilities allows us to write

$$\mathbb{E} \max_{Q_H(x) \leqslant 1} V_x = \mathbb{E} \max_{v \in T} \sum_{k=1}^{M} \varepsilon_k \max_{\{i,j\} \in e_k} \langle v, y_{ij}^{e_k} \rangle^2 \,, \tag{1.4}$$

for a family of vectors $\{y_{ij}^{e_k}\}$ that depends on our choice of edge conductances $c \in \mathbb{R}_+^F$ in G.

A central component of this approach is the existence of conductances that ensure two key properties:

1. $T \subseteq B_2^n := \{x \in \mathbb{R}^n : \|x\| \leqslant 1\}$,
2. $\|y_{ij}^{e_k}\| \leqslant O(\sqrt{n})$ for all $k = 1, \ldots, M$ and $\{i, j\} \in \binom{e_k}{2}$.

We return to a discussion of these properties in a moment.

Chaining bounds. Note that the right-hand side of (1.4) can be written as

$$\mathbb{E} \max_{v \in T} \sum_{k=1}^{M} \varepsilon_k N_k(v)^2,$$

where N_k is an ℓ_∞ norm on a subset of the coordinates of Av, and A is a matrix whose rows are the vectors $\{y_{ij}^{e_k}\}$. Thus in Section 2, we apply aspects of the generic chaining theory (see the extensive reference [Tal14]) to the analysis of such expected maxima.

For readers familiar with the theory, let us note that a bound of $|\tilde{E}| \leqslant O(\varepsilon^{-2} n (\log n)^3)$ in Theorem 1.1 follows from applying Dudley's entropy bound (cf. (2.4)) in a straightforward way. A bound of $|\tilde{E}| \leqslant O(\varepsilon^{-2} n (\log n)^2)$ follows from a deeper inequality of Talagrand (see Theorem 2.2 and Section 2.2) that exploits property (1) above, that T is a subset of the Euclidean unit ball.

Finally, in order to achieve $|\tilde{E}| \leqslant O(\varepsilon^{-2} \log(D) \cdot n \log n)$, we need to exploit further structure of the norms $\{N_k\}$ in a novel way. Our approach is modeled after Rudelson's geometric argument [Rud99a] which, roughly speaking, handles the case where each N_k is a 1-dimensional norm, as well as Talagrand's method of chaining via growth functionals (see Section 2.3 and Section 2.4).

To state this bound, let us consider arbitrary norms $N_1, \ldots, N_M$ on $\mathbb{R}^n$. Define:

$$\kappa := \mathbb{E} \max_{k \in [M]} N_k(g)\,,$$
$$\lambda := \max_{k \in [M]} \left(\mathbb{E}[N_k(g)^2]\right)^{1/2}\,,$$

where g is a standard n-dimensional Gaussian. In Section 2.4, we prove that for any $T \subseteq B_2^n$,

$$\mathbb{E} \sup_{x \in T} \sum_{k=1}^{M} \varepsilon_k N_k(x)^2 \lesssim \sup_{x \in T} \left(\sum_{j=1}^{M} N_j(x)^4\right)^{1/2} + \left(\lambda\sqrt{\log n} + \kappa\right) \cdot \sup_{x \in T} \left(\sum_{k=1}^{M} N_k(x)^2\right)^{1/2} \tag{1.5}$$

When $M = m$, each N_k is a 1-dimensional norm $N_k(x) := |\langle x, a_k\rangle|$ for some $a_k \in \mathbb{R}^n$, and $T = B_2^n$, this lemma recovers Rudelson's concentration bound for Bernoulli sums of rank-1 matrices [Rud99b] (as mentioned there, the inequality we state next is a consequence of the noncommutative Khintchine inequalities [LPP91]).

Observe that $N_k(x)^2 = \langle x, a_k\rangle^2 = \langle x, a_k a_k^* x\rangle$, and using $\|\cdot\|_{op}$ to denote the operator norm, the preceding bound asserts that

$$\mathbb{E}\left\|\sum_{k=1}^{m} \varepsilon_k a_k a_k^*\right\|_{op} = \mathbb{E} \max_{x \in B_2^n} \left\langle x, \left(\sum_{k=1}^{m} \varepsilon_k a_k a_k^*\right) x\right\rangle \leqslant O(\sqrt{\log(m+n)}) \max_{k \in [m]} \|a_k\| \cdot \left\|\sum_{k=1}^{m} a_k a_k^*\right\|_{op}^{1/2},$$

where we use $\lambda \leqslant O(1) \max_{k \in [m]} \|a_k\|$ and $\kappa \leqslant O(\sqrt{\log m}) \max_{k \in [m]} \|a_k\|$.

When applying (1.5) to hypergraph sparsification, one picks up an additional $\sqrt{\log D}$ factor because each N_k is an ℓ_∞ norm on a subset of at most D coordinates.

Remark 1.2. As far as we know, it is an open problem to replicate consequences of the noncommutative Khintchine bound for higher-rank matrices using chaining, i.e., in the setting where $N_k(x) = \|A_k x\|$ for matrices $A_1, \ldots, A_M$.

Choosing good conductances. In order to satisfy properties (1) and (2) above, one chooses nonnegative numbers

$$\left\{c_{ij}^e \geqslant 0 : \{i,j\} \in \tbinom{e}{2}, e \in E\right\}$$

for which

$$\sum_{\{i,j\} \in \binom{e}{2}} c_{ij}^e = w_e\,, \qquad \forall e \in E\,. \tag{1.6}$$

Define the edge conductances $c_{ij} := \sum_{e \in E : \{i,j\} \in \binom{e}{2}} c_{ij}^e$. As argued in Section 3.2, any such choice satisfies property (1).

Let R_{ij} denote the effective resistance between $\{i,j\} \in F$ in the weighted graph $G = (V, F, c)$. To satisfy property (2), it suffices that for all hyperedges $e \in E$, the effective resistances R_{ij} are the same for all pairs $\{i,j\} \in \binom{e}{2}$ with $c_{ij}^e > 0$. (This continues to hold even if the resistances are only comparable up to universal constant factors.)

Let J denote the all-ones matrix and consider maximizing the quantity

$$\log\det(L_G + J)$$

over all choices of (c_{ij}^e) satisfying (1.6). This quantity is a concave function of the conductances (c_{ij}^e) and the KKT conditions for the maximizer establish the desired property for the effective resistances. See Section 3.3.

This is essentially a reformulation and simplification of the method used in [KKTY21a] for establishing the existence of nice conductances $c : F \to \mathbb{R}_+$. It is also reminiscent of Barthe's method for analyzing the Gaussian maximizers of the Brascamp-Lieb (and reverse Brascamp-Lieb) inequalities [Bar98] (see also the treatment in [HM13]).

1.2 Notation

For two expressions A and B, we will use the equivalent notations $A \lesssim B$ and $A \leqslant O(B)$ to denote that there is a constant $C > 0$ such that $A \leqslant CB$. If A and B depend on some parameters $\alpha_1, \alpha_2, \ldots$, we use the notation $A \lesssim_{\alpha_1,\alpha_2,\ldots} B$ to denote that there is a number $C = C(\alpha_1, \alpha_2, \ldots)$ such that $A \leqslant CB$. We use $A \asymp B$ to denote the conjunction of $A \lesssim B$ and $B \lesssim A$.

A number of vector and matrix norms will appear in what follows. When $x \in \mathbb{R}^n$ is a vector, $\|x\|$ will always refer to the standard Euclidean norm of x. For a positive integer $M \geqslant 1$, we will sometimes use the notation $[M] := \{1, 2, \ldots, M\}$.

2 Extrema of random processes

2.1 Background on generic chaining

A space (T, d) is called a *K-quasimetric* if satisfies

1. $d(x, y) = d(y, x)$ for all $x, y \in T$.
2. $d(x, x) = 0$ for all $x \in T$.
3. There is a constant $K > 0$ such that

$$d(x, y) \leqslant K\left(d(x, z) + d(z, y)\right), \qquad \forall x, y, z \in T.$$

Say that (T, d) is a *quasimetric space* if (T, d) is a K-quasimetric for some $K > 0$.

Consider a distance d on T. A random process $\{V_x : x \in T\}$ is said to be *subgaussian with respect to d* if there is a number $\alpha > 0$ such that

$$\mathbb{P}\left(|V_x - V_y| > t\right) \leqslant \exp\left(-\alpha \frac{t^2}{d(x, y)^2}\right), \qquad t > 0. \tag{2.1}$$

The generic chaining functional. For a quasimetric space (T, d), let us recall Talagrand's generic chaining functional [Tal14, Def. 2.2.19]. Define $N_h := 2^{2^h}$. Then

$$\gamma_2(T, d) := \inf_{\{\mathcal{A}_h\}} \sup_{x \in T} \sum_{h=0}^{\infty} 2^{h/2} \mathrm{diam}_d(\mathcal{A}_h(x)), \tag{2.2}$$

where the infimum runs over all sequences $\{\mathcal{A}_h : h \geqslant 0\}$ of partitions of T satisfying $|\mathcal{A}_h| \leqslant N_h$ for each $h \geqslant 0$. Note that we use the notation $\mathcal{A}_h(x)$ for the unique set of $\mathcal{A}_h$ that contains x, and $\mathrm{diam}_d(S) := \sup_{x,y \in S} d(x,y)$ for $S \subseteq T$. The next theorem constitutes the generic chaining upper bound; see [Tal14, Thm 2.2.18].

Theorem 2.1. *If $\{V_x : x \in T\}$ is a centered subgaussian process satisfying (2.1) with respect to a K-quasimetric (T,d), then*

$$\mathbb{E} \sup_{x \in T} V_x \lesssim_{K,\alpha} \gamma_2(T,d)\,. \tag{2.3}$$

Define the entropy numbers $e_h(T,d) := \inf\{\sup_{t \in T} d(t, T_h) : T_h \subseteq T, |T_h| \leqslant 2^{2^h}\}$. This is the infimum of numbers $r > 0$ such that T can be covered by at most 2^{2^h} balls of radius r. A classical way of controlling $\gamma_2(T,d)$ is given by Dudley's entropy bound (see, e.g., [Tal14, Prop 2.2.10]):

$$\gamma_2(T,d) \lesssim \sum_{h \geqslant 0} 2^{h/2} e_h(T,d)\,. \tag{2.4}$$

But often additional structure of the space (T,d) allows one to improve on (2.4). The next lemma is a consequence of [Tal14, Thm 4.1.11 & (4.23)]. It actually holds whenever T is the unit ball of a uniformly 2-convex Banach space and d is induced by some (possibly different) norm.

Theorem 2.2. *Suppose that $T = B_2^n$ is the unit Euclidean ball in $\mathbb{R}^n$ and $\|\cdot\|_X$ is a norm on $\mathbb{R}^n$. Then,*

$$\gamma_2(T, \|\cdot\|_X) \lesssim \left(\sum_{h \geqslant 0} \left(2^{h/2} e_h(T, \|\cdot\|_X) \right)^2 \right)^{1/2}.$$

In order to bound the entropy numbers $e_h(B_2^n, \|\cdot\|_X)$, we will use the following classical fact; see, e.g., [LT11, (3.15)].

Lemma 2.3 (Dual Sudakov inequality)**.** *Let B_2^n denote the unit Euclidean ball, and suppose that $\|\cdot\|_X$ is a norm on $\mathbb{R}^n$. Then*

$$e_h(B_2^n, \|\cdot\|_X) \lesssim 2^{-h/2}\, \mathbb{E}\, \|g\|_X,$$

where g is a standard n-dimensional Gaussian.

Corollary 2.4. *If $\|\cdot\|_X$ is a norm on $\mathbb{R}^n$, then*

$$\gamma_2(B_2^n, \|\cdot\|_X) \lesssim \mathrm{diam}(B_2^n, \|\cdot\|_X) + \sqrt{\log n}\, \mathbb{E}\, \|g\|_X\,,$$

where g is a standard n-dimensional Gaussian.

Proof. A straightforward volume argument shows that any set of δ-separated points in $(B_2^n, \|\cdot\|)$ must have cardinality at most $(4/\delta)^n$, and therefore

$$e_h(T, \|\cdot\|) \leqslant 4 \cdot N_h^{-1/n} = 4 \cdot 2^{-2^h/n}\,.$$

Taking $L := \mathrm{diam}(B_2^n, \|\cdot\|_X)$, we have $e_h(B_2^n, \|\cdot\|_X) \leqslant L \cdot e_h(B_2^n, \|\cdot\|)$, and therefore

$$e_h(B_2^n, \|\cdot\|_X) \leqslant 4L \cdot (2^{-2^h/n}).$$

Denote $S := \sup_{h \geqslant 0} 2^{h/2} e_h(T, \|\cdot\|_X)$. Applying Theorem 2.2 yields, for any $h_0 \geqslant 0$,

$$\gamma_2(T, d) \lesssim S\sqrt{h_0} + 4L \left(\sum_{h \geqslant h_0} (2^{h/2} 2^{-2^h/n})^2 \right)^{1/2} .$$

Choosing $h_0 \geqslant 2 \log n$ bounds the latter sum by $O(1)$, yielding

$$\gamma_2(T, d) \lesssim S\sqrt{\log n} + L .$$

To conclude, use Lemma 2.3 to bound S. □

2.2 Warm up

The next lemma will allow us to establish the existence of hypergraph spectral sparsifiers with at most $O(\varepsilon^{-2} n (\log n)^2)$ hyperedges. It also provides a nice warm up for the more delicate arguments in Section 2.4.

Let $A : \mathbb{R}^n \to \mathbb{R}^m$ denote a linear operator. We use the notation

$$\|A\|_{2\to\infty} := \max_{\|x\| \leqslant 1} \|Ax\|_\infty .$$

This is equal to the maximum ℓ_2 norm of a row of A. Define the norm

$$\|x\|_A := \|Ax\|_\infty ,$$

and let us observe the following.

Lemma 2.5. *If g is a standard n-dimensional Gaussian, it holds that*

$$\mathbb{E}\, \|g\|_A \lesssim \|A\|_{2\to\infty} \sqrt{\log m} .$$

In particular, Lemma 2.3 gives

$$e_h(B_2^n, \|\cdot\|_A) \lesssim 2^{-h/2} \sqrt{\log m} \|A\|_{2\to\infty} .$$

Proof. If $a_1, \ldots, a_m$ are the rows of A and g is an n-dimensional Gaussian, then

$$\mathbb{E}\, \|Ag\|_\infty = \mathbb{E} \max_{i \in [m]} |\langle g, a_i \rangle| \lesssim \max_{i \in [m]} \|a_i\| \sqrt{\log m} = \|A\|_{2\to\infty} \sqrt{\log m} .$$ □

Additionally, let $\varphi_1, \varphi_2, \ldots, \varphi_M : \mathbb{R}^m \to \mathbb{R}$ be arbitrary functions.

Lemma 2.6. *For any subset $T \subseteq B_2^n$, it holds that*

$$\mathbb{E} \sup_{x \in T} \sum_{j=1}^M \varepsilon_j \varphi_j(Ax)^2 \lesssim \sqrt{\log m \log n}\, \|A\|_{2\to\infty} \cdot \sup_{\substack{j \in [M], \\ \|z - z'\|_\infty \leqslant 1}} |\varphi_j(z) - \varphi_j(z')| \cdot \sup_{x \in T} \left(\sum_{j=1}^M \varphi_j(Ax)^2 \right)^{1/2} ,$$

where $\varepsilon_1, \ldots, \varepsilon_M$ are i.i.d. Bernoulli ± 1 random variables.

Proof. Define

$$\alpha := \max_{j \in [M]} \sup_{\|z-z'\|_\infty \leqslant 1} |\varphi_j(z) - \varphi_j(z')|\,, \tag{2.5}$$

$$\beta := \sup_{x \in T} \left(\sum_{j=1}^{M} \varphi_j(Ax)^2 \right)^{1/2}, \tag{2.6}$$

$$V_x := \sum_{j=1}^{M} \varepsilon_j \varphi_j(Ax)^2\,,$$

and note that $\{V_x : x \in \mathbb{R}^n\}$ is a subgaussian process with respect to the distance

$$d(x,\hat{x}) := \left(\sum_{j=1}^{M} \left| \varphi_j(Ax)^2 - \varphi_j(A\hat{x})^2 \right|^2 \right)^{1/2}.$$

Thus in light of (2.3), it suffices to prove that

$$\gamma_2(T,d) \lesssim \sqrt{\log m \log n}\, \|A\|_{2\to\infty} \cdot \alpha\beta\,. \tag{2.7}$$

Note that for $x, \hat{x} \in T$,

$$\begin{aligned} d(x,\hat{x})^2 &= \sum_{j=1}^{M} \left(\varphi_j(Ax) - \varphi_j(A\hat{x})\right)^2 \left(\varphi_j(Ax) + \varphi_j(A\hat{x})\right)^2 \\ &\leqslant 2 \sum_{j=1}^{M} \left(\varphi_j(Ax) - \varphi_j(A\hat{x})\right)^2 \left(\varphi_j(Ax)^2 + \varphi_j(A\hat{x})^2\right) \\ &\overset{(2.5)}{\leqslant} 2\alpha^2 \|A(x-\hat{x})\|_\infty^2 \sum_{j=1}^{M} \left(\varphi_j(Ax)^2 + \varphi_j(A\hat{x})^2\right) \\ &\overset{(2.6)}{\leqslant} 4\alpha^2\beta^2 \|x-\hat{x}\|_A^2\,. \end{aligned} \tag{2.8}$$

In particular, we have

$$\gamma_2(T,d) \leqslant 2\alpha\beta \cdot \gamma_2(T, \|\cdot\|_A) \leqslant 2\alpha\beta \cdot \gamma_2(B_2^n, \|\cdot\|_A), \tag{2.9}$$

where the last inequality uses $T \subseteq B_2^n$.

Noting that $\mathrm{diam}(B_2^n, \|\cdot\|_A) \leqslant 2\|A\|_{2\to\infty}$, we can thus apply Lemma 2.5 and Corollary 2.4 with $\|\cdot\|_X = \|\cdot\|_A$ to conclude that

$$\gamma_2(B_2^n, \|\cdot\|_A) \lesssim \|A\|_{2\to\infty} \sqrt{\log m \log n}\,.$$

Combining this with (2.9) completes our verification of (2.7). □

In Section 2.4, we will obtain an improved bound by using convexity in a stronger way. In particular, we will assume that each of the functions φ_j in Lemma 2.6 is a norm on $\mathbb{R}^m$.

2.3 Growth functionals

Talagrand introduced a powerful way to control $\gamma_2(T, d)$ via the existence of certain growth functionals. For $x \in T$ and $\rho > 0$, define the ball

$$B_d(x, \rho) := \{y \in T : d(x, y) \leqslant \rho\}. \tag{2.10}$$

Definition 2.7 (Separated sets)**.** Let (T, d) denote a metric space and consider numbers $a > 0, r \geqslant 4$. Say that subsets $H_1, \ldots, H_m \subseteq T$ *are* (a, r)*-separated* if

$$H_\ell \subseteq B_d(x_\ell, a/r), \quad \ell = 1, \ldots, m,$$

where $x_1, \ldots, x_m \in T$ are points satisfying

$$a \leqslant d(x_\ell, x_{\ell'}) \leqslant ar, \quad \forall \ell \neq \ell'. \tag{2.11}$$

Definition 2.8 (The growth condition)**.** Consider nonnegative functionals $\{F_h : h \geqslant 0\}$ defined on subsets of a metric space (T, d) and which satisfy the following two conditions for every $h \geqslant 0$:

$$\begin{aligned} F_h(S) &\leqslant F_h(S'), \qquad \forall S \subseteq S' \subseteq T, \\ F_{h+1}(S) &\leqslant F_h(S), \qquad \forall\, S \subseteq T. \end{aligned}$$

Say that such functionals satisfy the *growth condition with parameters* $r \geqslant 4$ *and* $c^* > 0$ if for any integer $h \geqslant 0$ and $a > 0$, the following holds true with $m = N_{h+1}$: For each collection of subsets $H_1, \ldots, H_m \subseteq T$ that are (a, r)-separated, we have

$$F_h\left(\bigcup_{\ell \leqslant m} H_\ell\right) \geqslant c^* a 2^{h/2} + \min_{\ell \leqslant m} F_{h+1}(H_\ell). \tag{2.12}$$

Theorem 2.9 ([Tal14, Thm 2.3.16])**.** *Let (T, d) be a K-quasimetric space and consider a sequence of functionals $\{F_h\}$ satisfying the growth condition (cf. Definition 2.8) with parameters $r \geqslant 4$ and $c^* > 0$. Then,*

$$\gamma_2(T, d) \lesssim_K \frac{r}{c^*} F_0(T) + r \cdot \mathrm{diam}_d(T).$$

Remark 2.10 (Packing/covering duality)**.** For the reader encountering Definition 2.8 and Theorem 2.9 for the first time, the role of the functionals $\{F_h\}$ might appear mysterious. Some intuition can be gained by considering the duality between covering and packing: A set S in some metric space can be covered by m balls of radius $r > 0$ if it is impossible to find m points in S that are pairwise separated by distance r.

The quantity $\gamma_2(T, d)$ (cf. (2.2)) is a sort of multiscale covering functional. The growth functionals $\{F_h\}$ measure the "size" of packings of various cardinalities, and (2.12) asserts a form of packing impossibility. This makes Theorem 2.9 a multiscale analog of the simple packing/covering argument recalled above.

Those familiar with convex optimization and duality may find the approach of [BDOS21] instructive in this regard. It is shown that the corresponding *fractional* multiscale covering and packing values are equal by convex duality, and then a rounding argument establishes that the integral versions are equivalent up to constant factors.

We will use the following corollary of Theorem 2.9 that simplifies the construction of functionals if we have a bound on the growth rate of nets in (T, d).

Corollary 2.11. *Let (T, d) be a K-quasimetric and assume there are numbers $k, L \geqslant 1$ and $r \geqslant 4$ such that that for every $a > 0$,*

$$H_1, \ldots, H_m \subseteq T \text{ are } (a, r)\text{-separated} \implies m \leqslant \left(\frac{L}{a}\right)^k . \tag{2.13}$$

Let h_0 be the largest integer $h \geqslant 0$ such that

$$2^{2^h} \leqslant (2L)^{k(h-1)/2} . \tag{2.14}$$

Consider a sequence of functionals $\{F_0, F_1, \ldots, F_{h_0}\}$ satisfying the growth condition (2.12) with parameters r and $c^* > 0$. Then,

$$\gamma_2(T, d) \lesssim_K \frac{r}{c^*} F_0(T) + r \cdot \operatorname{diam}_d(T) \tag{2.15}$$

Proof. Define the numbers

$$c_j := c^* L \cdot 2^{-2^j/k} 2^{(j-1)/2}$$
$$C_0 := \sum_{j=h_0+1}^{\infty} c_j ,$$

and note that $C_0 \lesssim c^*$, since (2.14) is violated for every $h \geqslant h_0 + 1$.

Define a new family of functionals $\{\tilde{F}_h : h \geqslant 0\}$ so that for every $S \subseteq T$,

$$\tilde{F}_h(S) := F_h(S) + C_0 , \qquad h = 0, 1, \ldots, h_0 ,$$
$$\tilde{F}_h(S) := F_{h_0}(S) + C_0 - \sum_{j=h_0+1}^{h} c_j , \qquad h > h_0 .$$

By construction, these satisfy the growth condition Definition 2.8 since for $h \geqslant h_0$, if $H_1, \ldots, H_m \subseteq T$ are (a, r)-separated sets with $m = 2^{2^{h+1}}$, then

$$\tilde{F}_{h+1}\left(\bigcup_{\ell \leqslant m} H_\ell\right) \geqslant c_{h+1} + \tilde{F}_h\left(\bigcup_{\ell \leqslant m} H_\ell\right) \geqslant c_{h+1} + \min_{\ell \leqslant m} \tilde{F}_h(H_\ell) \geqslant c^* a 2^{h/2} + \min_{\ell \leqslant m} \tilde{F}_h(H_\ell) ,$$

where the last inequality uses the fact that $a \leqslant L 2^{-2^{h+1}/k}$ from (2.13). Moreover, we have

$$\tilde{F}_0(T) = F_0(T) + C_0 \leqslant F_0(T) + O(c^*) ,$$

and therefore we can apply Theorem 2.9 to $\{\tilde{F}_h\}$ to complete the proof. □

2.4 Further exploiting convexity

We will now use the growth functional approach (cf. Section 2.3) to prove a more elaborate upper bound under the additional assumption that our summands are derived from norms. This will allow us in Section 3 to find spectral ε-sparsifiers with $O\left(\frac{\log D}{\varepsilon^2} n \log n\right)$ hyperedges.

Let $N_1, N_2, \ldots, N_M$ be norms on $\mathbb{R}^n$ and define

$$\kappa := \mathbb{E} \max_{j \in [M]} N_j(g)\,,$$
$$\lambda := \max_{j \in [M]} \left(\mathbb{E}[N_j(g)^2]\right)^{1/2}\,,$$

where g is a standard n-dimensional Gaussian.

Lemma 2.12. *For any $T \subseteq B_2^n$, it holds that*

$$\mathbb{E} \sup_{x \in T} \sum_{j=1}^{M} \varepsilon_j N_j(x)^2 \lesssim \sup_{x \in T} \left(\sum_{j=1}^{M} N_j(x)^4\right)^{1/2} + \left(\lambda \sqrt{\log n} + \kappa\right) \cdot \sup_{x \in T} \left(\sum_{j=1}^{M} N_j(x)^2\right)^{1/2}\,,$$

where $\varepsilon_1, \ldots, \varepsilon_M$ are i.i.d. Bernoulli ± 1 random variables.

Before proving the lemma, let us illustrate a corollary that we will use to construct hypergraph sparsifiers. Consider a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$, and suppose that each N_i is a (weighted) ℓ_∞ norm on some subset $S_i \subseteq [m]$ of the coordinates:

$$N_i(z) = \max_{j \in S_i} w_j |(Az)_j|\,, \qquad w \in [0,1]^{S_i}\,. \tag{2.16}$$

Let $a_1, \ldots, a_m$ denote the rows of A, and observe that $(Ag)_j = \langle a_j, g\rangle$ is a normal random variable with variance $\|a_j\|^2$, and therefore

$$\mathbb{E}[N_i(g)]^2 = \mathbb{E} \max_{j \in S_i} w_j^2 |\langle a_j, g\rangle|^2 \lesssim \max_{j \in S_i} \|a_j\|^2 \cdot \sqrt{\log |S_i|}\,.$$

Similarly, we have

$$\kappa = \mathbb{E} \max_{i \in [M]} \max_{j \in S_i} w_j |\langle a_j, g\rangle| \leqslant \mathbb{E} \max_{i \in [m]} |\langle a_i, g\rangle| \lesssim \|A\|_{2\to\infty} \sqrt{\log m}\,,$$

and

$$\sum_{j=1}^{M} N_j(x)^4 \leqslant \|A\|_{2\to\infty}^2 \sum_{j=1}^{M} N_j(x)^2\,,$$

yielding the following.

Corollary 2.13. *If the norms $N_1, \ldots, N_M$ are of the form (2.16) for some $A : \mathbb{R}^n \to \mathbb{R}^m$ and subsets $S_1, \ldots, S_M \subseteq [m]$ with $\max_{i \in [M]} |S_i| \leqslant D$, then for any $T \subseteq B_2^n$, it holds that*

$$\mathbb{E} \sup_{x \in T} \sum_{j=1}^{M} \varepsilon_j N_j(x)^2 \lesssim \|A\|_{2\to\infty} \sqrt{\log(m+n) \log D} \cdot \sup_{x \in T} \left(\sum_{j=1}^{M} N_j(x)^2\right)^{1/2}\,,$$

where $\varepsilon_1, \ldots, \varepsilon_M$ are i.i.d. Bernoulli ± 1 random variables.

The proof of Lemma 2.12 is modeled after arguments of Rudelson [Rud99a] and Talagrand; see [Tal14, §16.7] and the historical notes in [Tal14, §16.10]. A version of the latter argument first appeared in [Rud99a], as a simplification of Rudelson's original construction of an explicit majorizing measure. In the proof of [Tal14, Prop 16.7.4], one encounters growth functionals of the form $F(S) = 1 - \inf\{\|u\| : u \in \mathrm{conv}(S)\}$, where $\|\cdot\|$ is a uniformly 2-convex norm. We recall this definition.

Definition 2.14 (Uniform p-convexity)**.** A Banach space Z is called *uniformly p-convex* if there is a number $\eta > 0$ such that for all $x, y \in Z$ with $\|x\|_Z, \|y\|_Z \leqslant 1$,

$$\left\|\frac{x+y}{2}\right\|_Z \leqslant 1 - \eta\|x-y\|_Z^p .$$

We remark that the statement of Lemma 2.12 actually holds when T is a subset of the unit ball of any uniformly 2-convex norm on $\mathbb{R}^n$ (with an implicit constant that depends on η).

We will instead employ functionals of the form

$$F(S) = 2 - \inf\left\{\|u\|^2 + \sum_{j=1}^{M} N_j(u)^2 : u \in \mathrm{conv}(S)\right\}.$$

Problematically, the norm $u \mapsto \left(\|u\|^2 + \sum_{j=1}^M N_j(u)^2\right)^{1/2}$ is potentially very far from uniformly 2-convex, thus we have to be careful in using only 2-convexity of the Euclidean norm, along with 2-convexity of the "outer" ℓ_2 norm of the N_j's. This requires application of the inequality $|N_j(x) - N_j(\hat{x})| \leqslant N_j(x-\hat{x})$ only at judiciously chosen points in the argument. We offer some further explanation in Remark 2.21 after the proof.

Proof of Lemma 2.12. For a set $S \subseteq \mathbb{R}^n$, let $\mathrm{conv}(S)$ denote the closed convex hull of S. Note that by convexity,

$$\sup_{x\in T}\left(\sum_{j=1}^M N_j(x)^2\right)^{1/2} = \sup_{x\in\mathrm{conv}(T)}\left(\sum_{j=1}^M N_j(x)^2\right)^{1/2}.$$

Therefore we may replace T by $\mathrm{conv}(T)$ and henceforth assume that T is compact and convex.

By scaling $\{N_j\}$, we may assume that

$$\sup_{x\in T}\sum_{j=1}^M N_j(x)^2 = 1\,. \tag{2.17}$$

Define $V_x := \sum_{j=1}^M \varepsilon_j N_j(x)^2$. Then $\{V_x : x \in \mathbb{R}^n\}$ is a subgaussian process with respect to the metric

$$\tilde{d}(x,\hat{x}) := \left(\sum_{j=1}^M |N_j(x)^2 - N_j(\hat{x})^2|^2\right)^{1/2},$$

therefore from (2.3), we have

$$\mathbb{E}\sup_{x\in T} V_x \lesssim \gamma_2(T,\tilde{d})\,. \tag{2.18}$$

Passing to a nicer distance. Define the related distance

$$d(x,\hat{x}) := \left(\sum_{j=1}^{M} N_j(x-\hat{x})^2 \left(N_j(x)^2 + N_j(\hat{x})^2\right)\right)^{1/2},$$

and note that for all $x,\hat{x} \in \mathbb{R}^n$,

$$\begin{aligned}\tilde{d}(x,\hat{x})^2 &= \sum_{j=1}^{M} \left(N_j(x) - N_j(\hat{x})\right)^2 \left(N_j(x) + N_j(\hat{x})\right)^2 \\ &\leqslant 2\sum_{j=1}^{M} N_j(x-\hat{x})^2 \left(N_j(x)^2 + N_j(\hat{x})^2\right) = 2\, d(x,\hat{x})^2\,.\end{aligned}$$

We will observe momentarily that

$$d(x,\hat{x}) \leqslant 2\sqrt{2}\,(d(x,y) + d(y,\hat{x}))\,, \qquad \forall x,\hat{x},y \in \mathbb{R}^n\,. \tag{2.19}$$

Since $\tilde{d} \leqslant \sqrt{2}d$ and d is a quasimetric, (2.3) gives

$$\mathbb{E}\sup_{x\in T} V_x \lesssim \gamma_2(T,d)\,,$$

and thus our goal is to establish that

$$\gamma_2(T,d) \lesssim \lambda\sqrt{\log n} + \kappa + \mathrm{diam}_d(T)\,. \tag{2.20}$$

Indeed, note that

$$\mathrm{diam}_d(T) \lesssim \sup_{x\in T}\left(\sum_{j=1}^{M} N_j(x)^4\right)^{1/2},$$

and therefore (2.20) yields our main claim.

Lemma 2.15. *For any metric space (X,D) and $x_0 \in X$, it holds that the distance*

$$\tilde{D}(x,\hat{x}) := D(x,\hat{x})\,(D(x,x_0) + D(\hat{x},x_0))$$

is a 2-quasimetric.

Proof. Define $\psi(x) := D(x,x_0)$ and consider $x,\hat{x},y \in X$. Then,

$$\begin{aligned}\tilde{D}(x,\hat{x}) &\leqslant (D(x,y) + D(\hat{x},y))\,(\psi(x) + \psi(\hat{x})) \\ &\leqslant D(x,y)\,(\psi(x) + \psi(y) + D(\hat{x},y)) + D(\hat{x},y)\,(\psi(\hat{x}) + \psi(y) + D(x,y)) \\ &\leqslant \tilde{D}(x,y) + \tilde{D}(\hat{x},y) + 2D(x,y)D(\hat{x},y)\,.\end{aligned}$$

Now use $2D(x,y)D(\hat{x},y) \leqslant D(x,y)^2 + D(\hat{x},y)^2 \leqslant \tilde{D}(x,y) + \tilde{D}(\hat{x},y)$, completing the proof. ◻

Applying the preceding lemma with $D(x,\hat{x}) = N_j(x-\hat{x})$ and $x_0 = 0$ shows that the distance $(x,\hat{x}) \mapsto N_j(x-\hat{x})(N_j(x)+N_j(\hat{x})^2)^{1/2}$ is a $2\sqrt{2}$-quasimetric for each $j = 1,\ldots,M$, and therefore d is a $2\sqrt{2}$-quasimetric on $\mathbb{R}^n$, verifying (2.19).

Balls in $(\mathbb{R}^n, d)$ are approximately convex. Recall the definition of the balls $B_d(x,\rho)$ from (2.10).

Lemma 2.16. *For any $x \in \mathbb{R}^n$ and $\rho > 0$, it holds that*

$$\operatorname{conv}(B_d(x,\rho)) \subseteq B_d(x,4\rho)\,.$$

Proof. For $y \in B_d(x,\rho)$, we have

$$\left(\sum_{j=1}^{M} N_j(x-y)^2 N_j(x)^2\right)^{1/2} \leqslant \rho\,, \tag{2.21}$$

as well as

$$\sqrt{\rho} \geqslant d(x,y)^{1/2} = \left(\sum_{j=1}^{M} N_j(x-y)^2 \left(N_j(x)^2 + N_j(y)^2\right)\right)^{1/4} \geqslant \left(\frac{1}{2}\sum_{j=1}^{M} N_j(x-y)^4\right)^{1/4}, \tag{2.22}$$

where the final inequality uses $N_j(x-y) \leqslant N_j(x) + N_j(y)$. Since the left-hand side of (2.21) and the right-hand side of (2.22) are both convex functions of y, these inequalities remain true for all $y \in \operatorname{conv}(B_d(x,\rho))$.

In particular, for any $y \in \operatorname{conv}(B_d(x,\rho))$, we can use $a^2 + b^2 \leqslant 4a^2 + 2(a-b)^2$ to write

$$\begin{aligned} d(x,y) &\leqslant \left(\sum_{j=1}^{M} N_j(x-y)^2 \left(4N_j(x)^2 + 2(N_j(x) - N_j(y))^2\right)\right)^{1/2} \\ &\leqslant 2\left(\sum_{j=1}^{M} N_j(x-y)^2 N_j(x)^2\right)^{1/2} + \sqrt{2}\left(\sum_{j=1}^{M} N_j(x-y)^4\right)^{1/2} \leqslant 4\rho\,. \end{aligned}$$

□

Covering estimates. Define now the following norms on $\mathbb{R}^n$:

$$\begin{aligned} \|x\|_{\mathcal{N}} &:= \max_{j\in[M]} N_j(x)\,, \\ \|x\|_{\mathcal{E}(u)} &:= \left(\sum_{j=1}^{M} N_j(x)^2 N_j(u)^2\right)^{1/2}, \quad u \in \mathbb{R}^n\,. \end{aligned}$$

Lemma 2.17. *For all $x,\hat{x},u \in \mathbb{R}^n$,*

$$d(x,\hat{x})^2 \leqslant 2\,\|x-\hat{x}\|_{\mathcal{N}}^2 \left(\sum_{j=1}^{M} \left(N_j(x) - N_j(u)\right)^2 + \sum_{j=1}^{M} \left(N_j(\hat{x}) - N_j(u)\right)^2\right) + 4\|x-\hat{x}\|_{\mathcal{E}(u)}^2\,.$$

Proof. Use the inequalities

$$N_j(x)^2 \leqslant 2(N_j(x) - N_j(u))^2 + 2N_j(u)^2\,, \qquad x, u \in \mathbb{R}^n$$

to write

$$\begin{aligned}\sum_{j=1}^{M} N_j(x-\hat{x})^2 N_j(x)^2 &\leqslant 2\|x-\hat{x}\|_{\mathcal{N}}^2 \sum_{j=1}^{M} \big(N_j(x) - N_j(u)\big)^2 + 2\sum_{j=1}^{M} N_j(x-\hat{x})^2 N_j(u)^2 \\ &= 2\|x-\hat{x}\|_{\mathcal{N}}^2 \sum_{j=1}^{M} \big(N_j(x) - N_j(u)\big)^2 + 2\|x-\hat{x}\|_{\mathcal{E}(u)}^2\,.\end{aligned}$$ □

Lemma 2.18. *It holds that*

$$\begin{aligned}e_h(B_2^n, \|\cdot\|_{\mathcal{N}}) &\lesssim 2^{-h/2}\kappa\,, \\ e_h(B_2^n, \|\cdot\|_{\mathcal{E}(u)}) &\lesssim 2^{-h/2}\lambda\,, \quad \forall u \in T\,.\end{aligned}$$

Proof. Both inequalities follow readily from Lemma 2.3: If g is a standard n-dimensional Gaussian, then

$$e_h(B_2^n, \|\cdot\|_{\mathcal{N}}) \lesssim 2^{-h/2}\,\mathbb{E}\,\|g\|_{\mathcal{N}} = 2^{-h/2}\kappa,$$

by the definition of κ. For the second inequality,

$$e_h(B_2^n, \|\cdot\|_{\mathcal{E}(u)}) \lesssim 2^{-h/2}\,\mathbb{E}\,\|g\|_{\mathcal{E}(u)}\,.$$

Now use convexity of the square to bound

$$\big(\mathbb{E}\,\|g\|_{\mathcal{E}(u)}\big)^2 \leqslant \mathbb{E}\,\|g\|_{\mathcal{E}(u)}^2 = \sum_{j=1}^{M} N_j(u)^2\,\mathbb{E}[N_j(g)^2] \leqslant \lambda^2\,,$$

where the final line uses the definition of λ and $\sum_{j=1}^{M} N_j(u)^2 \leqslant 1$ by (2.17), because $u \in T$. □

We also need a basic estimate that we will use to apply Corollary 2.11. Observe that for $x, \hat{x} \in T$,

$$d(x,\hat{x}) \overset{(2.17)}{\leqslant} \sqrt{2}\|x-\hat{x}\|_{\mathcal{N}} \leqslant \sqrt{2}\,(\|x\|_{\mathcal{N}} + \|\hat{x}\|_{\mathcal{N}}) \leqslant 2\sqrt{2}\,, \tag{2.23}$$

where the last inequality uses $\|x\|_{\mathcal{N}} \leqslant (\sum_{j=1}^{M} N_j(x)^2)^{1/2} \leqslant 1$ for $x \in T$, by (2.17).

Lemma 2.19. *For any $a > 0$, if $x_1, \ldots, x_K \in T$ satisfy $d(x_i, x_j) \geqslant a$ for $i \neq j$, then, $K \leqslant \left(\frac{6}{a}\right)^n$.*

Proof. As noted above, we have $\|x\|_{\mathcal{N}} \leqslant 1$ for $x \in T$, and (2.23) gives $\|x_i - x_j\|_{\mathcal{N}} \geqslant a/\sqrt{2}$ for $i \neq j$. Therefore by a simple volume argument (valid for any norm on $\mathbb{R}^n$):

$$K \leqslant \left(1 + \frac{2\sqrt{2}}{a}\right)^n \leqslant \left(\frac{6}{a}\right)^n,$$

where the last inequality follows because if $K \geqslant 2$, then (2.23) implies $a \leqslant 2\sqrt{2}$. □

The growth functionals. Define a norm on $\mathbb{R}^n$ by

$$\|\!|u|\!\| := \left(\|u\|^2 + \sum_{j=1}^{M} N_j(u)^2\right)^{1/2}. \tag{2.24}$$

Denote $r := 64$. Let h_0 be the largest integer so that $2^{2^{h_0}} \leqslant (12)^{n(h_0-1)/2}$, and note that $h_0 \leqslant O(\log n)$. Define

$$F_h(S) := 2 - \inf\left\{\|\!|u|\!\|^2 : u \in \mathrm{conv}(S)\right\} + \frac{\max(h_0+1-h,0)}{\log n}, \qquad h = 0,1,\ldots,h_0\,. \tag{2.25}$$

Recall that $T \subseteq B_2^n$ and, along with (2.17), this gives $\max_{u\in T} \|\!|u|\!\|^2 \leqslant 2$. Since $h_0 \leqslant O(\log n)$, we have $F_0(T) \leqslant O(1)$.

From Lemma 2.19, it holds that the packing assumption (2.13) is satisfied with $L = 6$ and $k = n$. Therefore if we can verify that our functionals satisfy the growth conditions (2.12) for $h = 0,1,\ldots,h_0$, then we will conclude from (2.15) that

$$\gamma_2(T,d) \lesssim \frac{1}{c^*} + \mathrm{diam}_d(T)\,. \tag{2.26}$$

Consideration of (a,r)-separated sets. Define $K := N_{h+1}$ and consider points $\{x_1,\ldots,x_K\} \subseteq T$ such that $d(x_\ell, x_{\ell'}) \geqslant a$ whenever $\ell \neq \ell'$, along with sets $H_\ell \subseteq T \cap B_d(x_\ell, a/r)$ for $\ell = 1,\ldots,K$.

Let z_0 be a minimizer of $\|\!|u|\!\|^2$ over $u \in \mathrm{conv}(\bigcup_{\ell\leqslant K} H_\ell)$, and note that $z_0 \in T$ since T is closed and convex. Define $\theta_0 := \|\!|z_0|\!\|^2$ and

$$\theta := \max_{\ell\leqslant K} \min\left\{\|\!|u|\!\|^2 : u \in \mathrm{conv}(H_\ell)\right\},$$

and for each $\ell \in [K]$, let $z_\ell \in \mathrm{conv}(H_\ell)$ be such that $\|\!|z_\ell|\!\|^2 \leqslant \theta$.

Note that $\mathrm{conv}(H_\ell) \subseteq \mathrm{conv}(B_d(x_\ell, a/r)) \subseteq B_d(x_\ell, 4a/r)$, where the latter inclusion follows from Lemma 2.16. Since $z_\ell \in \mathrm{conv}(H_\ell)$, we have $d(x_\ell, z_\ell) \leqslant 4a/r$ for all $\ell \in \{1,\ldots,K\}$. In particular for $\ell, \ell' \in \{1,\ldots,K\}$ with $\ell \neq \ell'$, we can use the quasimetric inequalities (2.19) to write

$$\begin{aligned} a \leqslant d(x_\ell, x_{\ell'}) &\leqslant 2\sqrt{2}\,(d(x_\ell, z_\ell) + d(z_\ell, x_{\ell'})) \\ &\leqslant 2\sqrt{2}\,\frac{4a}{r} + 8\,(d(z_\ell, z_{\ell'}) + d(z_{\ell'}, x_{\ell'})) \leqslant (8+2\sqrt{2})\frac{4a}{r} + 8\,d(z_\ell, z_{\ell'}). \end{aligned}$$

Using our choice $r = 64$, we conclude that that for $\ell \neq \ell'$,

$$d(z_\ell, z_{\ell'}) \geqslant \frac{a}{32}\,. \tag{2.27}$$

Observe that

$$F_h\left(\bigcup_{\ell\leqslant K} H_\ell\right) - \min_{\ell\leqslant K} F_{h+1}(H_\ell) = (2-\theta_0) - (2-\theta) + \frac{1}{\log n} = \theta - \theta_0 + \frac{1}{\log n}\,,$$

thus to verify that the growth condition Definition 2.8 holds for $\{F_h\}$, our goal is to show that

$$\theta - \theta_0 + \frac{1}{\log n} \gtrsim \frac{2^{h/2} a}{\kappa + \lambda\sqrt{\log n}}\,, \qquad h = 0, 1, \ldots, h_0\,. \tag{2.28}$$

This will confirm the growth condition with $c^* \asymp \left(\lambda\sqrt{\log n} + \kappa\right)^{-1}$, and therefore (2.26) yields our desired goal (2.20).

The next lemma exploits 2-uniform convexity of the ℓ_2 distance. Note that the claimed inequality would fail (in general) if the left-hand side were replaced by the larger quantity $|||z_0 - z_\ell|||^2$, as $|||\cdot|||$ is not necessarily 2-convex.

Lemma 2.20. *For every $\ell = 1, \ldots, K$, it holds that*

$$\|z_0 - z_\ell\|^2 + \sum_{j=1}^{M} \left(N_j(z_0) - N_j(z_\ell)\right)^2 \leqslant 4(\theta - \theta_0)\,.$$

Proof. Let us use

$$\left(\frac{a-b}{2}\right)^2 = \frac{1}{2}a^2 + \frac{1}{2}b^2 - \left(\frac{a+b}{2}\right)^2.$$

to write

$$\begin{aligned}\left\|\frac{z_0 - z_\ell}{2}\right\|^2 + \sum_{j=1}^{M}\left(\frac{N_j(z_0) - N_j(z_\ell)}{2}\right)^2 &= \frac{1}{2}\left(\|z_\ell\|^2 + \sum_{j=1}^{M} N_j(z_\ell)^2\right) + \frac{1}{2}\left(\|z_0\|^2 + \sum_{j=1}^{M} N_j(z_0)^2\right) \\ &\quad - \left\|\frac{z_0 + z_\ell}{2}\right\|^2 - \sum_{j=1}^{M}\left(\frac{N_j(z_0) + N_j(z_\ell)}{2}\right)^2.\end{aligned}$$

By convexity of the norm N_j, we have $\frac{1}{2}(N_j(z_0) + N_j(z_\ell)) \geqslant N_j(\frac{z_0+z_\ell}{2})$, so the preceding identity gives

$$\begin{aligned}\left\|\frac{z_0 - z_\ell}{2}\right\|^2 + \sum_{j=1}^{M}\left(\frac{N_j(z_0) - N_j(z_\ell)}{2}\right)^2 &\leqslant \frac{1}{2}|||z_\ell|||^2 + \frac{1}{2}|||z_0|||^2 - \left|\left|\left|\tfrac{z_0+z_\ell}{2}\right|\right|\right|^2 \\ &\leqslant |||z_\ell|||^2 - \left|\left|\left|\tfrac{z_0+z_\ell}{2}\right|\right|\right|^2 \\ &\leqslant \theta - \theta_0\,,\end{aligned}$$

where the inequality $\left|\left|\left|\frac{z_0+z_\ell}{2}\right|\right|\right|^2 \geqslant \theta_0$ follows from $\frac{z_0+z_\ell}{2} \in \mathrm{conv}(\bigcup_{\ell \leqslant K} H_\ell)$, since $z_0 \in \mathrm{conv}(\bigcup_{\ell\leqslant K} H_\ell)$ and $z_\ell \in \mathrm{conv}(H_\ell)$. □

Define $\rho := \theta - \theta_0$. One consequence of Lemma 2.20 is that

$$z_1, \ldots, z_K \in z_0 + 2\sqrt{\rho}B_2^n\,.$$

We can cover $z_0 + 2\sqrt{\rho}B_2^n$ by N_h sets that have $\|\cdot\|_{\mathcal{N}}$-diameter bounded by $2e_h(2\sqrt{\rho}B_2^n, \|\cdot\|_{\mathcal{N}})$. Since we have $K = N_{h+1} = N_h^2$ points $z_1, \ldots, z_K$, at least N_h of them $z_{i_1}, \ldots, z_{i_{N_h}}$ must lie in the same set of the cover. And by definition, these points cannot all have pairwise $\|\cdot\|_{\mathcal{E}(z_0)}$ distance greater than $e_h(2\sqrt{\rho}B_2^n, \|\cdot\|_{\mathcal{E}(z_0)})$. Therefore we must have at least two points z_ℓ and $z_{\ell'}$ with $\ell \neq \ell'$ and $\ell, \ell' \geqslant 1$, and such that

$$\|z_\ell - z_{\ell'}\|_{\mathcal{N}} \leqslant 2e_h(2\sqrt{\rho}B_2^n, \|\cdot\|_{\mathcal{N}}) \lesssim 2^{-h/2}\kappa\sqrt{\rho}\,,$$
$$\|z_\ell - z_{\ell'}\|_{\mathcal{E}(z_0)} \leqslant e_h(2\sqrt{\rho}B_2^n, \|\cdot\|_{\mathcal{E}(z_0)}) \lesssim 2^{-h/2}\lambda\sqrt{\rho}\,,$$

where the latter two estimates follow from Lemma 2.5 and Lemma 2.18, respectively.

Let us also note a second consequence of Lemma 2.20, that

$$\sum_{j=1}^{M}\left(N_j(z_0) - N_j(z_\ell)\right)^2 + \sum_{j=1}^{M}\left(N_j(z_0) - N_j(z_{\ell'})\right)^2 \leqslant 4\rho\,.$$

Using the three preceding inequalities in Lemma 2.17 yields

$$a^2 \overset{(2.27)}{\lesssim} d(z_\ell, z_{\ell'})^2 \lesssim 2^{-h}\rho^2\kappa^2 + 2^{-h}\rho\lambda^2 \leqslant \max\left(2^{-h}\kappa^2\rho^2, 2^{-h}\lambda^2\rho\right).$$

This implies

$$\rho \gtrsim \min\left(\frac{2^{h/2}a}{\kappa}, \frac{2^h a^2}{\lambda^2}\right).$$

Since it holds that

$$\frac{2^h a^2}{\lambda^2} + \frac{1}{\log n} \geqslant \frac{2^{h/2}a}{\lambda\sqrt{\log n}},$$

we conclude that

$$\rho + \frac{1}{\log n} \gtrsim \min\left(\frac{2^{h/2}a}{\kappa}, \frac{2^{h/2}a}{\lambda\sqrt{\log n}}\right) \gtrsim \frac{2^{h/2}a}{\lambda\sqrt{\log n} + \kappa}\,.$$

Recalling that $\rho = \theta - \theta_0$, we have established (2.28), completing the proof. □

Remark 2.21 (Discussion of the implicit partitioning)**.** It is often more intuitive to think about bounding $\gamma_2(T, d)$ by explicitly constructing the sequence of partitions $\{\mathcal{A}_h\}$ (recall (2.2)). This is a technical process that is aided significantly by Theorem 2.9, whose proof involves the construction of partitions from growth functionals.

Recall the norm $|||\cdot|||$ from (2.24) and for a subset $S \subseteq B_2^n$, define the quantity

$$\varphi(S) := 2 - \min\left\{|||x|||^2 : x \in \operatorname{conv}(S)\right\}.$$

Then $\varphi(S)$ can be considered as an approximate measure of the "size" of S, where sets of larger $\varphi(S)$ value tend to have a larger $\mathbb{E}\sup_{x\in S}\sum_{j=1}^{M}\varepsilon_j N_j(x)^2$ value.

Recall that $r := 64$. Consider a ball $B_d(x_0, \eta)$, and let $z_0 \in B_d(x_0, 4\eta)$ be such that $\varphi(B_d(x_0, \eta)) = 2 - |||z_0|||^2$. Let us think of z_0 as the "analytic center" of the ball $B_d(x_0, \eta)$. (We have to take $z_0 \in B_d(x_0, 4\eta)$ because the ball $B_d(x_0, \eta)$ is only approximately convex.)

Define the distance

$$\Delta(x, y) := \left(\|x - y\|^2 + \sum_{j=1}^{M}(N_j(x) - N_j(y))^2\right)^{1/2}, \qquad x, y \in \mathbb{R}^n\,.$$

For $x \in B_d(x_0, \eta)$, let $\hat{x} \in B_d(x, 4\eta/r^2)$ denote a point satisfying $\varphi(B_d(x, \eta/r^2)) = 2 - |||\hat{x}|||^2$. Then Lemma 2.20 gives

$$\Delta(z_0, \hat{x})^2 \lesssim \varphi\left(B_d(x_0, \eta)\right) - \varphi\left(B_d(x, \eta/r^2)\right). \tag{2.29}$$

In other words, either the φ-value of $B_d(x, \eta/r^2)$ is significantly smaller than that of $B_d(x_0, \eta)$, or $\hat{x}$ is close (in the distance Δ) to the analytic center z_0.

The second part of the argument involves bounding the number of centers that can be within a certain distance of z_0. Consider now any points $x_1, \ldots, x_M \in B_d(x_0, \eta)$ with $d(x_i, x_j) > \eta/r$ for $i \neq j$. Lemma 2.17 and the covering estimates on $e_h(B_2^n, \|\cdot\|_{\mathcal{E}(z_0)})$ and $e_h(B_2^n, \|\cdot\|_{\mathcal{N}})$ together give that for some constant $C > 0$,

$$\#\left\{i \geqslant 1 : \Delta(z_0, \hat{x}_i)^2 \leqslant \rho\right\} \leqslant \exp\left(\frac{C}{\eta^2}\left(\kappa^2\rho^2 + \lambda^2\rho\right)\right). \tag{2.30}$$

Now (2.29) and (2.30) imply that for any $\delta > 0$,

$$\#\left\{i \geqslant 1 : \varphi\left(B_d(x_i, \eta/r^2)\right) \geqslant \varphi\left(B_d(x_0, \eta)\right) - \delta\right\} \leqslant \exp\left(\frac{C}{\eta^2}\left(\kappa^2\delta^2 + \lambda^2\delta\right)\right). \tag{2.31}$$

This is the key tradeoff occuring in the argument: A bound on the number of pairwise separated "children" $B_d(x_i, \eta/r^2)$ of $B_d(x_0, \eta)$ that do not experience a significant reduction in their φ-value.

Employing this bound repeatedly, in a sufficiently careful manner, allows one to construct a sequence of partitions $\{\mathcal{A}_h\}$ that yields the desired upper bound on $\gamma_2(T, d)$. The role of Theorem 2.9 is to automate this process.

3 Hypergraph sparsification

Suppose $H = (V, E, w)$ is a weighted hypergraph and denote $n := |V|$. For a single hyperedge $e \in E$, let us recall the definitions

$$Q_e(x) := \max_{\{u,v\} \in \binom{e}{2}} (x_u - x_v)^2,$$

as well as the energy

$$Q_H(x) := \sum_{e \in E} w_e Q_e(x).$$

3.1 Sampling

Suppose we have a probability distribution $\mu \in (0,1]^E$ on hyperedges in H. Let us sample hyperedges $\tilde{E} = \{e_1, e_2, \ldots, e_M\}$ independently according to μ. The weighted hypergraph $\tilde{H} = (V, \tilde{E}, \tilde{w})$ is defined so that

$$Q_{\tilde{H}}(x) = \frac{1}{M} \sum_{k=1}^{M} \frac{w_{e_k}}{\mu_{e_k}} Q_{e_k}(x),$$

In particular, $\mathbb{E}[Q_{\tilde{H}}(x)] = Q_H(x)$ for all $x \in \mathbb{R}^V$. Recall that the hyperedge weights in $\tilde{H}$ are given by (1.2). To help us choose the distribution μ, we now introduce a Laplacian on an auxiliary graph.

An auxiliary Laplacian. Define the edge set $F := \bigcup_{e\in E}\binom{e}{2}$, and let $G = (V, F, c)$ be a weighted graph, where we will choose the edge conductances $c \in \mathbb{R}_+^F$ later. Denote by $L_G : \mathbb{R}^V \to \mathbb{R}^V$ the weighted Laplacian

$$L_G := \sum_{\{i,j\}\in F} c_{ij}(\chi_i - \chi_j)(\chi_i - \chi_j)^*, \tag{3.1}$$

where $\chi_1, \dots, \chi_n$ is the standard basis of $\mathbb{R}^n$. Let L_G^+ denote its Moore-Penrose pseudoinverse and define

$$\begin{aligned}
\mathsf{R}_{ij} &:= \|L_G^{+/2}(\chi_i - \chi_j)\|^2, && \{i,j\} \in F\,,\\
\mathsf{R}_{\max}(e) &:= \max\left\{\mathsf{R}_{ij} : \{i,j\} \in \tbinom{e}{2}\right\}, && e \in E\,,\\
Z &:= \sum_{e\in E} w_e \mathsf{R}_{\max}(e)\,,\\
\mu_e &:= \frac{w_e \mathsf{R}_{\max}(e)}{Z}\,, && e \in E\,.
\end{aligned} \tag{3.2}$$

Lemma 3.1. *Suppose it holds that*

$$\ker(L_G) = \operatorname{span}\{(1,1,\dots,1)\}\,, \tag{3.3}$$

$$\|x\|^2 \leqslant Q_H(L_G^{+/2}x)\,, \qquad \forall x \in \mathbb{R}^n\,. \tag{3.4}$$

Then for any $\varepsilon \in (0,1)$, there is a number

$$M_0 \lesssim \frac{\log D}{\varepsilon^2} Z \log n$$

such that for $M \geqslant M_0$, with probability at least $1/2$, the hypergraph $\tilde{H}$ is a spectral ε-sparsifier for H.

Proof. We may assume that $w_e > 0$ for each $e \in E$, as otherwise the corresponding edge can be removed from H. Note that $\mathsf{R}_{ij} > 0$ for all $\{i,j\} \in F$ by (3.3) and the fact that $\chi_i - \chi_j \notin \operatorname{span}\{(1,1,\dots,1)\}$. In particular, the sampling probabilities $\{\mu_e : e \in E\}$ defined in (3.2) are strictly positive.

By convexity,

$$\mathop{\mathbb{E}}_{\tilde{H}} \max_{v: Q_H(v)\leqslant 1} \left|Q_H(v) - Q_{\tilde{H}}(v)\right| \leqslant \mathop{\mathbb{E}}_{\tilde{H},\hat{H}} \max_{v: Q_H(v)\leqslant 1} |Q_{\tilde{H}}(v) - Q_{\hat{H}}(v)|\,, \tag{3.5}$$

where $\hat{H}$ is an independent copy of $\tilde{H}$.

The latter quantity can be written as

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\tilde{e},\hat{e}} \max_{v: Q_H(v)\leqslant 1} \left|\frac{1}{M}\sum_{i=1}^M \frac{w_{\tilde{e}_i}}{\mu_{\tilde{e}_i}} Q_{\tilde{e}_i}(v) - \frac{1}{M}\sum_{i=1}^M \frac{w_{\hat{e}_i}}{\mu_{\hat{e}_i}} Q_{\hat{e}_i}(v)\right|\\
&= \mathop{\mathbb{E}}_{\varepsilon} \mathop{\mathbb{E}}_{\tilde{e},\hat{e}} \max_{v: Q_H(v)\leqslant 1} \left|\frac{1}{M}\sum_{i=1}^M \varepsilon_i \left(\frac{w_{\tilde{e}_i}}{\mu_{\tilde{e}_i}} Q_{\tilde{e}_i}(v) - \frac{w_{\hat{e}_i}}{\mu_{\hat{e}_i}} Q_{\hat{e}_i}(v)\right)\right| \qquad (3.6)\\
&\leqslant 2 \mathop{\mathbb{E}}_{\tilde{H}} \mathop{\mathbb{E}}_{\varepsilon} \max_{v: Q_H(v)\leqslant 1} \left|\frac{1}{M}\sum_{i=1}^M \varepsilon_i \frac{w_{e_i}}{\mu_{e_i}} Q_{e_i}(v)\right|, \qquad (3.7)
\end{aligned}$$

where $\varepsilon_1, \ldots, \varepsilon_M$ are i.i.d. Bernoulli ± 1 random variables. Note that we can introduce signs in (3.6) because the distribution of $\frac{w_{\tilde{e}_i}}{\mu_{\tilde{e}_i}} Q_{\tilde{e}_i}(v) - \frac{w_{\hat{e}_i}}{\mu_{\hat{e}_i}} Q_{\hat{e}_i}(v)$ is symmetric.

For $e \in E$ and $\{i,j\} \in \binom{e}{2}$, define the vectors

$$\begin{aligned} y_{ij} &:= L_G^{+/2}(\chi_i - \chi_j) \\ y_{ij}^e &:= \sqrt{\frac{w_e}{\mu_e}}\, y_{ij} = \sqrt{\frac{Z}{\mathsf{R}_{\max}(e)}}\, y_{ij}\,. \end{aligned}$$

Then we have

$$\frac{w_e}{\mu_e} Q_e(L_G^{+/2}x) = \frac{w_e}{\mu_e} \max_{\{i,j\}\in\binom{e}{2}} |\langle L_G^{+/2}x, \chi_i - \chi_j\rangle|^2 = \max_{\{i,j\}\in\binom{e}{2}} \langle x, y_{ij}^e\rangle^2\,. \tag{3.8}$$

Define the values

$$S_{ij} := \max_{e\in E:\{i,j\}\in\binom{e}{2}} \|y_{ij}^e\|\,, \quad \{i,j\} \in F,$$

and the linear map $A : \mathbb{R}^n \to \mathbb{R}^F$ by $(Ax)_{\{i,j\}} := S_{ij}\langle x, y_{ij}/\|y_{ij}\|\rangle$.

For $k = 1, \ldots, M$, define the weighted ℓ_∞ norms

$$N_k(z) := \max\left\{ \left|(Az)_{\{i,j\}}\right| \frac{\|y_{ij}^{e_k}\|}{S_{ij}} : \{i,j\} \in \binom{e_k}{2}, S_{ij} > 0 \right\}.$$

It holds that

$$N_k(x) = \max_{\{i,j\}\in e_k} |\langle x, y_{ij}^{e_k}\rangle|\,,$$

so from (3.8), we have

$$Q_{\tilde{H}}(L_G^{+/2}x) = \frac{1}{M}\sum_{i=1}^M N_i(x)^2\,, \tag{3.9}$$

$$\frac{1}{M}\sum_{i=1}^M \varepsilon_i \frac{w_{e_i}}{\mu_{e_i}} Q_{e_i}(L_G^{+/2}x) = \frac{1}{M}\sum_{i=1}^M \varepsilon_i N_i(x)^2\,. \tag{3.10}$$

Thus we can write the quantity (3.7) as

$$2 \mathop{\mathbb{E}}_{\tilde{H}} \mathop{\mathbb{E}}_{\varepsilon} \max_{x:Q_H(L_G^{+/2}x)\leqslant 1} \left| \frac{1}{M}\sum_{i=1}^M \varepsilon_i N_i(x)^2 \right| \leqslant 4 \mathop{\mathbb{E}}_{\tilde{H}} \mathop{\mathbb{E}}_{\varepsilon} \max_{x:Q_H(L_G^{+/2}x)\leqslant 1} \frac{1}{M}\sum_{i=1}^M \varepsilon_i N_i(x)^2,$$

Define $T := \{x \in \mathbb{R}^n : Q_H(L_G^{+/2}x) \leqslant 1\}$ and note that from (3.4), we have $T \subseteq B_2^n$. Now apply Corollary 2.13 to bound

$$\mathop{\mathbb{E}}_{\varepsilon} \max_{x\in T} \frac{1}{M}\sum_{i=1}^M \varepsilon_i N_i(x)^2 \lesssim \frac{\|A\|_{2\to\infty}\sqrt{\log n \log D}}{M^{1/2}} \max_{x\in T} \left(\frac{1}{M}\sum_{i=1}^M N_i(x)^2\right)^{1/2}. \tag{3.11}$$

Note also that

$$\max_{x\in T} \frac{1}{M}\sum_{i=1}^{M} N_i(x)^2 = \max_{v:Q_H(v)\leqslant 1} \frac{1}{M}\sum_{i=1}^{M} N_i\left(L_G^{1/2}v\right)^2 = \max_{v:Q_H(v)\leqslant 1} Q_{\tilde{H}}(v)\,.$$

where the first equality follows from the fact that $Q_H(x) = Q_H(\hat{x})$ when $x - \hat{x} \in \ker(L_G)$, and the second inequality uses this and an application of (3.9) with $x = L_G^{1/2}v$.

Recalling our starting point (3.5), it follows that for some universal constant $C > 0$,

$$\begin{aligned}\tau := \mathop{\mathbb{E}}_{\tilde{H}} \max_{v:Q_H(v)\leqslant 1} \left|Q_H(v) - Q_{\tilde{H}}(v)\right| &\leqslant C\frac{\|A\|_{2\to\infty}\sqrt{\log n \log D}}{M^{1/2}} \mathop{\mathbb{E}}_{\tilde{H}}\left(\max_{v:Q_H(v)\leqslant 1} Q_{\tilde{H}}(v)\right)^{1/2} \\ &\leqslant C\frac{\|A\|_{2\to\infty}\sqrt{\log n \log D}}{M^{1/2}} \left(\mathop{\mathbb{E}}_{\tilde{H}} \max_{v:Q_H(v)\leqslant 1} Q_{\tilde{H}}(v)\right)^{1/2},\end{aligned}$$

where the last inequality is by concavity of the square root.

Observe that

$$\max_{v:Q_H(v)\leqslant 1} Q_{\tilde{H}}(v) \leqslant \max_{v:Q_H(v)\leqslant 1} \left(\left|Q_H(v) - Q_{\tilde{H}}(v)\right| + Q_H(v)\right) \leqslant 1 + \max_{v:Q_H(v)\leqslant 1} |Q_H(v) - Q_{\tilde{H}}(v)|\,,$$

and therefore we have

$$\tau \leqslant C\frac{\|A\|_{2\to\infty}\sqrt{\log n \log D}}{M^{1/2}}\,(1+\tau)^{1/2}\,.$$

It follows that if $M \geqslant (2C\|A\|_{2\to\infty}\sqrt{\log n \log D})^2$, then $\tau \leqslant 4C\frac{\|A\|_{2\to\infty}\sqrt{\log n \log D}}{M^{1/2}}$.

For $0 < \varepsilon < 1$, choosing

$$M := \frac{4C^2 \log D}{\varepsilon^2}\|A\|_{2\to\infty}^2 \log n$$

gives

$$\mathop{\mathbb{E}}_{\tilde{H}} \max_{v:Q_H(v)\leqslant 1} \left|Q_H(v) - Q_{\tilde{H}}(v)\right| = \tau \leqslant \varepsilon\,.$$

The proof is complete once we observe that

$$\|A\|_{2\to\infty}^2 = \max_{\{i,j\}\in F} S_{ij}^2 = \max_{e\in E, \{i,j\}\in\binom{e}{2}} \|y_{ij}^e\|^2 = Z \max_{\{i,j\}\in\binom{e}{2}} \frac{\mathsf{R}_{ij}}{\mathsf{R}_{\max}(e)} \leqslant Z\,. \qquad \square$$

3.2 Choosing conductances

We are therefore left to find edge conductances in the graph $G = (V, F, c)$ so that (3.3) and (3.4) hold and Z is small. To this end, let us choose nonnegative numbers

$$\left\{c_{ij}^e \geqslant 0 : \{i,j\} \in \binom{e}{2}, e \in E\right\}$$

such that

$$\sum_{\{i,j\}\in\binom{e}{2}} c_{ij}^e = w_e, \quad \forall e \in E\,. \tag{3.12}$$

For $\{i,j\} \in F$, we then define our edge conductance

$$c_{ij} := \sum_{e \in E : \{i,j\} \in \binom{e}{2}} c_{ij}^e \,. \tag{3.13}$$

In this case,

$$\begin{aligned}
\|L_G^{1/2} v\|^2 = \langle v, L_G v\rangle &= \sum_{\{i,j\} \in F} c_{ij}(v_i - v_j)^2 \\
&= \sum_{e \in E} \sum_{\{i,j\} \in \binom{e}{2}} c_{ij}^e (v_i - v_j)^2 \\
&\leqslant \sum_{e \in E} \sum_{\{i,j\} \in \binom{e}{2}} c_{ij}^e \max_{\{i,j\} \in \binom{e}{2}} (v_i - v_j)^2 \\
&\overset{(3.12)}{\leqslant} \sum_{e \in E} w_e \max_{\{i,j\} \in \binom{e}{2}} (v_i - v_j)^2 = Q_H(v)\,.
\end{aligned}$$

Taking $v = L_G^{+/2} x$ gives

$$\|x\|^2 \leqslant Q_H(L_G^{+/2} x),$$

verifying (3.4).

Lemma 3.2 (Foster's Network Theorem)**.** *It holds that* $\sum_{\{i,j\} \in F} c_{ij} \mathsf{R}_{ij} \leqslant n - 1$.

Proof. Recall that $\mathsf{R}_{ij} = \langle \chi_i - \chi_j, L_G^+(\chi_i - \chi_j)\rangle$ and $L_G = \sum_{\{i,j\} \in F} c_{ij}(\chi_i - \chi_j)(\chi_i - \chi_j)^*$. It follows that

$$\sum_{\{i,j\} \in F} c_{ij} \mathsf{R}_{ij} = \sum_{\{i,j\} \in F} \operatorname{tr}(c_{ij}(\chi_i - \chi_j)(\chi_i - \chi_j)^* L_G^+) = \operatorname{tr}(L_G L_G^+) \leqslant n - 1\,,$$

since $\operatorname{rank}(L_G) \leqslant n - 1$. □

Define

$$K := \max_{e \in E} \max_{\{i,j\} \in \binom{e}{2}} \frac{\mathsf{R}_{\max}(e)}{\mathsf{R}_{ij}} \mathbb{1}_{\{c_{ij}^e > 0\}} \tag{3.14}$$

so that

$$Z = \sum_{e \in E} w_e \mathsf{R}_{\max}(e) = \sum_{e \in E} \sum_{\{i,j\} \in \binom{e}{2}} c_{ij}^e \mathsf{R}_{\max}(e) \leqslant K \sum_{e \in E} \sum_{\{i,j\} \in \binom{e}{2}} c_{ij}^e \mathsf{R}_{ij} \leqslant K(n-1)\,,$$

where the last inequality uses (3.13) and Lemma 3.2. In conjunction with Lemma 3.1, we have proved the following.

Lemma 3.3. *Suppose there is a choice of conductances so that* (3.12) *holds and* $\ker(L_G) = \operatorname{span}\{(1, \ldots, 1)\}$. *Then for any* $\varepsilon > 0$, *there is a spectral* ε*-sparsifier for* H *with at most* $O(K \frac{\log D}{\varepsilon^2} n \log n)$ *hyperedges, where* K *is defined in* (3.14).

3.3 Balanced effective resistances

We will exhibit conductances satisfying (3.3), (3.12), and (3.14) with $K \leqslant 1$. To this end, we may assume that the weighted hypergraph $H = (V, E, w)$ has strictly positive edge weights and that the (unweighted) graph $G_0 = (V, F)$ is connected.

Define $\hat{F} := \{(e, \{i, j\}) : e \in E, \{i, j\} \in \binom{e}{2}\}$, and consider vectors $\left(c_{ij}^e : e \in E, \{i, j\} \in \binom{e}{2}\right) \in \mathbb{R}_+^{\hat{F}}$. Define the convex set

$$\mathsf{K} := \mathbb{R}_+^{\hat{F}} \cap \left\{ \sum_{\{i,j\} \in \binom{e}{2}} c_{ij}^e = w_e : e \in E \right\}.$$

We use $\mathcal{S}_+^n$ and $\mathcal{S}_{++}^n$ for the cones of positive semidefinite (resp., positive definite) $n \times n$ matrices. Define $c_{ij} := \sum_{e:\{i,j\} \in \binom{e}{2}} c_{ij}^e$ and denote the linear function $L_G : \mathbb{R}_+^F \to \mathcal{S}_+^n$ by

$$L_G\left((c_{ij})\right) := \sum_{\{i,j\} \in F} c_{ij} (\chi_i - \chi_j)(\chi_i - \chi_j)^* .$$

Let J be the all-ones matrix and consider the objective

$$\Phi\left((c_{ij})\right) := -\log\det\left(L_G\left((c_{ij})\right) + J\right) .$$

Note that $X \mapsto -\log\det(X)$ is a convex function on the cone $\mathcal{S}_+^n$ of $n \times n$ positive semidefinite matrices (see, e.g., [BV04, §3.1]) and takes the value $+\infty$ on $\mathcal{S}_+^n \setminus \mathcal{S}_{++}^n$. Consider finally the convex optimization problem:

$$\min\left\{\Phi\left((c_{ij})\right) : (c_{ij}^e) \in \mathsf{K}\right\} . \tag{3.15}$$

Since G_0 is connected, it holds that if $\left(c_{ij}\right) \in \mathbb{R}_{++}^F$, then $\ker(L_G)$ is the span of $(1, 1, \ldots, 1)$, and therefore $L_G\left((c_{ij})\right) + J \in \mathcal{S}_{++}^n$. Therefore Φ is finite on the strictly positive orthant $\mathbb{R}_{++}^F$.

Lemma 3.4. *The value of* (3.15) *is finite and there is a feasible point in the relative interior of* K.

Proof. It is straightforward to check that the maximum of eigenvalue of L_G is bounded by $2\sum_{\{i,j\} \in \binom{e}{2}} c_{ij} = 2\sum_{e \in E} w_e$, hence the value of (3.15) is finite. Moreover, the vector defined by $c_{ij}^e := \frac{1}{|\binom{e}{2}|} w_e$ is feasible and lies in $\mathbb{R}_{++}^{\hat{F}}$ since the weights w_e are strictly positive. $\square$

Note that $\Phi\left((c_{ij})\right)$ is only finite if $\ker(L_G + J) = \{0\}$, i.e., if $\ker(L_G) = \operatorname{span}\{(1, \ldots, 1)\}$. Therefore an optimal solution to (3.15) automatically satisfies (3.3).

Let us write the corresponding Lagrangian as

$$g\left((c_{ij}^e); \alpha, \beta\right) = -\log\det\left(L_G\left((c_{ij})\right) + J\right) + \sum_{e \in E} \alpha_e \left(\sum_{\{i,j\} \in \binom{e}{2}} c_{ij}^e - w_e \right) - \sum_{e \in E} \sum_{\{i,j\} \in \binom{e}{2}} \beta_{ij}^e c_{ij}^e$$

Lemma 3.4 allows one to conclude that there are vectors $(\hat{c}_{ij}^e)$, $\hat{\alpha}, \hat{\beta}$ with $\hat{\beta} \geqslant 0$ and such that the KKT conditions hold; see [Roc70, Thm 28.2]. In particular, for all $e \in E$ and $\{i, j\} \in \binom{e}{2}$, we have

$$\partial_{c_{ij}^e}\, g\left((\hat{c}_{ij}^e); \hat{\alpha}, \hat{\beta}\right) = 0 , \tag{3.16}$$

$$\hat{\beta}^e_{ij} > 0 \implies \hat{c}^e_{ij} = 0\,. \tag{3.17}$$

By the rank-one update formula for the determinant, we have

$$\partial_{c^e_{ij}} \log\det(L_G + J) = \langle \chi_i - \chi_j, (L_G + J)^{-1}(\chi_i - \chi_j)\rangle\,.$$

Define $\hat{L}_G := L_G\left((\hat{c}_{ij})\right)$. Define $\hat{\mathsf{R}}_{ij} := \langle \chi_i - \chi_j, \hat{L}_G^+(\chi_i - \chi_j)\rangle$. Taking the derivative of g with respect to each c^e_{ij} and using (3.16) gives

$$\hat{\mathsf{R}}_{ij} = \langle \chi_i - \chi_j, (\hat{L}_G + J)^{-1}(\chi_i - \chi_j)\rangle = \hat{\alpha}_e - \hat{\beta}^e_{ij}, \qquad \forall e \in E, \{i,j\} \in \binom{e}{2}\,,$$

where the first equality uses the fact that the eigenvectors of $\hat{L}_G$ and J are orthogonal and $\chi_i - \chi_j \in \ker J$.

Note that since $\hat{\beta} \geqslant 0$ coordinate-wise, this implies that

$$\hat{\mathsf{R}}_{\max}(e) := \max_{\{i,j\}\in\binom{e}{2}} \hat{\mathsf{R}}_{ij} \leqslant \hat{\alpha}_e\,.$$

Moreover, if $\hat{c}^e_{ij} > 0$, then $\hat{\beta}^e_{ij} = 0$ (cf. (3.17)), and in that case $\hat{\mathsf{R}}_{ij} = \hat{\alpha}_e = \hat{\mathsf{R}}_{\max}(e)$.

We conclude that the edge conductances $\hat{c}^e_{ij}$ yield $K \leqslant 1$ in (3.14), and therefore Lemma 3.3 gives a sparsifier with $O(\frac{\log D}{\varepsilon^2} n \log n)$ edges, completing the proof of Theorem 1.1.

Acknowledgements

I am grateful to Thomas Rothvoss for many suggestions and comments on preliminary drafts.

References

[Bar98] Franck Barthe. On a reverse form of the Brascamp-Lieb inequality. *Invent. Math.*, 134(2):335–361, 1998.

[BDOS21] Sander Borst, Daniel Dadush, Neil Olver, and Makrand Sinha. Majorizing measures for the optimizer. In *12th Innovations in Theoretical Computer Science Conference*, volume 185 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 73, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021.

[BST19] Nikhil Bansal, Ola Svensson, and Luca Trevisan. New notions and constructions of sparsification for graphs and hypergraphs. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 910–928. IEEE Computer Society, 2019.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[CKN20] Yu Chen, Sanjeev Khanna, and Ansh Nagda. Near-linear size hypergraph cut sparsifiers. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 61–72. IEEE, 2020.

[HM13] Moritz Hardt and Ankur Moitra. Algorithms and hardness for robust subspace recovery. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *COLT 2013 - The 26th Annual Conference on Learning Theory, June 12-14, 2013, Princeton University, NJ, USA*, volume 30 of *JMLR Workshop and Conference Proceedings*, pages 354–375. JMLR.org, 2013.

[JLS22] A. Jambulapati, Y. Liu, and A. Sidford. Chaining, group leverage score overestimates, and fast spectral hypergraph sparsification. Preprint at arXiv:2209.10539, 2022.

[KKTY21a] Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Spectral hypergraph sparsifiers of nearly linear size. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 1159–1170. IEEE, 2021.

[KKTY21b] Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Towards tight bounds for spectral sparsification of hypergraphs. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 598–611. ACM, 2021.

[LPP91] Françoise Lust-Piquard and Gilles Pisier. Noncommutative Khintchine and Paley inequalities. *Ark. Mat.*, 29(2):241–260, 1991.

[LT11] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[Roc70] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[Rud99a] M. Rudelson. Almost orthogonal submatrices of an orthogonal matrix. *Israel J. Math.*, 111:143–155, 1999.

[Rud99b] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999.

[SS11] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011.

[ST11] Daniel A. Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM J. Comput.*, 40(4):981–1025, 2011.

[SY19] Tasuku Soma and Yuichi Yoshida. Spectral sparsification of hypergraphs. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2570–2581. SIAM, 2019.

[Tal14] Michel Talagrand. *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Heidelberg, 2014. Modern methods and classical problems.