Abstract

Spectral Sparsification and Restricted Invertibility

Nikhil Srivastava
2010

In this thesis we prove the following two basic statements in linear algebra. Let B be an arbitrary $n \times m$ matrix where $m \geq n$ and suppose $0 < \epsilon < 1$ is given.

1. **Spectral Sparsification.** There is a nonnegative diagonal matrix $S_{m \times m}$ with at most $\lceil n/\epsilon^2 \rceil$ nonzero entries for which $(1 - \epsilon)^2 BB^T \preceq BSB^T \preceq (1 + \epsilon)^2 BB^T$. Thus the spectral behavior of BB^T is captured by a *weighted* subset of the columns of B, of size proportional to its rank n.

2. **Restricted Invertibility.** There is a diagonal $S_{m \times m}$ with at least $k = (1 - \epsilon)^2 \frac{\|B\|_F^2}{\|B\|_2^2}$ nonzero entries, all equal to 1, for which BSB^T has k eigenvalues greater than $\epsilon^2 \frac{\|B\|_F^2}{m}$. Thus there is a large coordinate restriction of B (i.e., a submatrix of its columns, given by S), of size proportional to its *numerical rank* $\frac{\|B\|_F^2}{\|B\|_2^2}$, which is well–invertible. This improves a theorem of Bourgain and Tzafriri [14].

We give deterministic algorithms for constructing the promised diagonal matrices S in time $O(mn^3/\epsilon^2)$ and $O((1 - \epsilon)^2 mn^3)$, respectively.

By applying (1) to the class of Laplacian matrices of graphs, we show that every graph on n vertices can be spectrally approximated by a weighted graph with $O(n)$ edges, thus generalizing the concept of expanders, which are constant–degree approx–imations of the complete graph. Our quantitative bounds are within a factor of two of those achieved by the celebrated Ramanujan graphs. We then present a second graph sparsification algorithm based on random sampling, which produces weaker sparsifiers with $O(n \log n)$ edges but runs in nearly–linear time.

We also prove a refinement of (1) for the special case of B arising from John's decompositions of the identity, which allows us to show that every convex body is close to one which has very few contact points with its minimum volume ellipsoid.

Spectral Sparsification and Restricted Invertibility

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Nikhil Srivastava

Dissertation Director: Daniel A. Spielman

May 2010

for Mamma, Papa, Neha, Moni Didi,
and Cundi.

Acknowledgments

This thesis and graduate school in general turned out better than I could have imagined, and that is because of many people. First I want to thank Dan Spielman, who I was outrageously lucky to have as my advisor. Dan tolerated and encouraged me in the early days, when I had no idea what a singular value was and was often paralyzed by fear. Homer (Simpson) said trying is the first step towards failure; without Dan's support I might still be stuck in that well. Dan was generous in every respect — with time, freedom, ideas, advice, and *MATLAB* refutations of conjectures (which should have been my job). He was happy for me when I succeeded, and he smiled like a kid at my first conference talk. Without his heroic optimism (and of course, his bold mathematics), we would certainly have given up on some of the stronger theorems in this thesis. Knowing Dan is a pleasure and an honor, both personally and intellectually, and unquestionably one of the best things in my life.

I am forever indebted to my two great undergraduate advisors at Union: Alan Taylor drew me to mathematics with his beautiful lectures, went out of his way to get me involved in research, supplied wise maxims which I live by to this day, and was generally an awesome role model; Peter Heinegg taught me how to read and write and (along with Rosie Heinegg) cut weeds and paint furniture — I ended up minoring in English after trying to take every class that he offered, and they were all great. I also learned a lot from lectures and collaborations with Chris Fernandes, Brian Postow, Steve Sargent, and Bill Zwicker.

At Yale, I thank Dana Angluin for being my friend and unofficial guide starting the day I arrived, David Pollard for teaching the best course I have ever taken, and Joan Feigenbaum for convincing me to come here in the first place. I thank my committee members Vladmir Rokhlin, Steve Zucker, and Shang-Hua Teng for checking out my thesis, and Linda Dobb for helping me submit it. I thank my pal Lev Reyzin, with whom I had fun and symbiotic collaborations and some of my first successes in research in graduate school, as well as my first sprats. I was also very fortunate to work with Josh Batson and Adam Marcus, and some of our joint results appear here.

I thank Ravi Kannan and Rina Panigrahy, who mentored me during wonderful summer internships at Microsoft Research in Bangalore and Silicon Valley. Moses Charikar invited me to give one of my first talks; I was lucky to run into him again at SVC, where we had several exciting discussions about graphs. I enjoyed hanging out and working with many interns during those summers — Hari, Moritz, Debmalya, Yi, Madhur — and also with Amit Deshpande, Ittai Abraham, and Parikshit Gopalan.

Contents

Chapter 1

Introduction

Matrices are central objects in Mathematics and Computer Science. They can be used to represent a wide variety of data, for instance: graphs, point configurations, polyhedra, covariances of random variables, systems of linear equations, Markov chains, quantum states, recurrences, second order derivatives, and general linear transformations. Depending on the context, we may emphasize different features of a matrix A: the entries (a_{ij}) themselves, the linear function $x \mapsto Ax$, the quadratic form $x \mapsto x^T Ax$, the eigendecomposition/invariant subspaces, and so on; every square matrix comes equipped with all of these interpretations, and each can be enlightening in its own way.

In this thesis, we will investigate the extent to which arbitrary matrices can be approximated by 'sparse' or 'structured' ones. We will show that for a certain large class of matrices this is always possible in a certain useful sense. We will then use that fact and related ones to obtain new results in:

- **Graph Theory (Chapter 4).** Every undirected weighted graph G on n vertices can be approximated by a weighted graph H which has $O(n)$ edges. This generalizes the concept of expander graphs, which are constant-degree approximations of the complete graph — in fact, the quantitative bounds which we achieve are within a factor of two of those achieved by the celebrated Ramanujan graphs. Besides being of inherent interest, the graph H, which we call a *sparsifier* can be used to speed up computations on G. We give a polynomial time algorithm for constructing H with $O(n)$ edges, as well as a nearly-linear time algorithm for constructing H with $O(n \log n)$ edges.

- **Asymptotic Convex Geometry (Chapter 5).** Every convex body K in $\mathbb{R}^n$ is close to a body H which has at most $O(n)$ contact points with the minimum volume ellipsoid which contains it. This improves a result of Rudelson [38], who showed the existence of H with $O(n \log n)$ contact points.

- **Functional/Numerical Analysis (Chapter 6).** Every matrix with large trace has a *coordinate* subspace of size proportional to its *numerical rank* on which it is well-invertible (i.e., close to an isometry). Qualitatively, this fact has been

1

known for some time as the Restricted Invertibility Theorem of Bourgain and Tzafriri [14]. Our construction is significantly simpler and supplies much sharper quantitative bounds.

A notable feature of our proofs is that they use only elementary linear algebra and yield deterministic polynomial time algorithms for constructing the desired objects. All previous results in these directions were weaker and had considerably more complex proofs which were probabilistic or nonconstructive.

All of our theorems rest on two foundational results, which we will motivate and introduce in the next two sections. In each case, we will start by asking a basic question about matrices, show that the spectral theorem provides a weak answer to that question, and then present the result of this thesis as a more satisfying answer.

1.1   Spectral Sparsification

The object of study in this section is a positive semidefinite matrix A of rank n written as a sum of outer products

$$A = \sum_{i \leq m} v_i v_i^T,$$

where the number of terms m may be much more than the rank n. Here we think of the v_i as being 'elementary', 'meaningful', or 'interesting' directions arising from the combinatorial, probabilistic, geometric, or other origin of the matrix. The following examples illustrate how such a representation can provide a more natural coordinate system for A than simply considering its entries.

Example 1.1. (Graph Laplacians). Suppose $G = (V, E, w)$ is an undirected weighted graph on n vertices. Then the *Laplacian* of G is defined as

$$L_G = \sum_{ij \in E} w_{ij}(e_i - e_j)(e_i - e_j)^T.$$

Thus the space of Laplacians of weighted graphs on n vertices is simply the set of all nonnegative linear combinations of

$$\left\{ (e_i - e_j)(e_i - e_j)^T : i, j \in [n], i \neq j \right\}$$

which correspond to the possible edges between vertices.

Example 1.2. (Covariance Matrices). Suppose $X \subset \mathbb{R}^n$ is a discrete set of points and we are interested in probability distributions p on X. Then the set of covariance matrices

$$A = \mathbb{E}_p x x^T = \sum_{x \in X} p(x) x x^T$$

is the convex hull of $\{xx^T : x \in X\}$.

More generally, such a representation as a sum of outer products can arise from any $n \times m$ matrix B with columns b_i by considering $A = BB^T = \sum_{i \leq m} b_i b_i^T$; we often do this, for instance, when computing the singular values of B.

The question we seek to answer is:

Question 1.3. When does A admit a *sparse* representation as a sum of outer products?

Being able to write A as the sum of a *small number* of outer products may be useful in succinctly storing, computing with, and understanding the behavior of A. One answer to this question is readily supplied by the spectral theorem.

Answer 1.4. (Sparse Representation in Eigenbasis). As A is symmetric and positive semidefinite, all of its eigenvalues are nonnegative real numbers. Thus there are efficiently computable nonnegative weights $\lambda_1 \geq \ldots, \geq \lambda_n$ (the eigenvalues) and vectors $u_1, \ldots u_n$ (the eigenvectors) for which

$$A = \sum_{i \leq n} \lambda_i u_i u_i^T,$$

and we have written A compactly as a weighted sum of n outer products, which is the minimal number. $\square$

The above representation has several unique and desirable geometric properties — for instance, the eigenvectors are orthogonal and cleanly decouple the action of A into separate components — which account for the immense power of the spectral theorem. However, it has the serious drawback that the eigenvectors $\{u_j\}_{j \leq n}$ need not have meaningful interpretations in terms of the elementary directions $\{v_i\}_{i \leq m}$; indeed, decomposing the action of A along eigenspaces may be quite unnatural in terms of the v_i. To see this more concretely, consider that Answer 1.4 allows us to represent the Laplacian of any graph on n vertices as a sum of n outer products in its eigenvectors u_j; however, the u_j do not in general correspond to *edges* $(e_i - e_j)$, and so this does *not* tell us anything about, say, the graph being equivalent a sparser graph with n edges. In essence, as soon as we start talking about $u_j u_j^T$ which do not come from our set of elementary directions, we have left the space of matrices which correspond to graphs.

To remedy the above situation, we consider a more demanding task: given a rank n matrix $A = \sum_{i \leq m} v_i v_i^T$, we seek to represent A by a shorter sum of the elementary $v_i v_i^T$, ideally of length comparable to n as attained in Answer 1.4. Unfortunately it is not always the case that there are sparse weights $s_i \geq 0$ for which $A = \sum_i s_i v_i v_i^T$ exactly. What we show is that such weights do always exist (and can be computed efficiently) if we are willing to settle for an *approximation* of A rather than an exact representation. Let us take a moment to describe the notion of approximation that we use.

Definition 1.5 (Matrix Approximation). The positive semidefinite cone $\mathcal{S}_n$ comes with a natural partial order, given by

$$A \succeq B \qquad \text{iff} \qquad A - B \in \mathcal{S}_n.$$

This induces a strong notion of *multiplicative approximation* for such matrices: we say that $\tilde{A}$ is a κ-approximation for A if

$$A \preceq \tilde{A} \preceq \kappa \cdot A.$$

The above compares favorably with more common notions of approximation that appear in the literature, such as requiring $\|A - \tilde{A}\|$ to be small in some norm; we defer a more thorough discussion to Section 3.3.

We are now in a position to state our first main theorem.

Theorem 1.6 (Spectral Sparsification). *Suppose $0 < \epsilon < 1$ and*

$$A = \sum_{i \leq m} v_i v_i^T$$

are given, with $v_i \in \mathbb{R}^n$. Then there are nonnegative weights $\{s_i\}_{i \leq m}$, at most $\lceil n/\epsilon^2 \rceil$ of which are nonzero, for which

$$(1 - \epsilon)^2 A \preceq \tilde{A} = \sum_{i \leq m} s_i v_i v_i^T \preceq (1 + \epsilon)^2 A.$$

There is an algorithm which computes the weights s_i in deterministic $O(n^3 m/\epsilon^2)$ time.

In short, the theorem says that if A can be written as *any* nonnegative linear combination of elementary (rank one) matrices, then it can be approximated by a *sparse* nonnegative linear combination of the same elementary matrices. The number of terms $\lceil n/\epsilon^2 \rceil$ is only a constant factor greater than that promised by the spectral theorem in Fact 1.4. In fact, we show in Section 3.1.3 that the sparsity/approximation tradeoff which we achieve cannot be improved by more than a factor of 4.

We mention in passing that the above bears some syntactic resemblance to the well-studied sparse approximation problem for vectors (see, for instance, [48]), in which we are given a *dictionary* of elementary vectors $\{v_i\}_{i \leq m} \subseteq \mathbb{R}^n$ and a signal vector $x \in \mathbb{R}^n$ and we seek a sparse set of coefficients s_i for which $x = \sum_i s_i v_i$. We remark that the analogue of Answer 1.4 in this case would be to write x as a sparse sum of just one vector, namely itself. However, there are important differences between this setting and ours, such as requiring the coefficients to be nonnegative, and we do not dwell on the comparison further in this thesis.

We prove Theorem 1.6 in Chapter 3, and we call $\tilde{A}$ a *spectral sparsifier* for A. The procedure we give for finding $\tilde{A}$ is iterative: given $A = \sum_i v_i v_i^T$, it selects some multiple of *one* outer product $v_i v_i^T$ to add to $\tilde{A}$ in each step, and sparsity is ensured by proving that the running sum converges to a matrix with the appropriate spectral properties in a small number of steps, linear in n. The rule for selecting a particular v_i in any step is greedy, and based on two 'barrier' functions which blow up if one tries to choose a v_i which affects the eigenvalues in an undesirable manner. The analysis of why the process works is local in that we just prove that each step by itself brings us closer to approximating A.

At the end of Chapter 3, in Section 3.2, we observe that a simple randomized method for spectral sparsification follows immediately from a probability inequality of Rudelson [39] regarding sums of random rank one matrices. The approximation $\tilde{A}$ produced by that method has a weaker sparsity guarantee of $O(n \log n/\epsilon^2)$ terms, but it can be computed more quickly than the stronger one promised in Theorem 1.6.

In Chapter 4 we motivate and define a spectral notion of approximation for undirected weighted graphs, which corresponds exactly to restricting attention to the class of Laplacian matrices described in Example 1.1. We observe that Theorem 1.6 immediately indicates a way to deterministically construct near–optimal sparsifiers of such graphs. Then we describe a faster algorithm based on the random sampling approach discussed in Section 3.2, which gives sparsifiers of somewhat worse quality but runs in nearly linear time.

In Chapter 5, we prove a refinement of Theorem 1.6 which provides additional guarantees in the case when we are given elementary vectors v_i whose mean is zero. We apply this to prove a sparsification result for John's decompositions of the identity, which characterize the contact points of convex bodies with their minimum volume ellipsoids, and use that to improve a result of Rudelson on the matter [38].

1.2   Restricted Invertibility

The results of this section are best appreciated in the slightly more general setting where we consider an arbitrary $n \times m$ matrix B, which we view as a linear operator from $\mathbb{R}^m$ to $\mathbb{R}^n$. An important step towards understanding the behavior of such a B is to determine whether/where it is invertible, i.e., where the function $x \mapsto Bx$ is injective. It is well known that the invertibilty of B is characterized by the subspaces rowspan$(B) \subseteq \mathbb{R}^m$ and colspan$(B) \subseteq \mathbb{R}^m$, both of dimension rank$(B)$, for which the restricted map B : rowspan$(B) \rightarrow$ colspan(B) is a bijection. We summarize this information as:

Fact 1.7. Every matrix B is invertible precisely on a linear subspace of dimension equal to its rank, and this subspace is spanned by any maximal linearly independent subset of its columns.

However, both rank and invertibility are fragile notions in the sense that adding a tiny amount of noise to a degenerate matrix such as $B = \mathbf{1}\mathbf{1}^T$ immediately makes it full rank and invertible everywhere. A more quantitative and robust measure of invertibility is supplied by the *least singular value*, which is defined for any rectangular matrix as:

$$\sigma_{min}(B) = \min_{\|x\|=1} \|Bx\| = \min_{x \neq 0} \left(\frac{x^T B^T B x}{x^T x} \right)^{1/2} = \sqrt{\lambda_{min}(B^T B)}.$$

The least singular value tells us how much B can shrink a vector in the worst case. Thus a matrix is invertible precisely when $\sigma_{min}(B) > 0$, and *well-invertible* when $\sigma_{min}(B) > c$ for some constant $c \gg 0$.

5

The corresponding quantitative notion of rank turns out to be the *numerical rank*, which is defined as:
$$\text{nrank}(B) = \frac{\|B\|_F^2}{\|B\|_2^2} = \frac{\text{Tr}(B^T B)}{\|B^T B\|_2}.$$

We think of nrank as counting the number of directions in which B is large. Notice that we always have $\text{nrank}(B) \le \text{rank}(B)$, with equality when B is an orthonormal basis. Also observe that this definition is robust against perturbations, so that in the degenerate example mentioned earlier $\text{nrank}(\mathbf{1}\mathbf{1}^T + \text{noise}) \simeq 1$ rather than n.

We may now ask if a quantitative version of Fact 1.7 holds, namely:

Question 1.8. Given a matrix $B_{n \times m}$, is there a subset of columns of size equal to the *numerical rank* so that the restriction of B to these columns is *well-invertible*?

As in the previous section, a partial answer is provided by the spectral theorem.

Answer 1.9. (Well-Invertibility on Large Eigenspace.) Let $A_{n \times n} = BB^T$, so that the eigenvalues λ_i of A are equal to the squares of the singular values, σ_i^2, of B. Recalling that $\text{Tr}(A) = \sum_i \lambda_i$ and that each $\lambda_i \le \|A\|_2$, we can deduce by Markov's inequality that A has at least
$$k = \frac{1}{2}\frac{\text{Tr}(A)}{\|A\|_2} = \frac{1}{2}\text{nrank}(B)$$
eigenvalues $\lambda_1, \dots, \lambda_k$ greater than

$$\theta = \frac{1}{2}\frac{\text{Tr}(A)}{n} = \frac{1}{2}\frac{\|B\|_F^2}{n},$$

which we assume for normalization is at least a constant.

Let $\mathfrak{L} \subset \mathbb{R}^n$ denote the span of the corresponding eigenvectors $u_1, \dots, u_k$, and let $\mathfrak{D}$ be its preimage $B^{-1}\mathfrak{L} \subset \mathbb{R}^m$. Then the restricted linear map $B' : \mathfrak{D} \to \mathfrak{L}$ has singular values $\sigma_1, \dots, \sigma_k$, all greater than $\theta^{1/2}$, and is therefore well-invertible. Moreover, these subspaces have dimension k, which is proportional to the numerical rank of B. $\qquad\qquad\square$

Thus we have shown that every matrix B is well-invertible on a subspace of size proportional to its numerical rank, and this subspace is spanned by the top eigen-vectors of BB^T. Again, the answer provided by the spectral theorem is geometrically satisfying in certain ways — for instance, the subspace $\mathfrak{L}$ which we restrict to is in-variant under BB^T— but it fundamentally disrespects the structure of the matrix B in others. In particular, it does not supply a restriction to *columns* of B as in Fact 1.7.

The truly satisfying answer to Question 1.8 would be an appropriately large subset $S \subset [m]$ of the columns of B for which the *coordinate* restriction

$$B_S : \mathbb{R}^S \to \text{colspan}(B_S)$$

is well-invertible, where B_S denotes the $n \times |S|$ submatrix of B with columns in S. In 1987, Bourgain and Tzafriri [14] proved that this is true upto constant factors

6

when $m = n$ and the columns of B have unit length. Their result was used to great effect in geometric functional analysis and convex geometry. It can be seen as a Ramsey type theorem, in that it says that every matrix satisfying some reasonable properties contains a large submatrix (corresponding to the coordinate subspace) which is 'structured' in the sense that it is close to an isometry by virtue of not shrinking or stretching vectors too much. It can also be viewed as an 'average case vs. worst case' kind of statement, in that it says that if B is well-invertible for an average vector (i.e., has large trace), then there is some coordinate restriction of B which is well-invertible for every vector.

Bourgain and Tzafriri's proof used probabilistic and functional-analytic techniques and was nonconstructive. Moreover, the constant factors it achieved were far from practical — in particular, the bounds were off from those in Answer 1.9 by about 10^{-72} — and so the theorem was not very influential in numerical analysis even though it is qualitatively related to the Column Subset Selection problem and the Rank-Revealing QR Factorization [47]. These constants were improved in recent works by Tropp [47] (who also gave a randomized polynomial time algorithm), Casazza [16], and others in certain regimes, but they remained far from optimal.

The second main theorem of this thesis is an elementary constructive proof of the following generalization of Bourgain and Tzafriri's theorem, with a sharp constant of 1.

Theorem 1.10 (Restricted Invertibility). *Given an $n \times m$ matrix B, there exists a subset $S \subset [m]$ of linearly independent columns of size*

$$|S| \geq (1 - \epsilon)^2 \frac{\|B\|_F^2}{\|B\|_2^2} = (1 - \epsilon)^2 \mathrm{nrank}(B)$$

for which the coordinate restriction B_S has least singular value at least

$$\sigma_{min}^2(B_S) \geq \epsilon^2 \frac{\|B\|_F^2}{m}.$$

Moreover, S can be found deterministically in $O((1 - \epsilon)^2 n^3 m)$ time.

In Chapter 6, we use a variant of the barrier method of Chapter 3 (with one barrier) to prove Theorem 1.10. Theorem 1.6 guarantees the existence of a sparse set of scalars $s_i \geq 0$ for which $\sum_{i \leq m} s_i v_i v_i^T$ approximates A; here, we are able to guarantee that the scalars are are either 0 or 1, corresponding to whether or not a column is chosen to be in S[1]. At a high level, this becomes possible because we are only interested in a lower bound on the spectrum of the vectors we select (this is what is meant by 'well-invertible'), as opposed to both upper and lower bounds as in Theorem 1, and this gives us more freedom to choose the weights s_i.

---

[1]This connection becomes clearer if we state Theorem 1.10 in terms of $A = BB^T = \sum_{i \leq m} b_i b_i^T$, for which it guarantees a subset $S \subset [m]$ of size at least $(1 - \epsilon)^2 \frac{\mathrm{Tr}(A)}{\|A\|_2}$ satisfying $\sum_{i \in S} b_i b_i^T \succeq \epsilon^2 \frac{\mathrm{Tr}(A)}{m} I_{\mathrm{span}(b_i : i \in S)}$.

1.3 Summary

Thus, our two main results can be seen as 'coordinate' versions of Answers 1.4 and 1.9 — they supply comparable quantitative conclusions to those implied by the spectral theorem, but in terms of columns / elementary vectors v_i rather than eigenvectors u_j. They share the same iterative method of proof, which we will refer to as the *barrier method*. Unlike previous approaches, this method immediately yields deterministic polynomial time algorithms. Because of the pliability of the method as well as the central place of linear algebra in mathematics, we are able to apply our results to prove interesting theorems in diverse areas.

1.4 Bibliographic Note

Many of the results presented in this dissertation have been published or submitted for publication. The new results in Chapter 3 are joint work with Joshua Batson and Daniel Spielman, and appeared in [8]. The results of Chapter 4 are joint work with Daniel Spielman and appeared in [41]. Chapter 6 is also joint work with Daniel Spielman, and has been submitted for publication.

Chapter 2

Preliminaries

2.1 Linear Algebra

Most of the proofs in this thesis rely only on elementary facts about symmetric matrices. All of this material can be found in any textbook on linear algebra, such as [27].

We will use uppercase letters such as A, B, X to denote matrices and lowercase letters such as v, w, x to denote vectors. In what follows, A is an $n \times n$ matrix with entries in $\mathbb{R}$ and all vectors are column vectors (matrices of dimension $n \times 1$).

2.1.1 Basic Notions

- **Entries.** We will use parentheses as in $A(i, j)$ and $v(i)$ to denote the entries of matrices and vectors, respectively.

- **Transpose, Norm, Inner Product, and Outer Product.** The transpose of an $m \times n$ matrix X is the $n \times m$ matrix $X^T(i, j) = X(j, i)$.

  If $v, w \in \mathbb{R}^n$ are vectors then the *inner product* is a real number given by

  $$\langle v, w \rangle = v^T w = \sum_{i \leq n} v(i) w(i).$$

  The *Euclidean norm* of a vector is denoted by $\|v\| = \sqrt{v^T v}$.

  The *outer product* vw^T is an $n \times n$ matrix with entries $v(i)w(j)$.

- **Spectral Theorem.** If A is symmetric then there are real numbers $\lambda_1, \ldots \lambda_n$ and orthonormal vectors $u_1, \ldots u_n \in \mathbb{R}^n$ for which

  $$Au_i = \lambda_i u_i$$

  and

  $$A = \sum_{i \leq n} u_i u_i^T.$$

These are called the *eigenvalues* and *eigenvectors*, respectively.

- **Trace.** The *trace* of a matrix is the sum of its diagonal entries; equivalently, the sum of its eigenvalues:

$$\text{Tr}(A) = \sum_{i \leq n} A(i, i) = \sum_{i \leq n} \lambda_i.$$

- **Determinant.** The *determinant* of a matrix is given by

$$\det(A) = \prod_{i \leq n} \lambda_i,$$

the product of the eigenvalues.

- **Characteristic Polynomial.** The *characteristic polynomial*

$$p_A(x) = \det(xI - A)$$

has zeros equal to the eigenvalues $\lambda_1, \ldots \lambda_n$.

- **Image and Kernel.** If A is symmetric then the subspaces $\text{im}(A) = \{Ax : x \in \mathbb{R}^n\}$ and $\ker(A) = \{x : Ax = 0\}$ are orthogonal to each other.

- **Rank.** The rank of A is equal to the number of linearly independent columns of A, the dimension of its image, and the number of nonzero eigenvalues λ_i.

- **Matrix Product.** If A has columns $a_1, \ldots, a_n \in \mathbb{R}^n$ and B has rows $b_1^T, \ldots, b_n^T \in \mathbb{R}^n$, then

$$AB = \sum_{i \leq n} a_i b_i^T.$$

- **Matrix Norms.** The *spectral norm* or *operator norm* of a matrix is given by

$$\|A\|_2 = \sup_{\|x\|=1} \|Ax\|.$$

When A is symmetric then $\|A\|_2 = \max_i \lambda_i$, the largest eigenvalue.

The *Frobenius norm* is given by

$$\|A\|_F = \left(\sum_{i,j \leq n} A(i, j)^2 \right)^{1/2} = \left(\text{Tr}(A^T A) \right)^{1/2} = \left(\sum_{i \leq n} \lambda_i^2 \right)^{1/2}.$$

When we omit the subscript in $\| \cdot \|$, we refer to the spectral norm.

- **Quadratic form and Entrywise Product.** Let

$$A \bullet B = \sum_{i,j \leq n} A(i,j) \cdot B(i,j) = \mathrm{Tr}(A^T B)$$

  denote the *entrywise* product.

  The *quadratic form* of A is the function $v \mapsto v^T A v$ from $\mathbb{R}^n \to \mathbb{R}$, and can be written as

$$v^T A v = A \bullet v v^T.$$

- **Courant-Fisher Theorem.** The eigenvalues $\lambda_1 \leq \lambda_2, \ldots \leq \lambda_n$ of a symmetric matrix A have the following variational characterization:

$$\lambda_i = \max_{\mathcal{S}:\dim(\mathcal{S})=n-i+1} \min_{v \in \mathcal{S} \setminus \{0\}} \frac{v^T A v}{v^T v} = \min_{\mathcal{S}:\dim(\mathcal{S})=i} \max_{v \in \mathcal{S} \setminus \{0\}} \frac{v^T A v}{v^T v},$$

  where $\mathcal{S}$ ranges over subspaces of $\mathbb{R}^n$.

2.1.2   The Pseudoinverse

If $A_{n \times n}$ is symmetric then we can diagonalize it and write

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T$$

where $\lambda_1, \ldots, \lambda_k$ are the nonzero eigenvalues of A and $u_1, \ldots, u_k$ are a corresponding set of orthonormal eigenvectors. The *Moore-Penrose Pseudoinverse* of A is then defined as

$$A^+ = \sum_{i=1}^{k} \frac{1}{\lambda_i} u_i u_i^T.$$

The pseudoinverse behaves as the inverse on $\ker(A)^\perp$. Notice that $\ker(A) = \ker(A^+)$ and that

$$A A^+ = A^+ A = \sum_{i=1}^{k} u_i u_i^T,$$

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

2.1.3   Formulas for Rank–one Updates

The following well-known and easily verified identity describes the behavior of the inverse of a matrix under rank–one updates (see [27, Section 2.1.3]). As many of our constructions involve iteratively building a matrix by adding one outer product at a time, we will use it frequently.

Lemma 2.1 (Sherman–Morrison Formula). *If A is a nonsingular $n \times n$ matrix and $\mathbf{v}$ is a vector, then*

$$(A + \mathbf{v}\mathbf{v}^T)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{v}\mathbf{v}^T A^{-1}}{1 + \mathbf{v}^T A^{-1}\mathbf{v}}.$$

There is a related formula describing the change in the *determinant* of a matrix under the same update:

Lemma 2.2 (Matrix Determinant Lemma). *If A is nonsingular and $\mathbf{v}$ is a vector, then*

$$\det(A + \mathbf{v}\mathbf{v}^T) = \det(A)(1 + \mathbf{v}^T A^{-1}\mathbf{v}).$$

2.1.4 Positive Semidefinite Matrices

In this thesis we will often focus on symmetric *positive semidefinite matrices* over $\mathbb{R}$, which have the following equivalent characterizations.

Definition 2.3. A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is *positive semidefinite (PSD)* if

1. There is a matrix $B \in \mathbb{R}^{n \times m}$ for which $A = BB^T$. Thus we can write

$$A = \sum_{i \leq m} b_i b_i^T$$

   where b_i are the columns of B.

2. A is in the conical hull of the set of rank one symmetric outer products, i.e.

$$A \in \text{cone}\left(vv^T : v \in \mathbb{R}^n\right) = \left\{\sum_{i \leq m} a_i v_i v_i^T : a_i \geq 0, v_i \in \mathbb{R}^n\right\}.$$

3. For every vector $v \in \mathbb{R}^n$ the quadratic form $v^T A v$ is nonnegative.

4. The eigenvalues of A are nonnegative real numbers.

We will denote the cone of $n \times n$ PSD matrices by $\mathcal{S}_n$. This cone comes with a natural partial order given by

$$A \succeq B \qquad \text{iff} \qquad A - B \in \mathcal{S}_n.$$

2.2 Graphs and Laplacians

Let $G = (V, E, w)$ be a connected weighted undirected graph with n vertices and m edges and edge weights $w_e \geq 0$. If we orient the edges of G arbitrarily, we can write

its Laplacian as $L = B^T W B$, where $B_{m \times n}$ is the *signed edge-vertex incidence matrix*, given by

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

and $W_{m \times m}$ is the diagonal matrix with $W(e, e) = w_e$. To put this another way, the row b_e^T of B corresponding to an edge $e = (u, v)$ is given by

$$b_e = \chi_u - \chi_v$$

where χ_u is the canonical basis vector with a 1 in position u[1]. Thus the Laplacian admits an expansion as a sum of outer products

$$L = \sum_{e \in E} w_e b_e b_e^T = \sum_{uv \in E} w_{uv} (\chi_u - \chi_v)(\chi_u - \chi_v)^T.$$

It is immediate that L is positive semidefinite since:

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

and that G is connected if and only if $\ker(L) = \ker(W^{1/2} B) = \text{span}(\mathbf{1})$.

---

[1]We are using χ_u to denote canonical vectors here rather than the more conventional e_i, e_j to avoid confusion with edges named e.

Chapter 3

Spectral Sparsification

In the first section we will prove our main spectral sparsification result, Theorem 1.6, restated shortly for convenience. In the second section we will describe a weaker version of this result, due to Rudelson [39], which is based on random sampling and is computationally more efficient.

3.1 Strong Sparsification

Let us begin by recalling the statement of Theorem 1.6.

Theorem 3.1 (Spectral Sparsification, copy of Theorem 1.6). *Suppose $0 < \epsilon < 1$ and*

$$A = \sum_{i \leq m} v_i v_i^T$$

are given, with $v_i \in \mathbb{R}^n$. Then there are nonnegative weights $\{s_i\}_{i \leq m}$, at most $\lceil n/\epsilon^2 \rceil$ of which are nonzero, for which

$$(1 - \epsilon)^2 A \preceq \tilde{A} = \sum_{i \leq m} s_i v_i v_i^T \preceq (1 + \epsilon)^2 A. \tag{3.1}$$

There is an algorithm which computes the weights s_i in deterministic $O(n^3 m/\epsilon^2)$ time.

For conceptual and notational convenience, we will actually prove the following equivalent statement, which amounts to saying that it is sufficient to consider the case when $A = I$.

Theorem 3.2. *Suppose $d > 1$ and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ are vectors[1] in $\mathbb{R}^n$ with*

$$\sum_{i \leq m} \mathbf{v}_i \mathbf{v}_i^T = I$$

---

[1]We use boldface to denote the vectors $\mathbf{v}_i$ in order to avoid confusion with lowercase letters u, l which denote real numbers in the subsequent proofs.

Then there exist scalars $s_i \geq 0$ with $|\{i : s_i \neq 0\}| \leq dn$ so that

$$I \preceq \sum_{i \leq m} s_i v_i v_i^T \preceq \left(\frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}} \right) I.$$

Theorem 3.1 can be deduced immediately from this.

Proof of Theorem 3.1. Assume without loss of generality that A has full rank and

$$A = \sum_{i \leq m} w_i w_i^T.$$

Setting $v_i = A^{-1/2} w_i$, we see that

$$\sum_i v_i v_i^T = A^{-1/2} \left(\sum_{i \leq m} w_i w_i^T \right) A^{-1/2} = I,$$

satisfying the requirement of Theorem 3.2. We can now set $d = 1/\epsilon^2$ and obtain scalars $s_i \geq 0$, at most $\lceil n/\epsilon^2 \rceil$ nonzero, for which

$$I \preceq \sum_{i \leq m} s_i v_i v_i^T \preceq \left(\frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}} \right) I = \left(\frac{1 + \epsilon}{1 - \epsilon} \right)^2 I.$$

Thus if we take $\tilde{A} = (1 - \epsilon)^2 \sum_i s_i w_i w_i^T$ then

$$I \preceq (1 - \epsilon)^{-2} A^{-1/2} \tilde{A} A^{-1/2} \preceq \left(\frac{1 + \epsilon}{1 - \epsilon} \right)^2 I,$$

which is equivalent to the desired conclusion. $\qquad \square$

The rest of this section is devoted to proving Theorem 3.2. The proof is constructive and yields a deterministic polynomial time algorithm for finding the scalars s_i.

Given vectors $\{v_i\}$, our goal is to choose a small set of coefficients s_i so that $A = \sum_i s_i v_i v_i^T$ is well-conditioned. We will build the matrix A in steps, starting with $A = 0$ and adding one outer product $s_i v_i v_i^T$ at a time. Before beginning the proof, it will be instructive to study how the eigenvalues and characteristic polynomial of a matrix evolve upon the addition of a vector. This discussion should provide some intuition for the structure of the proof, and demystify the origin of the 'Twice-Ramanujan'[2] number $\frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}$ which appears in our final result.

---

[2] See Section 4.1

3.1.1 Intuition for the Proof

It is well known that the eigenvalues of $A + \mathbf{vv}^T$ interlace those of A. In fact, the new eigenvalues can be determined exactly by looking at the characteristic polynomial of $A + \mathbf{vv}^T$, which is computed using Lemma 2.2 as follows:

$$p_{A+\mathbf{vv}^T}(x) = \det(xI - A - \mathbf{vv}^T) = p_A(x)\left(1 - \sum_j \frac{\langle \mathbf{v}, u_j \rangle^2}{x - \lambda_i}\right),$$

where λ_i are the eigenvalues of A and u_j are the corresponding eigenvectors. The polynomial $p_{A+\mathbf{vv}^T}(x)$ has two kinds of zeros λ:

1. Those for which $p_A(\lambda) = 0$. These are equal to the eigenvalues λ_j of A for which the added vector $\mathbf{v}$ is orthogonal to the corresponding eigenvector u_j, and which do not therefore 'move' upon adding $\mathbf{vv}^T$.

2. Those for which $p_A(\lambda) \neq 0$ and

$$f(\lambda) = \left(1 - \sum_j \frac{\langle \mathbf{v}, u_j \rangle^2}{\lambda - \lambda_j}\right) = 0.$$

These are the eigenvalues which have moved and strictly interlace the old eigenvalues. The above equation immediately suggests a simple physical model which gives intuition as to where these new eigenvalues are located.
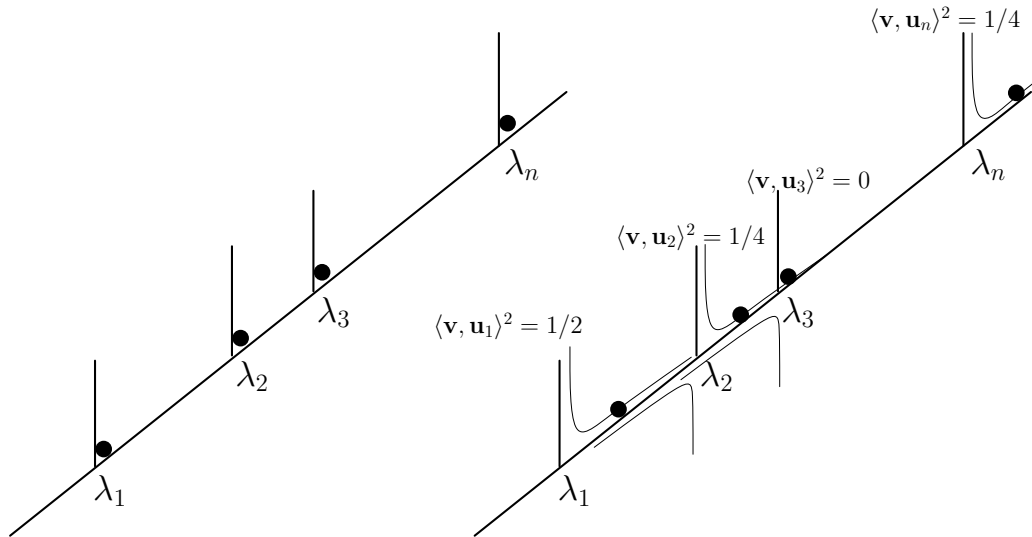

Figure 3.1: Physical model of interlacing eigenvalues.

Physical Model. We interpret the eigenvalues λ as charged particles lying on a slope. On the slope are n fixed, chargeless barriers located at the initial eigenvalues λ_j, and each particle is resting against one of the barriers under

the influence of gravity. Adding the vector $\mathbf{v}\mathbf{v}^T$ corresponds to placing a charge of $\langle \mathbf{v}, u_j \rangle^2$ on the barrier corresponding to λ_j. The charges on the barriers repel those on the eigenvalues with a force that is proportional to the charge on the barrier and inversely proportional to the distance from the barrier — i.e., the force from barrier j is given by

$$\frac{\langle \mathbf{v}, u_j \rangle^2}{\lambda - \lambda_j},$$

a quantity which is positive for λ_j 'below' λ, which are pushing the particle 'upward', and negative otherwise. The eigenvalues move up the slope until they reach an equilibrium in which the repulsive forces from the barriers cancel the effect of gravity, which we take to be a $+1$ in the downward direction. Thus the equilibrium condition corresponds exactly to having the total 'downward pull' $f(\lambda)$ equal to zero.

With this physical model in mind, we begin to consider what happens to the eigenvalues of A when we add a *random* vector from our set $\{\mathbf{v}_i\}$. The first observation is that for any eigenvector u_j (in fact for any vector at all), the expected projection of a randomly chosen $\mathbf{v} \in \{\mathbf{v}_i\}_{i \leq m}$ is

$$\mathbb{E}_\mathbf{v}\langle \mathbf{v}, u_j \rangle^2 = \frac{1}{m}\sum_i \langle \mathbf{v}_i, u_j \rangle^2 = \frac{1}{m}u_j^T \left(\sum_i \mathbf{v}_i\mathbf{v}_i^T \right) u_j = \frac{\|u_j\|^2}{m} = \frac{1}{m}.$$

Of course, this does not mean that there is any single vector $\mathbf{v}_i$ in our set that realizes this 'expected behavior' of equal projections on the eigenvectors. But if we were to add such a vector [3] in our physical model, we would add equal charges of $1/m$ to each of the barriers, and we would expect all of the eigenvalues of A to drift forward 'steadily'. In fact, one might expect that after sufficiently many iterations of this process, the eigenvalues would all march forward together, with no eigenvalue too far ahead or too far behind, and we would end up in a position where $\lambda_{max}/\lambda_{min}$ is bounded.

In fact, this intuition turns out to be correct. Adding a vector with equal projections changes the characteristic polynomial in the following manner:

$$p_{A + \mathbf{v}_{\text{avg}}\mathbf{v}_{\text{avg}}^T}(x) = p_A(x) \left(1 - \sum_j \frac{1/m}{x - \lambda_j} \right) = p_A(x) - (1/m)p_A'(x),$$

---

[3]For concreteness, we remark that this 'average' vector would be precisely

$$\mathbf{v}_{\text{avg}} = \frac{1}{\sqrt{m}} \sum_j u_j.$$

since $p_A'(x) = \sum_j \prod_{i \neq j}(x - \lambda_i)$. If we start with $A = 0$, which has characteristic polynomial $p_0(x) = x^n$, then after k iterations of this process we obtain the polynomial

$$p_k(x) = (I - (1/m)D)^k x^n$$

where D is the derivative with respect to x. Fortunately, iterating the operator $(I - \alpha D)$ for any $\alpha > 0$ generates polynomials which are members of a standard orthogonal family – the *associated Laguerre polynomials* [20]. These polynomials are very well–studied and the locations of their zeros are known; in particular, after $k = dn$ iterations the ratio of the largest to the smallest zero is known [20] to approach

$$\frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}} \qquad \text{as } n \to \infty,$$

which is exactly what we want.

To prove the theorem, we will show that we can choose a sequence of actual vectors that realizes the expected behavior (i.e. the behavior of repeatedly adding $\mathbf{v}_{\text{avg}}$), as long as we are allowed to add arbitrary fractional amounts of the $\mathbf{v}_i \mathbf{v}_i^T$ via the weights $s_i \geq 0$. We will control the eigenvalues of our matrix by maintaining two barriers as in the physical model, and keeping the eigenvalues between them. The lower barrier will 'repel' the eigenvalues forward; the upper one will make sure they do not go too far. The barriers will move forward at a steady pace. By maintaining that the total 'repulsion' at every step of this process is bounded, we will be able to guarantee that there is always some multiple of a vector to add that allows us to continue the process.

3.1.2 Proof by Barrier Functions

We begin by defining two 'barrier' potential functions which measure the quality of the eigenvalues of a matrix. These potential functions are inspired by the inverse law of repulsion in the physical model discussed in the last section.

Definition 3.3. For $u, l \in \mathbb{R}$ and A a symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, define:

$$\Phi^u(A) \stackrel{\text{def}}{=} \text{Tr}(uI - A)^{-1} = \sum_i \frac{1}{u - \lambda_i} \qquad \text{(Upper potential).}$$

$$\Phi_l(A) \stackrel{\text{def}}{=} \text{Tr}(A - lI)^{-1} = \sum_i \frac{1}{\lambda_i - l} \qquad \text{(Lower potential).}$$

As long as $A \prec uI$ and $A \succ lI$ (i.e., $\lambda_{\max}(A) < u$ and $\lambda_{\min}(A) > l$), these potential functions measure how far the eigenvalues of A are from the barriers u and l. In particular, they blow up as any eigenvalue approaches a barrier, since then $uI - A$ (or $A - lI$) approaches a singular matrix. Their strength lies in that they reflect the locations of all the eigenvalues simultaneously: for instance, $\Phi^u(A) \leq 1$ implies that no λ_i is within distance one of u, no 2 λ_i's are at distance 2, no k are at distance k,

and so on. In terms of the physical model, the upper potential $\Phi^u(A)$ is equal to the total repulsion of the eigenvalues of A from the upper barrier u, while $\Phi_l(A)$ is the analogous quantity for the lower barrier.

To prove the theorem, we will build the sum $\sum_i s_i \mathbf{v}_i \mathbf{v}_i^T$ iteratively, adding one vector at a time. Specifically, we will construct a sequence of matrices

$$0 = A^{(0)}, A^{(1)}, \ldots, A^{(Q)}$$

along with positive constants[4] $u_0, l_0, \delta_U, \delta_L, \epsilon_U$ and ϵ_L which satisfy the following conditions:

(a) Initially, the barriers are at $u = u_0$ and $l = l_0$ and the potentials are

$$\Phi^{u_0}(A^{(0)}) = \epsilon_U \quad \text{and} \quad \Phi_{l_0}(A^{(0)}) = \epsilon_L.$$

(b) Each matrix is obtained by a rank–one update of the previous one — specifically by adding a positive multiple of an outer product of some $\mathbf{v}_i$.

$$A^{(q+1)} = A^{(q)} + t\mathbf{v}\mathbf{v}^T \quad \text{for some } \mathbf{v} \in \{\mathbf{v}_i\} \text{ and } t \geq 0.$$

(c) If we increment the barriers u and l by δ_U and δ_L respectively at each step, then the upper and lower potentials do not increase. For every $q = 0, 1, \ldots Q$,

$$\Phi^{u+\delta_U}(A^{(q+1)}) \leq \Phi^u(A^{(q)}) \leq \epsilon_U \quad \text{for } u = u_0 + q\delta_U.$$

$$\Phi_{l+\delta_L}(A^{(q+1)}) \leq \Phi_l(A^{(q)}) \leq \epsilon_L \quad \text{for } l = l_0 + q\delta_L.$$

(d) No eigenvalue ever jumps across a barrier. For every $q = 0, 1, \ldots Q$,

$$\lambda_{\max}(A^{(q)}) < u_0 + q\delta_U \quad \text{and} \quad \lambda_{\min}(A^{(q)}) > l_0 + q\delta_L.$$

To complete the proof we will choose $u_0, l_0, \delta_U, \delta_L, \epsilon_U$ and ϵ_L so that after $Q = dn$ steps, the condition number of $A^{(Q)}$ is bounded by

$$\frac{\lambda_{\max}(A^{(Q)})}{\lambda_{\min}(A^{(Q)})} \leq \frac{u_0 + dn\delta_U}{l_0 + dn\delta_L} = \frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}}.$$

By construction, $A^{(Q)}$ is a weighted sum of at most dn of the vectors, as desired.

The main technical challenge is to show that conditions (b) and (c) can be satisfied simultaneously — i.e., that there is always a choice of $\mathbf{v}\mathbf{v}^T$ to add to the current matrix which allows us to shift *both* barriers up by a constant without increasing either potential. We achieve this in the following three lemmas.

---

[4]On first reading, we suggest the reader follow the proof with the assignment $\epsilon_U = \epsilon_L = 1$, $u_0 = n$, $l_0 = -n$, $\delta_U = 2$, $\delta_L = 1/3$. This will provide the bound $(6d + 1)/(d - 1)$, and eliminates the need to use Claim 3.7.

The first lemma concerns shifting the upper barrier. If we shift u forward to $u + \delta_U$ without changing the matrix A, then the upper potential $\Phi^u(A)$ decreases since the eigenvalues λ_i do not move and u moves away from them. This gives us room to add some multiple of a vector $t\mathbf{v}\mathbf{v}^T$, which will move the λ_i towards l and increase the potential, counteracting the initial decrease due to shifting. The following lemma quantifies exactly how much of a given $\mathbf{v}\mathbf{v}^T$ we can add without increasing the potential beyond its original value before shifting.

Lemma 3.4 (Upper Barrier Shift). *Suppose $\lambda_{\max}(A) < u$, and $\mathbf{v}$ is any vector. If*

$$\frac{1}{t} \geq \frac{\mathbf{v}^T((u + \delta_U)I - A)^{-2}\mathbf{v}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \mathbf{v}^T((u + \delta_U)I - A)^{-1}\mathbf{v} \stackrel{\text{def}}{=} \mathbb{U}_A(\mathbf{v})$$

then

$$\Phi^{u+\delta_U}(A + t\mathbf{v}\mathbf{v}^T) \leq \Phi^u(A) \quad \text{and} \quad \lambda_{\max}(A + t\mathbf{v}\mathbf{v}^T) < u + \delta_U.$$

That is, if we add t times $\mathbf{v}\mathbf{v}^T$ to A and shift the upper barrier by δ_U, then we do not increase the upper potential.

We remark that $\mathbb{U}_A(\mathbf{v})$ is linear in the outer product $\mathbf{v}\mathbf{v}^T$.

Proof. Let $u' = u + \delta_U$. By the Sherman–Morrison formula, we can write the updated potential as:

$$\Phi^{u+\delta_U}(A + t\mathbf{v}\mathbf{v}^T) = \text{Tr}(u'I - A - t\mathbf{v}\mathbf{v}^T)^{-1}$$

$$= \text{Tr}\left((u'I - A)^{-1} + \frac{t(u'I - A)^{-1}\mathbf{v}\mathbf{v}^T(u'I - A)^{-1}}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}\right)$$

$$= \text{Tr}(u'I - A)^{-1} + \frac{t\text{Tr}(\mathbf{v}^T(u'I - A)^{-1}(u'I - A)^{-1}\mathbf{v})}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}$$

$$\text{since Tr is linear and } \text{Tr}(XY) = \text{Tr}(YX)$$

$$= \Phi^{u+\delta_U}(A) + \frac{t\mathbf{v}^T(u'I - A)^{-2}\mathbf{v}}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}$$

$$= \Phi^u(A) - (\Phi^u(A) - \Phi^{u+\delta_U}(A)) + \frac{\mathbf{v}^T(u'I - A)^{-2}\mathbf{v}}{1/t - \mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}$$

As $\mathbb{U}_A(\mathbf{v}) > \mathbf{v}^T(u'I - A)^{-1}\mathbf{v}$, the last term is finite for $1/t \geq \mathbb{U}_A(\mathbf{v})$. By now substituting any $1/t \geq \mathbb{U}_A(\mathbf{v})$ we find $\Phi^{u+\delta_U}(A + t\mathbf{v}\mathbf{v}^T) \leq \Phi^u(A)$. This also tells us that $\lambda_{\max}(A + t\mathbf{v}\mathbf{v}^T) < u + \delta_U$, as if this were not the case, then there would be some positive $t' \leq t$ for which $\lambda_{\max}(A + t'\mathbf{v}\mathbf{v}^T) = u + \delta_U$. But, at such a t', $\Phi^{u+\delta_U}(A + t'\mathbf{v}\mathbf{v}^T)$ would blow up, and we have just established that it is finite. $\square$

The second lemma is about shifting the lower barrier. Here, shifting l forward to $l + \delta_L$ while keeping A fixed has the opposite effect — it increases the lower potential $\Phi_l(A)$ since the barrier l moves towards the eigenvalues λ_i. Adding a multiple of a vector $t\mathbf{v}\mathbf{v}^T$ will move the λ_i forward and away from the barrier, decreasing the

potential. Here, we quantify exactly how much of a given $\mathbf{v}\mathbf{v}^T$ we need to add to compensate for the initial increase from shifting l, and return the potential to its original value before the shift.

Lemma 3.5 (Lower Barrier Shift). *Suppose $\lambda_{\min}(A) > l$, $\Phi_l(A) \leq 1/\delta_L$, and $\mathbf{v}$ is any vector. If*

$$0 < \frac{1}{t} \leq \frac{\mathbf{v}^T(A - (l + \delta_L)I)^{-2}\mathbf{v}}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - \mathbf{v}^T(A - (l + \delta_L)I)^{-1}\mathbf{v} \overset{\text{def}}{=} \mathbb{L}_A(\mathbf{v})$$

then

$$\Phi_{l+\delta_L}(A + t\mathbf{v}\mathbf{v}^T) \leq \Phi_l(A) \quad and \quad \lambda_{\min}(A + t\mathbf{v}\mathbf{v}^T) > l + \delta_L.$$

That is, if we add t times $\mathbf{v}\mathbf{v}^T$ to A and shift the lower barrier by δ_L, then we do not increase the lower potential.

Proof. First, observe that $\lambda_{\min}(A) > l$ and $\Phi_l(A) \leq 1/\delta_L$ imply that $\lambda_{\min}(A) > l + \delta_L$. So, for every $t > 0$, $\lambda_{\min}(A + t\mathbf{v}\mathbf{v}^T) > l + \delta_L$.

Now proceed as in the proof for the upper potential. Let $l' = l + \delta_L$. By Sherman–Morrison, we have:

$$\Phi_{l+\delta_L}(A + t\mathbf{v}\mathbf{v}^T) = \text{Tr}(A + t\mathbf{v}\mathbf{v}^T - l'I)^{-1}$$

$$= \text{Tr}\left((A - l'I)^{-1} - \frac{t(A - l'I)^{-1}\mathbf{v}\mathbf{v}^T(A - l'I)^{-1}}{1 + t\mathbf{v}^T(A - l')^{-1}\mathbf{v}}\right)$$

$$= \text{Tr}(A - l'I)^{-1} - \frac{t\text{Tr}(\mathbf{v}^T(A - l'I)^{-1}(A - l'I)^{-1}\mathbf{v})}{1 + t\mathbf{v}^T(A - l'I)^{-1}\mathbf{v}}$$

$$= \Phi_{l+\delta_L}(A) - \frac{t\mathbf{v}^T(A - l'I)^{-2}\mathbf{v}}{1 + t\mathbf{v}^T(A - l'I)^{-1}\mathbf{v}}$$

$$= \Phi_l(A) + (\Phi_{l+\delta_L}(A) - \Phi_l(A)) - \frac{\mathbf{v}^T(A - l'I)^{-2}\mathbf{v}}{1/t + \mathbf{v}^T(A - l'I)^{-1}\mathbf{v}}$$

Rearranging shows that $\Phi_{l+\delta_L}(A + t\mathbf{v}\mathbf{v}^T) \leq \Phi_l(A)$ when $1/t \leq \mathbb{L}_A(\mathbf{v})$. $\qquad\square$

The third lemma identifies the conditions under which we can find a single $t\mathbf{v}\mathbf{v}^T$ which allows us to maintain both potentials while shifting barriers, and thereby continue the process. The proof that such a vector exists is by an averaging argument, so this can be seen as the step in which we relate the behavior of actual vectors to the behavior of the expected vector $\mathbf{v}_{\text{avg}}$. Notice that the use of variable weights t, from which the eventual s_i arise, is crucial to this part of the proof.

Lemma 3.6 (Both Barriers). *If $\lambda_{\max}(A) < u$, $\lambda_{\min}(A) > l$, $\Phi^u(A) \leq \epsilon_U$, $\Phi_l(A) \leq \epsilon_L$, and $\epsilon_U, \epsilon_L, \delta_U$ and δ_L satisfy*

$$0 \leq \frac{1}{\delta_U} + \epsilon_U \leq \frac{1}{\delta_L} - \epsilon_L \tag{3.2}$$

then there exists an i and positive t for which

$$\mathbb{L}_A(\mathbf{v}_i) \geq 1/t \geq \mathbb{U}_A(\mathbf{v}_i), \quad \lambda_{\max}(A + t\mathbf{v}_i\mathbf{v}_i^T) < u + \delta_U,$$

$$and \quad \lambda_{\min}(A + t\mathbf{v}_i\mathbf{v}_i^T) > l + \delta_L.$$

Proof. We will show that

$$\sum_i \mathbb{L}_A(\mathbf{v}_i) \geq \sum_i \mathbb{U}_A(\mathbf{v}_i),$$

from which the claim will follow by Lemmas 3.4 and 3.5. We begin by bounding

$$\sum_i \mathbb{U}_A(\mathbf{v}_i) = \frac{\sum_i \mathbf{v}_i^T((u + \delta_U)I - A)^{-2}\mathbf{v}_i}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \sum_i \mathbf{v}_i^T((u + \delta_U)I - A)^{-1}\mathbf{v}_i$$

$$= \frac{((u + \delta_U)I - A)^{-2} \bullet (\sum_i \mathbf{v}_i\mathbf{v}_i^T)}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + ((u + \delta_U)I - A)^{-1} \bullet \left(\sum_i \mathbf{v}_i\mathbf{v}_i^T\right)$$

$$= \frac{\text{Tr}((u + \delta_U)I - A)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \text{Tr}((u + \delta_U)I - A)^{-1}$$

$$\qquad \text{since } \sum_i \mathbf{v}_i\mathbf{v}_i^T = I \text{ and } X \bullet I = \text{Tr}(X)$$

$$= \frac{\sum_i(u + \delta_U - \lambda_i)^{-2}}{\sum_i(u - \lambda_i)^{-1} - \sum_i(u + \delta_U - \lambda_i)^{-1}} + \Phi^{u+\delta_U}(A)$$

$$= \frac{\sum_i(u + \delta_U - \lambda_i)^{-2}}{\delta_U \sum_i(u - \lambda_i)^{-1}(u + \delta_U - \lambda_i)^{-1}} + \Phi^{u+\delta_U}(A)$$

$$\leq 1/\delta_U + \Phi^{u+\delta_U}(A),$$

$$\quad \text{as } \sum_i(u - \lambda_i)^{-1}(u + \delta_U - \lambda_i)^{-1} \geq \sum_i(u + \delta_U - \lambda_i)^{-2}$$

$$\leq 1/\delta_U + \Phi^u(A) \leq 1/\delta_U + \epsilon_U.$$

On the other hand, we have

$$\sum_i \mathbb{L}_A(\mathbf{v}_i) = \frac{\sum_i \mathbf{v}_i^T ((A - (l + \delta_L))^{-2} \mathbf{v}_i}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - \sum_i \mathbf{v}_i^T (A - (l + \delta_L)I)^{-1} \mathbf{v}_i$$

$$= \frac{(A - (l + \delta_L)I)^{-2} \bullet (\sum_i \mathbf{v}_i \mathbf{v}_i^T)}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - (A - (l + \delta_L)I)^{-1} \bullet \left(\sum_i \mathbf{v}_i \mathbf{v}_i^T \right)$$

$$= \frac{\mathrm{Tr}(A - (l + \delta_L)I)^{-2}}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - \mathrm{Tr}(A - (l + \delta_L)I)^{-1}$$

$$\text{since } \sum_i \mathbf{v}_i \mathbf{v}_i^T = I \text{ and } X \bullet I = \mathrm{Tr}(X)$$

$$= \frac{\sum_i (\lambda_i - l - \delta_L)^{-2}}{\sum_i (\lambda_i - l - \delta_L)^{-1} - \sum_i (\lambda_i - l)^{-1}} - \sum_i (\lambda_i - l - \delta_L)^{-1}.$$

$$\geq 1/\delta_L - \sum_i (\lambda_i - l)^{-1} = 1/\delta_L - \epsilon_L,$$

by Claim 3.7.

Putting these together, we find that

$$\sum_i \mathbb{U}_A(\mathbf{v}_i) \leq \frac{1}{\delta_U} + \epsilon_U \leq \frac{1}{\delta_L} - \epsilon_L \leq \sum_i \mathbb{L}_A(\mathbf{v}_i),$$

as desired. □

Claim 3.7. *If $\lambda_i > l$ for all i, $0 \leq \sum_i (\lambda_i - l)^{-1} \leq \epsilon_L$, and $1/\delta_L - \epsilon_L \geq 0$, then*

$$\frac{\sum_i (\lambda_i - l - \delta_L)^{-2}}{\sum_i (\lambda_i - l - \delta_L)^{-1} - \sum_i (\lambda_i - l)^{-1}} - \sum_i \frac{1}{\lambda_i - l - \delta_L}$$

$$\geq \frac{1}{\delta_L} - \sum_i \frac{1}{\lambda_i - l}. \tag{3.3}$$

Proof. We have

$$\delta_L \leq 1/\epsilon_L \leq \lambda_i - l,$$

for every i. So, the denominator of the left-most term on the left-hand side is positive,

and the claimed inequality is equivalent to

$$\sum_i (\lambda_i - l - \delta_L)^{-2} \geq \left(\sum_i \frac{1}{\lambda_i - l - \delta_L} - \sum_i \frac{1}{\lambda_i - l} \right) \left(\frac{1}{\delta_L} + \sum_i \frac{1}{\lambda_i - l - \delta_L} - \sum_i \frac{1}{\lambda_i - l} \right)$$

$$= \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right) \left(\frac{1}{\delta_L} + \delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)$$

$$= \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} + \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)^2 ,$$

which, by moving the first term on the RHS to the LHS, is just

$$\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \geq \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)^2 .$$

By Cauchy–Schwartz,

$$\left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)^2 \leq \left(\delta_L \sum_i \frac{1}{\lambda_i - l} \right) \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right)$$

$$\leq (\delta_L \epsilon_L) \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right)$$

$$\text{since } \sum (\lambda_i - l)^{-1} \leq \epsilon_L$$

$$\leq 1 \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right) ,$$

$$\text{since } \frac{1}{\delta_L} - \epsilon_L \geq 0,$$

and so (3.3) is established. $\qquad\square$

Proof of Theorem 3.2. All we need to do now is set $\epsilon_U, \epsilon_L, \delta_U,$ and δ_L in a manner that satisfies Lemma 3.6 and gives a good bound on the condition number. Then, we can take $A^{(0)} = 0$ and construct $A^{(q+1)}$ from $A^{(q)}$ by choosing any vector $\mathbf{v}_i$ with

$$\mathbb{L}_{A^{(q)}}(\mathbf{v}_i) \geq \mathbb{U}_{A^{(q)}}(\mathbf{v}_i)$$

(such a vector is guaranteed to exist by Lemma 3.6) and setting $A^{(q+1)} = A^{(q)} + t\mathbf{v}_i \mathbf{v}_i^T$ for any $t \geq 0$ satisfying:

$$\mathbb{L}_{A^{(q)}}(\mathbf{v}_i) \geq \frac{1}{t} \geq \mathbb{U}_{A^{(q)}}(\mathbf{v}_i).$$

It is sufficient to take

$$\delta_L = 1 \qquad\qquad \epsilon_L = \frac{1}{\sqrt{d}} \qquad\qquad l_0 = -n/\epsilon_L$$

$$\delta_U = \frac{\sqrt{d}+1}{\sqrt{d}-1} \qquad \epsilon_U = \frac{\sqrt{d}-1}{d+\sqrt{d}} \qquad u_0 = n/\epsilon_U.$$

We can check that:

$$\frac{1}{\delta_U} + \epsilon_U = \frac{\sqrt{d}-1}{\sqrt{d}+1} + \frac{\sqrt{d}-1}{\sqrt{d}(\sqrt{d}+1)} = 1 - \frac{1}{\sqrt{d}} = \frac{1}{\delta_L} - \epsilon_L$$

so that (3.2) is satisfied.

The initial potentials are $\Phi^{\frac{n}{\epsilon_U}}(0) = \epsilon_U$ and $\Phi_{\frac{n}{\epsilon_L}}(0) = \epsilon_L$. After dn steps, we have

$$\begin{aligned}\frac{\lambda_{\max}(A^{(dn)})}{\lambda_{\min}(A^{(dn)})} &\le \frac{n/\epsilon_U + dn\delta_U}{-n/\epsilon_L + dn\delta_L}\\ &= \frac{\frac{d+\sqrt{d}}{\sqrt{d}-1} + d\frac{\sqrt{d}+1}{\sqrt{d}-1}}{d-\sqrt{d}}\\ &= \frac{d + 2\sqrt{d} + 1}{d - 2\sqrt{d} + 1},\end{aligned}$$

as desired. □

The Algorithm. To turn this proof into an algorithm, one must first compute the vectors $\mathbf{v}_i$, which can be done in time $O\left(n^2 m\right)$. For each iteration of the algorithm, we must compute $((u + \delta_U)I - A)^{-1}$, $((u + \delta_U)I - A)^{-2}$, and the same matrices for the lower potential function. This computation can be performed in time $O\left(n^3\right)$. Finally, we can decide which vector to add in each iteration by computing $\mathbb{U}_A(\mathbf{v}_i)$ and $\mathbb{L}_A(\mathbf{v}_i)$ for each $\mathbf{v}_i$, which can be done in time $O\left(n^2 m\right)$. As we run for dn iterations, the total time of the algorithm is $O\left(dn^3 m\right)$.

3.1.3 Optimality

It can be shown that the approximation guarantee of Theorem 3.2 is optimal up to a constant factor of four. As the matrices which demonstrate this are the Laplacians of complete graphs, we present them in Section 4.1.2 along with the results on graph sparsification.

3.2 Weak Sparsification by Random Sampling

The results of the previous section more or less settle the question of spectral spar–sification as we have formulated it. However, the method presented is not compu–

tationally practical for large matrices as it requires computing inverses, which is a slow operation. In this section we briefly discuss an alternate approach which was discovered by Rudelson and Vershynin in [40]. The approach, which is based on random sampling and can be computationally faster and more robust, but loses a factor of $\log n$ in sparsity.

The starting point is to recall the Chernoff bound from probability theory (see, for instance, [34]). Suppose Y is a bounded random variable, $|Y| \leq M$, and we take q independent copies of it, $Y_1, \ldots, Y_q$. Then the empirical mean $\tilde{\mu} := \frac{1}{q} \sum_{i \leq q} Y_i$ concentrates very strongly about the mean $\mu := \mathbb{E} Y$:

$$\mathbb{P} \left(\left| \frac{1}{q} \sum_{i \leq q} Y_i - \mathbb{E} Y \right| > \epsilon \right) \leq \exp \left(\frac{-q \epsilon^2}{4 M^2} \right).$$

Thus, if we take $q = 8 M^2 / \epsilon^2$ samples, we can expect $\tilde{\mu}$ to be within ϵ of μ with probability at least $1 - \exp(-2)$. This can be interpreted as a kind of sparsification, since μ may be a sum of many (even infinitely many) terms whereas $\tilde{\mu}$ has only q terms.

It was discovered by Rudelson [39] and later by Ahlswede and Winter [3] that the same phenomenon continues to hold if Y is a random $n \times n$ matrix, but with a (necessary) blowup of $O(\log n)$ in the number of samples q that is required to attain concentration. Here is the slightly stronger version described in lecture notes of Vershynin [50]; note the normalization $\mathbb{E} Y = I$.

Theorem 3.8. *Suppose Y is a random $n \times n$ matrix satisfying $\mathbb{E} Y = I_n$ and $\|Y\|_2 \leq M$. If $Y_1, \ldots, Y_q$ are independent samples of Y, then*

$$\mathbb{P} \left(\left\| \frac{1}{q} \sum_{i \leq q} Y_i - I_n \right\|_2 > \epsilon \right) \leq n \cdot \exp \left(\frac{-q \epsilon^2}{4 M} \right).$$

Thus, taking $q = 8 M \log n / \epsilon^2$ samples of Y should suffice to obtain a good approximation to I. This suggests a strategy for sparsification which we formalize in the following theorem, analogous to Theorem 3.2.

Theorem 3.9. *Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are vectors in $\mathbb{R}^n$ with*

$$\sum_{i \leq m} \mathbf{v}_i \mathbf{v}_i^T = I$$

and $0 < \epsilon < 1$ is a constant. Let Y be the random rank one matrix given by

$$Y = \frac{n}{\|\mathbf{v}_i\|_2^2} \mathbf{v}_i \mathbf{v}_i^T \qquad \text{with probability } p_i \propto \|\mathbf{v}_i\|^2.$$

If we take $q = 8n \log n / \epsilon^2$ independent samples $Y_1, \ldots, Y_q$ then

$$(1 - \epsilon)I \preceq \frac{1}{q} \sum_{i \leq q} Y_i \preceq (1 + \epsilon)I$$

with probability at least $1/2$.

Proof. Note that $\sum_i \|\mathbf{v}_i\|^2 = \text{Tr}(I) = n$, so that $p_i = \|\mathbf{v}_i\|^2/n$. Thus

$$\mathbb{E}Y = \sum_i \frac{\|\mathbf{v}_i\|^2}{n} \frac{n}{\|\mathbf{v}_i\|^2} \mathbf{v}_i \mathbf{v}_i^T = I$$

and invoking Theorem 3.8 with $M = n$ gives the desired conclusion. $\square$

An important feature of this sampling procedure is that it is quite robust with regards to the sampling probabilities p_i; in particular, any constant–factor approximation to the distribution $p_i \propto \|v_i\|^2$ also works.

Corollary 3.10. *Suppose $\sum_{i \leq m} \mathbf{v}_i \mathbf{v}_i^T = I$, $\{p_i\}_{i \leq m}$ are given by $p_i \propto \|\mathbf{v}_i\|^2$, and q_i are numbers satisfying $q_i \geq p_i / \alpha$ and $\sum_{i \leq m} q_i \leq \alpha$ for some factor $\alpha \geq 1$. Then taking $q = 8\alpha^2 n \log n / \epsilon^2$ independent samples $Y_1, \ldots, Y_q$ from the distribution*

$$Y = \frac{\sum_i q_i}{q_i} \mathbf{v}_i \mathbf{v}_i^T \qquad \text{with probability } q_i.$$

yields a sparsifier in that

$$(1 - \epsilon)I \preceq \frac{1}{q} \sum_{i \leq q} Y_i \preceq (1 + \epsilon)I$$

with probability at least $1/2$.

Proof. Following the proof of Theorem 3.9, the norm of Y is now bounded by

$$\left\| \frac{\sum_i q_i}{q_i} \mathbf{v}_i \mathbf{v}_i^T \right\| \leq \left\| \frac{\alpha \sum_i p_i}{(p_i/\alpha)} \mathbf{v}_i \mathbf{v}_i^T \right\| \leq \alpha^2 n$$

so we need to take $M = \alpha^2 n$ instead of just n. $\square$

The flexibility of using approximate probabilities for sampling is important in applications, where the exact $\|v_i\|^2$ may be difficult to compute. This is, in fact, the case in the application to graph sparsification which we will see at the end of the next chapter.

3.3 Other Notions of Matrix Approximation

There is a large body of work on sparse [6, 2] and low-rank [23, 2, 40, 21, 22] approximations for general matrices. The algorithms in this literature provide guarantees of the form $\|A - \tilde{A}\|_2 \leq \epsilon$, where A is the original matrix and $\tilde{A}$ is obtained by entrywise or columnwise sampling of A. This is analogous to satisfying (3.1) only for vectors x in the span of the dominant eigenvectors of A, and corresponds to an 'additive' rather than a multiplicative approximation.

The above constructions have the advantage of being simple and quickly computable, but are not as useful in many contexts where much of the interesting information lies in the smaller eigenspaces. For instance, if we were to use these sparsifiers on Laplacian matrices of graphs (discussed in the next chapter), they would only preserve the large cuts whereas our spectral sparsifiers would preserve all the cuts.

Chapter 4

Sparsification of Graphs

In this chapter we will apply the spectral sparsification theorems of Chapter 3 to the class of matrices corresponding to Laplacians of undirected weighted graphs; more concretely, this is exactly the cone of matrices generated by

$$\left\{ (e_i - e_j)(e_i - e_j)^T : i, j \in [n], i \neq j \right\}$$

as discussed in Example 1.1. Let us begin by motivating the problem of graph sparsification and examining why the spectral notion of approximation for Laplacian matrices corresponds to a combinatorially interesting notion of approximation for graphs.

A sparsifier of a graph $G = (V, E, w)$ is a sparse graph H that is similar to G in some useful way. Many notions of similarity have been considered. For example, Chew's [18] spanners have the property that the distance between every pair of vertices in H is approximately the same as in G. Benczur and Karger's [9] cut-sparsifiers have the property that the weight of the boundary of every set of vertices is approximately the same in G as in H. The spectral notion of similarity which we consider was first introduced by Spielman and Teng [42, 45]: we say that H is a κ-approximation of G if for all $x \in \mathbb{R}^V$,

$$x^T L_G x \leq x^T L_H x \leq \kappa \cdot x^T L_G x, \tag{4.1}$$

where L_G and L_H are the Laplacian matrices of G and H. This matches exactly our definition from Chapter 1 of κ-approximation for PSD matrices.

Recall from Chapter 2 that

$$x^T L_G x = \sum_{(u,v) \in E} w_{u,v} (x_u - x_v)^2,$$

where $w_{u,v}$ is the weight of edge (u, v) in G. By considering vectors x that are the characteristic vectors of sets, one can see that condition (4.1) is strictly[1] stronger than the cut condition of Benczur and Karger. It is worth mentioning that although

---

[1]To see strictness, take G to be a cycle on $[n]$ and H to be a path on $[n]$. Then every cut in H is within half of its weight in G, but the vector $x = (1, 2, \ldots, n)$ witnesses the fact that H is not a spectral sparsifier for G.

the spectral condition is stronger, it can be verified easily in polynomial time by computing the eigenvalues of $(L_H^+)^{1/2}L_G(L_H^+)^{1/2}$, whereas this is not clear for the cut condition as there are exponentially many cuts.

Theoretical Motivation. Spectral graph sparsifiers preserve many interesting properties of graphs, in addition to the weights of cuts. The Courant–Fischer Theorem tells us that

$$\lambda_i = \max_{S:\dim(S)=k} \min_{x \in S} \frac{x^T L x}{x^T x}.$$

Thus, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of L_G and $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ are the eigenvalues of L_H, then we have

$$\lambda_i \leq \tilde{\lambda}_i \leq \kappa \lambda_i,$$

and the eigenspaces spanned by corresponding eigenvalues are related. As the eigenvalues of the normalized Laplacian are given by

$$\lambda_i = \max_{S:\dim(S)=k} \min_{x \in S} \frac{x^T D^{-1/2} L D^{-1/2} x}{x^T x},$$

and are the same as the eigenvalues of the walk matrix $D^{-1}L$, we obtain the same relationship between the eigenvalues of the walk matrix of the original graph and its sparsifier. Many properties of graphs and random walks are known to be revealed by their spectra (see for example any text on Spectral Graph Theory [13, 19, 25]). While the existence of sparse subgraphs which retain these properties is interesting its own right, we also expect it to be useful in theoretical applications. We remark that the study of sparse approximations of the complete graph, otherwise known as expanders, has been hugely successful over the past few decades, with significant impact in such diverse fields as robust network design, coding theory, and derandomization. We will discuss expanders in more detail in Section 4.1.

Practical Motivation. There is also a practical reason for studying graph sparsification: If H is close to G in some appropriate metric, then H can be used as a proxy for G in computations without introducing too much error. At the same time, since H has very few edges, computation with and storage of H should be cheaper. Thus a fast algorithm for graph sparsification can be used as a preprocessing step to speed up computations on graphs. This is an especially useful primitive to have given the increasingly large graphs we encounter in practice, for instance from internet and scientific data. Here are three outstanding examples of how different kinds of fast graph sparsification have been used to speed up computations.

1. Thorup and Zwick [46] gave a fast random sampling algorithm for constructing *multiplicative spanners*; specifically, on input G and k their algorithm produces a subgraph $H \subset G$ with $O(n^{1+1/k})$ edges for which

$$\text{dist}_G(u, v) \leq \text{dist}_H(u, v) \leq k \cdot \text{dist}_G(u, v) \qquad \forall u, v \in G,$$

where dist is the shortest path distance. They used these spanners to construct

fast and compact oracles for pairwise distance queries between vertices in a graph.

2. Spielman and Teng [45] gave a algorithm which on input G with m edges produces in $\tilde{O}(m)$ time a spectral $(1+\epsilon)$-approximation H with $O(n \log^c n)$ edges for some large constant c. They used these sparsifiers to construct preconditioners for symmetric diagonally-dominant matrices, which led to the first nearly-linear time solvers for such systems of equations.

3. Benczur and Karger gave a nearly-linear time procedure which takes a graph G and a parameter $\epsilon > 0$, and outputs a weighted subgraph H with $O(n \log n/\epsilon^2)$ edges such that the weight of every cut in H is within a factor of $(1 \pm \epsilon)$ of its weight in G. This was used to turn Goldberg and Tarjan's $\widetilde{O}(mn)$ max-flow algorithm [26] into an $\widetilde{O}(n^2)$ algorithm for approximate st-mincut, and appeared more recently as the first step of an $\widetilde{O}(n^{3/2} + m)$-time $O(\log^2 n)$ approximation algorithm for sparsest cut [32].

Two Sparsification Algorithms. The strong and weak sparsification theorems of Chapter 3 yield graph sparsification algorithms with very different characteristics. Roughly speaking, the first one is mainly of theoretical interest while the second is more likely to be useful in practice.

- Theorem 3.1 implies that every G has, for every $\epsilon > 0$, a weighted subgraph[2] H with $\lceil n/\epsilon^2 \rceil$ edges for which

$$(1 - \epsilon)^2 L_G \preceq L_H \preceq (1 + \epsilon)^2 L_G.$$

  The procedure given in the proof of that theorem shows that H can be computed in deterministic time $O(mn^3/\epsilon^2)$, which is prohibitively slow for large graphs.

  The significance of these sparsifiers lies in their small size — they have constant average degree for any constant ϵ, and can thus be seen as generalizations of expander graphs, which are constant degree approximations of the complete graph. In Section 4.1 we dwell on this analogy, and prove that our bounds cannot be improved by more than a factor of 4.

- Theorem 3.9 shows that every G has, for every $\epsilon > 0$, a weighted subgraph H with $O(n \log n/\epsilon^2)$ edges which is a spectral $(1 + \epsilon)$-approximation. This construction is randomized and involves sampling the edges of G independently with certain appropriate probabilities p_e, which turn out to be exactly the *effective resistances* of the edges. In Section 4.2, we show how to compute the effective resistances in nearly linear time, leading to an algorithm which is both fast and conceptually simple, although it is not optimal with respect to the sparsity/approximation ratio.

---

[2]i.e., the set of edges of H is a subset of the edges of G, but the weights are different

4.1 Twice–Ramanujan Sparsifiers

In the case where G is the complete graph, excellent spectral sparsifiers are supplied by *Ramanujan Graphs* [33, 35]. These are d-regular graphs H all of whose non-zero Laplacian eigenvalues lie between $d - 2\sqrt{d-1}$ and $d + 2\sqrt{d-1}$. Thus, if we take a Ramanujan graph on n vertices and multiply the weight of every edge by $n/(d - 2\sqrt{d-1})$, we obtain a graph that κ-approximates the complete graph, for

$$\kappa = \frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}}.$$

In this section, we observe that Theorem 3.1 implies that every graph can be approximated at least this well[3] by a graph with only twice as many edges as the Ramanujan graph (as a d-regular graph has $dn/2$ edges). The following is essentially a direct restatement of Theorem 3.1, but with the substitution $d = 1/\epsilon^2$ to emphasize the average degree of the graphs produced.

Theorem 4.1. *For every $d > 1$, every undirected weighted graph $G = (V, E, w)$ on n vertices contains a weighted subgraph $H = (V, F, \tilde{w})$ with $\lceil d(n-1) \rceil$ edges (i.e., average degree at most $2d$) that satisfies:*

$$x^T L_G x \leq x^T L_H x \leq \left(\frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}} \right) \cdot x^T L_G x \qquad \forall x \in \mathbb{R}^V.$$

We remark that while the edges of H are a subset of the edges of G, the weights of edges in H and G will typically be different. In fact, there exist unweighted graphs G for which every good spectral sparsifier H must contain edges of widely varying weights [45].

The significance of Ramanujan graphs lies in their exact optimality: no $d-$regular graph can exhibit sharper concentration of eigenvalues (or larger spectral gap) [36]. All known constructions of Ramanujan graphs are number-theoretic and rely on heavy tools from algebra to prove their correctness. There have been many attempts (e.g. [37, 12]) at elementary constructions of such graphs, but none have come within a constant factor of the optimal dependence on d. In the remainder of this section, we argue that Theorem 4.1 should be seen as a step towards this goal. First we discuss why the weighted graphs produced by our construction should be considered expanders at all. Next, we extend the Alon–Boppana bound [36] for the extremality of Ramanujan graphs to include weighted graphs.

---

[3]Strictly speaking, our approximation constant is only better than the Ramanujan bound $\kappa = \frac{d+2\sqrt{d-1}}{d-2\sqrt{d-1}}$ in the regime $d \geq \frac{1+\sqrt{5}}{2}$. This includes the actual Ramanujan graphs, for which d is an integer greater than 2.

4.1.1 Expanders: Sparsifiers of the Complete Graph

In the case that G is a complete graph, our construction produces graphs H that are expanders. However, these expanders are slightly unusual in that their edges have weights, they may be irregular, and the weighted degrees of vertices can vary slightly. This may lead one to ask whether they should really be considered expanders.

In this section, we argue that they should be, as they may easily be shown to have the properties that define expanders. In particular, H has high edge-conductance, random walks mix rapidly on H and converge to an almost-uniform distribution, and H satisfies the Expander Mixing Property (see [5] or [30, Lemma 2.5]). High edge-conductance and rapid mixing would not be so interesting if the weighted degrees were not nearly uniform — for example, the star graph has both of these properties, but the random walk on the star graph converges to a very non-uniform distribution, and the star does not satisfy the Expander Mixing Property. For the convenience of the reader, we include a proof that H has the Expander Mixing Property below.

Lemma 4.2. *Let $L_H = (V, E, w)$ be a graph that $(1 + \epsilon)$-approximates L_G, the complete graph on V. Then, for every pair of disjoint sets S and T,*

$$\left| w(S, T) - \left(1 + \frac{\epsilon}{2}\right) |S| |T| \right| \leq n(\epsilon/2)\sqrt{|S| |T|},$$

where $w(S, T)$ denotes the sum of the weights of edges between S and T.

Proof. We have

$$-\frac{\epsilon}{2} L_G \preceq L_H - \left(1 + \frac{\epsilon}{2}\right) L_G \preceq \frac{\epsilon}{2} L_G,$$

so we can write

$$L_H = \left(1 + \frac{\epsilon}{2}\right) L_G + M,$$

where M is a matrix of norm at most $(\epsilon/2) \|L_G\| \leq n\epsilon/2$. Let x be the characteristic vector of S, and let y be the characteristic vector of T. We have

$$-w(S, T) = x^T L_H y.$$

As G is the complete graph and S and T are disjoint, we also know

$$x^T L_G y = -|S| |T|.$$

Thus,

$$x^T L_H y = \left(1 + \frac{\epsilon}{2}\right) x^T L_G y + x^T M y$$

$$= -\left(1 + \frac{\epsilon}{2}\right) |S| |T| + x^T M y.$$

The lemma now follows by observing that

$$x^T M y \leq \|M\| \|x\| \|y\| \leq n(\epsilon/2)\sqrt{|S| |T|}.$$

4.1.2 A Lower Bound

As the graphs we produce are irregular and weighted, it is also not immediately clear that we should be comparing κ with the Ramanujan bound of

$$\frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}} = 1 + \frac{4}{\sqrt{d}} + O(1/d). \tag{4.2}$$

It is known[4] that no d-regular graph of uniform weight can κ-approximate a complete graph for κ asymptotically better than (4.2) [36]. While we believe that no graph of *average* degree d can be a κ-approximation of a complete graph for κ asymptotically better than (4.2), we are unable to show this at the moment and prove instead the following weaker claim.

Proposition 4.3. *Let G be the complete graph on vertex set V, and let $H = (V, E, w)$ be a weighted graph with n vertices and a vertex of degree d. If H κ-approximates G, then*

$$\kappa \geq 1 + \frac{2}{\sqrt{d}} - O\left(\frac{\sqrt{d}}{n}\right).$$

Proof. We use a standard approach. Suppose H is a κ-approximation of the complete graph. We will construct vectors x^* and y^* orthogonal to the $\mathbf{1}$ vector so that

$$\frac{y^{*T} L_H y^*}{x^{*T} L_H x^*} \frac{\|x^*\|^2}{\|y^*\|^2}$$

is large, and this will give us a lower bound on κ.

Let v_0 be the vertex of degree d, and let its neighbors be $v_1, \ldots, v_d$. Suppose v_i is connected to v_0 by an edge of weight w_i, and the total weight of the edges between v_i and vertices other than $v_0, v_1, \ldots, v_d$ is δ_i. We begin by considering vectors x and y with

$$x(u) = \begin{cases} 1 & \text{for } u = v_0, \\ 1/\sqrt{d} & \text{for } u = v_i, \ i \geq 1, \\ 0 & \text{for } u \notin \{v_0, \ldots, v_d\} \end{cases}$$

---

[4]While lower bounds [36] on the spectral gap of d-regular graphs focus on showing that the second-smallest eigenvalue is asymptotically at most $d - 2\sqrt{d-1}$, the same proofs by test functions can be used to show that the largest eigenvalue is at asymptotically least $d + 2\sqrt{d-1}$.

$$y(u) = \begin{cases} 1 & \text{for } u = v_0, \\ -1/\sqrt{d} & \text{for } u = v_i, \, i \geq 1, \\ 0 & \text{for } u \notin \{v_0, \dots, v_d\} \end{cases}$$

These vectors are not orthogonal to $\mathbf{1}$, but we will take care of that later. It is easy to compute the values taken by the quadratic form at x and y:

$$x^T L_H x = \sum_{i=1}^{d} w_i (1 - 1/\sqrt{d})^2 + \sum_{i=1}^{d} \delta_i (1/\sqrt{d} - 0)^2$$

$$= \sum_{i=1}^{d} w_i + \sum_{i=1}^{d} (\delta_i + w_i)/d - 2 \sum_{i=1}^{d} w_i/\sqrt{d}$$

and

$$y^T L_H y = \sum_{i=1}^{d} w_i (1 + 1/\sqrt{d})^2 + \sum_{i=1}^{d} \delta_i (-1/\sqrt{d} - 0)^2$$

$$= \sum_{i=1}^{d} w_i + \sum_{i=1}^{d} (\delta_i + w_i)/d + 2 \sum_{i=1}^{d} w_i/\sqrt{d}.$$

The ratio in question is thus

$$\frac{y^T L_H y}{x^T L_H x} = \frac{\sum_i w_i + \sum_i (\delta_i + w_i)/d + 2 \sum_i w_i/\sqrt{d}}{\sum_i w_i + \sum_i (\delta_i + w_i)/d - 2 \sum_i w_i/\sqrt{d}}$$

$$= \frac{1 + \frac{1}{\sqrt{d}} \frac{2 \sum_i w_i}{\sum_i w_i + \sum_i (\delta_i + w_i)/d}}{1 - \frac{1}{\sqrt{d}} \frac{2 \sum_i w_i}{\sum_i w_i + \sum_i (\delta_i + w_i)/d}}.$$

Since H is a κ-approximation, all weighted degrees must lie between n and $n\kappa$, which gives

$$\frac{2 \sum_i w_i}{\sum_i w_i + \sum_i (\delta_i + w_i)/d} = \frac{2}{1 + \frac{\sum_i (\delta_i + w_i)/d}{\sum_i w_i}} \geq \frac{2}{1 + \kappa}.$$

Therefore,

$$\frac{y^T L_H y}{x^T L_H x} \geq \frac{1 + \frac{1}{\sqrt{d}} \frac{2}{1+\kappa}}{1 - \frac{1}{\sqrt{d}} \frac{2}{1+\kappa}}. \tag{4.3}$$

Let x^* and y^* be the projections of x and y respectively orthogonal to the $\mathbf{1}$ vector. Then

$$\|x^*\|^2 = \|x\|^2 - \langle x, \mathbf{1}/\sqrt{n} \rangle^2 = 2 - \frac{(1 + \sqrt{d})^2}{n}$$

and

$$\|y^*\|^2 = \|y\|^2 - \langle y, \mathbf{1}/\sqrt{n}\rangle^2 = 2 - \frac{(1 - \sqrt{d})^2}{n}$$

so that as $n \to \infty$

$$\frac{\|x^*\|^2}{\|y^*\|^2} = 1 - O\left(\frac{\sqrt{d}}{n}\right). \tag{4.4}$$

Combining (4.3) and (4.4), we conclude that asymptotically:

$$\frac{y^{*T} L_H y^*}{x^{*T} L_H x^*} \frac{\|x^*\|^2}{\|y^*\|^2} \geq \frac{1 + \frac{1}{\sqrt{d}}\frac{2}{1+\kappa}}{1 - \frac{1}{\sqrt{d}}\frac{2}{1+\kappa}} \left(1 - O\left(\frac{\sqrt{d}}{n}\right)\right)$$

But by our assumption the LHS is at most κ, so we have

$$\kappa \geq \frac{1 + \frac{1}{\sqrt{d}}\frac{2}{1+\kappa}}{1 - \frac{1}{\sqrt{d}}\frac{2}{1+\kappa}} \left(1 - O\left(\frac{\sqrt{d}}{n}\right)\right)$$

which on rearranging gives

$$\kappa \geq 1 + \frac{2}{\sqrt{d}} - O\left(\frac{\sqrt{d}}{n}\right)$$

as desired. $\qquad \square$

4.2   Sparsification by Effective Resistances

In this section, we show how to construct spectral sparsifiers with $O(n \log n/\epsilon^2)$ edges in nearly–linear time, thus improving on both [10] and [43]. Our sparsifiers are sub–graphs of the original graph and can be computed in $\widetilde{O}(m)$ time by random sampling, where the sampling probabilities are given by the effective resistances of the edges in the graph.

Suppose G is a connected weighted graph with Laplacian $L_G = \sum_{e \in E} w_e b_e b_e^T$, where the b_e are incidence vectors of edges. The existence of a sparsifier H with $O(n \log n/\epsilon^2)$ edges follows immediately by applying Theorem 3.9 to the vectors

$$v_e = w_e^{1/2}(L^+)^{\frac{1}{2}} b_e$$

for which it is easily checked that

$$\sum_e v_e v_e^T = (L^+)^{\frac{1}{2}} \left(\sum_e w_e b_e b_e^T\right) (L^+)^{\frac{1}{2}} = (L^+)^{\frac{1}{2}} L (L^+)^{\frac{1}{2}} = I_{\mathrm{im}(L)}.$$

In particular, Theorem 3.9 tells us to sample the vectors v_e, which correspond to edges

36

of G with, probabilities $p_e \propto \|v_e\|^2$. In Section 4.2.1, we observe that these lengths correspond to the *effective resistances* R_e of edges in G (multiplied by their edge weights w_e), establishing the following conceptually simple and satisfying description of the sparsification algorithm.

$H = $ **ReffSparsify**(G)

Let $q = 8n \log n / \epsilon^2$ as in Theorem 3.9. Choose a random edge e of G with probability p_e proportional to $w_e R_e$, and add e to H with weight $w_e / q p_e$. Take q samples independently with replacement, summing weights if an edge is chosen more than once.

To turn this into a *fast* algorithm, we must compute the effective resistances R_e efficiently. In Section 4.2.2, we show how to compute approximate effective resistances in nearly-linear time, which is essentially optimal. The tools we use to do this are Spielman and Teng's nearly-linear time solver [43, 44] and the Johnson–Lindenstrauss Lemma [31, 1]. Specifically, we prove the following theorem, in which R_{uv} denotes the effective resistance between vertices u and v.

Theorem 4.4. *There is an $\widetilde{O}(m \log r / \epsilon^2)$ time algorithm which on input $\epsilon > 0$ and $G = (V, E, w)$ with $r = w_{max} / w_{min}$ computes a $(24 \log n / \epsilon^2) \times n$ matrix $\widetilde{Z}$ such that with probability at least $1 - 1/n$*

$$(1 - \epsilon) R_{uv} \leq \|\widetilde{Z}(\chi_u - \chi_v)\|^2 \leq (1 + \epsilon) R_{uv}$$

for every pair of vertices $u, v \in V$.

Since $\widetilde{Z}(\chi_u - \chi_v)$ is simply the difference of the corresponding two columns of $\widetilde{Z}$, we can query the approximate effective resistance between any pair of vertices (u, v) in time $O(\log n / \epsilon^2)$, and for all the edges in time $O(m \log n / \epsilon^2)$. By Corollary 3.10, this yields an $\widetilde{O}(m \log r / \epsilon^2)$ time for sparsifying graphs, as advertised.

4.2.1 Electrical Flows

Let us begin by identifying $G = (V, E, w)$ with an electrical network on n nodes in which each edge e corresponds to a link of conductance w_e (i.e., a resistor of resistance $1/w_e$), oriented arbitrarily as as in Section 2.2. We will use the same notation as [29] to describe electrical flows on graphs: for a vector $\mathbf{i}_{\text{ext}}(u)$ of currents injected at the vertices, let $\mathbf{i}(e)$ be the currents induced in the edges (in the direction of orientation) and $\mathbf{v}(u)$ the potentials induced at the vertices. By Kirchoff's current law, the sum of the currents entering a vertex is equal to the amount injected at the vertex:

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

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

$$\mathbf{i} = W B \mathbf{v}.$$

Combining these two facts, we obtain

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

If $\mathbf{i}_{\text{ext}} \perp \text{span}(\mathbf{1}) = \ker(L)$ — i.e., if the total amount of current injected is equal to the total amount extracted — then we can write

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

by the definition of L^+ in Section 2.1.2.

Recall that the *effective resistance* between two vertices u and v is defined as the potential difference induced between them when a unit current is injected at one and extracted at the other. We will derive an algebraic expression for the effective resistance in terms of L^+. To inject and extract a unit current across the endpoints of an edge $e = (u, v)$, we set $\mathbf{i}_{\text{ext}} = b_e = (\chi_v - \chi_u)$, which is clearly orthogonal to $\mathbf{1}$. The potentials induced by $\mathbf{i}_{\text{ext}}$ at the vertices are given by $\mathbf{v} = L^+b_e$; to measure the potential difference across $e = (u, v)$, we simply multiply by b_e on the left:

$$\mathbf{v}(v) - \mathbf{v}(u) = (\chi_v - \chi_u)^T\mathbf{v} = b_e^T L^+ b_e.$$

It follows that the effective resistance across e is given by $R_e = b_e^T L^+ b_e = \|(L^+)^{\frac{1}{2}} b_e\|^2$. Thus we conclude that

$$\|v_e\|^2 = \|w_e^{1/2}(L^+)^{\frac{1}{2}} b_e\|^2 = w_e R_e,$$

as desired.

4.2.2   Computing Approximate Resistances Quickly

It is not clear how to compute all the effective resistances $\{R_e\}$ exactly and efficiently. In this section, we show that one can compute constant factor approximations to all the R_e in time $\widetilde{O}(m \log r)$. In fact, we do something stronger: we build a $O(\log n) \times n$ matrix $\widetilde{Z}$ from which the effective resistance between any two vertices (including vertices not connected by an edge) can be computed in $O(\log n)$ time.

Proof of Theorem 4.4. If u and v are vertices in G, then following the discussion in the previous section, the effective resistance between u and v can be written as:

$$\begin{aligned}
R_{uv} &= (\chi_u - \chi_v)^T L^+(\chi_u - \chi_v) \\
&= (\chi_u - \chi_v)^T L^+ L L^+(\chi_u - \chi_v) \\
&= ((\chi_u - \chi_v)^T L^+ B^T W^{1/2})(W^{1/2} B L^+(\chi_u - \chi_v)) \\
&= \|W^{1/2} B L^+(\chi_u - \chi_v)\|_2^2.
\end{aligned}$$

Thus effective resistances are just pairwise distances between vectors in $\{W^{1/2} B L^+ \chi_v\}_{v \in V}$. By the Johnson–Lindenstrauss Lemma, these distances are preserved if we project the

vectors onto a subspace spanned by $O(\log n)$ random vectors. For concreteness, we use the following version of the Johnson-Lindenstrauss Lemma due to Achlioptas [1].

Lemma 4.5. *Given fixed vectors $v_1 \ldots v_n \in \mathbb{R}^d$ and $\epsilon > 0$, let $Q_{k \times d}$ be a random $\pm 1/\sqrt{k}$ matrix (i.e., independent Bernoulli entries) with $k \geq 24 \log n / \epsilon^2$. Then with probability at least $1 - 1/n$*

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

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

Our goal is now to compute the projections $\{QW^{1/2}BL^+\chi_v\}$. We will exploit the linear system solver of Spielman and Teng [43, 44], which we recall satisfies:

Theorem 4.6 (Spielman-Teng). *There is an algorithm $x = \texttt{STSolve}(L, y, \delta)$ which takes a Laplacian matrix L, a column vector y, and an error parameter $\delta > 0$, and returns a column vector x satisfying*

$$\|x - L^+ y\|_L \leq \epsilon \|L^+ y\|_L,$$

where $\|y\|_L = \sqrt{y^T L y}$. The algorithm runs in expected time $\widetilde{O}\left(m \log(1/\delta)\right)$, where m is the number of non-zero entries in L.

Let $Z = QW^{1/2}BL^+$. We will compute an approximation $\widetilde{Z}$ by using $\texttt{STSolve}$ to approximately compute the rows of Z. Let the column vectors z_i and $\tilde{z}_i$ denote the ith rows of Z and $\widetilde{Z}$, respectively (so that z_i is the ith column of Z^T). Now we can construct the matrix $\widetilde{Z}$ in the following three steps.

1. Let Q be a random $\pm 1/\sqrt{k}$ matrix of dimension $k \times n$ where $k = 24 \log n / \epsilon^2$.

2. Compute $Y = QW^{1/2}B$. Note that this takes $2m \times 24 \log n / \epsilon^2 + m = \widetilde{O}(m/\epsilon^2)$ time since B has $2m$ entries and $W^{1/2}$ is diagonal.

3. Let y_i, for $1 \leq i \leq k$, denote the rows of Y, and compute $\tilde{z}_i = \texttt{STSolve}(L, y_i, \delta)$ for each i.

We now prove that, for our purposes, it suffices to call $\texttt{STSolve}$ with

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

Lemma 4.7. *Suppose*

$$(1 - \epsilon)R_{uv} \leq \|Z(\chi_u - \chi_v)\|^2 \leq (1 + \epsilon)R_{uv},$$

for every pair $u, v \in V$. If for all i,

$$\|z_i - \tilde{z}_i\|_L \leq \delta \|z_i\|_L, \tag{4.5}$$

where

$$\delta \leq \frac{\epsilon}{3}\sqrt{\frac{2(1-\epsilon)w_{min}}{(1+\epsilon)n^3 w_{max}}} \tag{4.6}$$

then

$$(1-\epsilon)^2 R_{uv} \leq \|\widetilde{Z}(\chi_u - \chi_v)\|^2 \leq (1+\epsilon)^2 R_{uv},$$

for every uv.

Proof. Consider an arbitrary pair of vertices u, v. It suffices to show that

$$\left| \|Z(\chi_u - \chi_v)\| - \|\tilde{Z}(\chi_u - \chi_v)\| \right| \leq \frac{\epsilon}{3} \|Z(\chi_u - \chi_v)\| \tag{4.7}$$

since this will imply

$$\left| \|Z(\chi_u - \chi_v)\|^2 - \|\tilde{Z}(\chi_u - \chi_v)\|^2 \right|$$

$$= \left| \|Z(\chi_u - \chi_v)\| - \|\tilde{Z}(\chi_u - \chi_v)\| \right| \cdot \left| \|Z(\chi_u - \chi_v)\| + \|\tilde{Z}(\chi_u - \chi_v)\| \right|$$

$$\leq \frac{\epsilon}{3} \cdot \left(2 + \frac{\epsilon}{3}\right) \|Z(\chi_u - \chi_v)\|^2.$$

As G is connected, there is a simple path P connecting u to v. Applying the triangle inequality twice, we obtain

$$\left| \|Z(\chi_u - \chi_v)\| - \left\|\widetilde{Z}(\chi_u - \chi_v)\right\| \right| \leq \left\|(Z - \widetilde{Z})(\chi_u - \chi_v)\right\|$$

$$\leq \sum_{ab \in P} \left\|(Z - \widetilde{Z})(\chi_a - \chi_b)\right\|.$$

We will upper bound this later term by considering its square:

$$\left(\sum_{ab \in P} \left\| (Z - \tilde{Z})(\chi_a - \chi_b) \right\| \right)^2 \leq n \sum_{ab \in P} \left\| (Z - \tilde{Z})(\chi_a - \chi_b) \right\|^2 \qquad \text{by Cauchy-Schwarz}$$

$$\leq n \sum_{ab \in E} \left\| (Z - \tilde{Z})(\chi_a - \chi_b) \right\|^2$$

$$= n \left\| (Z - \tilde{Z})B^T \right\|_F^2 \qquad \text{writing this as a Frobenius norm}$$

$$= n \left\| B(Z - \tilde{Z})^T \right\|_F^2$$

$$\leq \frac{n}{w_{min}} \left\| W^{1/2} B(Z - \tilde{Z})^T \right\|_F^2$$

$$\text{since } \| W^{-1/2} \|_2 \leq 1/\sqrt{w_{min}}$$

$$\leq \delta^2 \frac{n}{w_{min}} \left\| W^{1/2} B Z^T \right\|_F^2$$

$$\text{since } \| W^{1/2} B(z_i - \tilde{z}_i) \|^2 \leq \delta^2 \| W^{1/2} B z_i \|^2 \text{ by (4.5)}$$

$$= \delta^2 \frac{n}{w_{min}} \sum_{ab \in E} w_{ab} \| Z(\chi_a - \chi_b) \|^2$$

$$\leq \delta^2 \frac{n}{w_{min}} \sum_{ab \in E} w_{ab}(1 + \epsilon) R_{ab}$$

$$\leq \delta^2 \frac{n(1 + \epsilon)}{w_{min}}(n - 1) \qquad \text{since } \sum_{ab \in E} w_{ab} R_{ab} = n - 1.$$

On the other hand,

$$\| Z(\chi_u - \chi_v) \|^2 \geq (1 - \epsilon) R_{uv} \geq \frac{2(1 - \epsilon)}{n w_{max}},$$

by Proposition 4.8. Combining these bounds, we have

$$\frac{\left| \| Z(\chi_u - \chi_v) \| - \left\| \tilde{Z}(\chi_u - \chi_v) \right\| \right|}{\| Z(\chi_u - \chi_v) \|} \leq \delta \left(\frac{n(1 + \epsilon)}{w_{min}}(n - 1) \right)^{1/2} \cdot \left(\frac{n w_{max}}{2(1 - \epsilon)} \right)^{1/2}$$

$$\leq \frac{\epsilon}{3} \qquad \text{by (4.6),}$$

as desired. $\qquad \square$

Proposition 4.8. *If $G = (V, E, w)$ is a connected graph, then for all $u, v \in V$,*

$$R_{uv} \geq \frac{2}{n w_{max}}.$$

Proof. By Rayleigh's monotonicity law (see [13]), each resistance R_{uv} in G is at least

41

the corresponding resistance R'_{uv} in $G' = w_{max} \times K_n$ (the complete graph with all edge weights w_{max}) since G' is obtained by increasing weights (i.e., conductances) of edges in G. But by symmetry each resistance R'_{uv} in G' is exactly

$$\frac{\sum_{uv} R'_{uv}}{\binom{n}{2}} = \frac{(n-1)/w_{max}}{n(n-1)/2} = \frac{2}{nw_{max}}.$$

Thus $R_{uv} \geq \frac{2}{nw_{max}}$ for all $u, v \in V$. $\qquad\square$

Thus the construction of $\widetilde{Z}$ takes $\widetilde{O}(m \log(1/\delta)/\epsilon^2) = \widetilde{O}(m \log r/\epsilon^2)$ time. We can then find the approximate resistance $\|\widetilde{Z}(\chi_u - \chi_v)\|^2 \approx R_{uv}$ for any $u, v \in V$ in $O(\log n/\epsilon^2)$ time simply by subtracting two columns of $\widetilde{Z}$ and computing the norm of their difference. $\qquad\square$

Using the above procedure, we can compute arbitrarily good approximations to the effective resistances $\{R_e\}$ which we need for sampling in nearly–linear time. By Corollary 3.10, any constant factor approximation yields a sparsifier, so we are done.

Chapter 5

Application to Contact Points

In this chapter we will prove a refinement of the Spectral Sparsification Theorem of Chapter 3 and use it to make progress on a problem in high dimensional convex geometry. The proof will again use two 'barrier' potential functions to iteratively build a sparse weighted sum of rank one outer products with desirable spectral properties. The additional twist in this case is that we require the vectors produced to have small *mean*; in Theorem 5.5, we show how to do this at some cost in both sparsity and control on the spectrum. We view this as an important first step in extending the barrier method to more constrained settings and exploring the scope of its applicability.

5.1 Introduction

Let K be an arbitrary convex body in $\mathbb{R}^n$ and let $\mathcal{E}$ be a minimal volume ellipsoid containing K. Then the *contact points* of K are the points of intersection of $\mathcal{E}$ and K. The ellipsoid $\mathcal{E}$ is unique and characterized by a celebrated theorem of F. John [7], which says that if K is embedded via an affine transformation in $\mathbb{R}^n$ so that $\mathcal{E}$ is the standard Euclidean ball B_2^n, then there are $m \leq N = n(n + 3)/2$ contact points $x_1, \ldots, x_m \in K \cap B_2^n$ and nonnegative weights $c_1, \ldots c_m$ for which

$$\sum_i c_i x_i = 0 \qquad \text{(mean zero)} \qquad (5.1)$$

$$\sum_i c_i x_i x_i^T = I \qquad \text{(inertia matrix identity)} \qquad (5.2)$$

Moreover, any convex body K' containing $x_1, \ldots, x_m$ must have B_2^n inside its John ellipsoid. We refer to a system $(c_i, x_i)_{i \leq m}$ which satisfies (5.1) and (5.2) as a *John's decomposition* of the identity.

The study of contact points has been fruitful in Convex Geometry, for instance in understanding the behavior of volume ratios of symmetric and nonsymmetric convex bodies [7] and in estimating distances between convex bodies and the cube or simplex [38, 24]. In this chapter, we consider the number of contact points of a convex body.

Define a distance d between two (not necessarily symmetric) convex bodies K and H in $\mathbb{R}^n$ as follows[1]:

$$d(K, H) = \inf_{T \in GL(n), u \in \mathbb{R}^n} \{c : H + u \subset TK \subset c(H + u)\}$$

and let $\mathcal{K}$ be the space of all convex bodies equipped with the topology induced by d. Gruber [28] proved that the set of K having fewer than $N = n(n+3)/2$ contact points is of the first Baire category in $\mathcal{K}$. However, Rudelson has shown that every K is arbitrarily close to a body which has a much smaller number of contact points.

Theorem 5.1 (Rudelson [39]). *Suppose K is a convex body in $\mathbb{R}^n$ and $\epsilon > 0$. Then there is a convex body H such that $d(H, K) \leq 1+\epsilon$ and H has at most $m \leq Cn \log n/\epsilon^2$ contact points, where C is a universal constant.*

In this chapter, we show that the $\log n$ factor in Rudelson's theorem is unnecessary in many cases. For symmetric convex bodies, we obtain exactly the same distance guarantee $d(H, K) \leq 1 + \epsilon$ but with a much smaller number $m \leq 32n/\epsilon^2$ of contact points of H. For arbitrary convex bodies, we show a somewhat weaker result that only guarantees an H within constant distance $d(H, K) \leq 2.24$, with $m \leq Cn$ contact points for some universal C. Thus Rudelson's $O(n \log n)$ bound is still the best known in the regime $d(H, K) < 2.24$ for nonsymmetric bodies.

Our approach for constructing H is the same as Rudelson's, and consists of two steps:

1. Given a John's decomposition $(c_i, x_i)_{i \leq m}$ for K, extract a small subsequence of points x_i which are *approximately* a John's decomposition. To be precise, choose a set of scalars b_i, at most $s = O(n)$ of which are nonzero and a small 'recentering' vector u for which

$$\sum_i b_i(x_i + u) = 0 \qquad \left\| I - \sum_i b_i(x_i + u)(x_i + u)^T \right\| \leq \epsilon.$$

2. Map the points $(b_i, x_i + u)_{i \leq s}$ in the approximate John's decomposition to an exact John's decomposition $(a_i, u_i)_{i \leq s}$, and show that it characterizes the John Ellipsoid of a body H that is close to K.

The source of our improvement is a new method for extracting the approximate decomposition in step (1). Whereas the b_i were chosen by random methods in Rudelson's work, we now use the deterministic procedure described in Theorem 3.1, restated here for convenience:

---

[1]When K and H are symmetric then we can take $u = 0$ and d becomes the usual Banach–Mazur distance.

Theorem 5.2 (Spectral Sparsification of the Identity, copy of Theorem 3.2). *Suppose $d > 1$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are vectors in $\mathbb{R}^n$ with*

$$\sum_{i \leq m} \mathbf{v}_i \mathbf{v}_i^T = I.$$

Then there exist scalars $s_i \geq 0$ with $|\{i : s_i \neq 0\}| \leq dn$ so that

$$I \preceq \sum_{i \leq m} s_i \mathbf{v}_i \mathbf{v}_i^T \preceq \left(\frac{\sqrt{d} + 1}{\sqrt{d} - 1} \right)^2 I.$$

A sharp result regarding the contact points of *symmetric* convex bodies can be derived as an immediate corollary of Theorem 5.2 and Rudelson's proof of Theorem 1.1 [38].

Corollary 5.3. *If K is a symmetric convex body in $\mathbb{R}^n$ and $\epsilon > 0$, then there exists a body H such that $H \subset K \subset (1 + \epsilon)H$ and H has at most $m \leq 32n/\epsilon^2$ contact points with its John Ellipsoid.*

Proof. Suppose K is a symmetric convex body whose John ellipsoid is B_2^n, and let $X = \{x_1, \ldots x_m\}$ be contact points satisfying (5.1,5.2) with weights $c_1, \ldots c_m$. Since K is symmetric we can assume that $x_i \in X \iff -x_i \in X$, and that the corresponding weights c_i are equal.

We will extract an approximate John's decomposition from X. Apply Theorem 5.2 to the vectors $\mathbf{v}_i = \sqrt{c_i} x_i$ with parameter $d = 16/\epsilon^2$ to obtain scalars s_i, and let $Y \subset X$ be the set of x_i with nonzero s_i. We are now guaranteed that

$$I \preceq \sum_{x_i \in Y} s_i c_i x_i x_i^T \preceq \left(\frac{4/\epsilon + 1}{4/\epsilon - 1} \right)^2 I$$

with $|Y| \leq 16n/\epsilon^2$. Notice that by an easy calculation

$$\left(\frac{4/\epsilon + 1}{4/\epsilon - 1} \right)^2 \leq 1 + \epsilon$$

for sufficiently small epsilon.

In order to obtain a John's decomposition from these vectors, we need to ensure the mean zero condition (5.1). This is achieved easily by taking a negative copy of each vector in Y and halving the scalars s_i, since

$$\sum_{x_i \in Y} (s_i/2) c_i x_i + (s_i/2) c_i (-x_i) = 0$$

and

$$\sum_{x_i \in Y} (s_i/2) c_i x_i x_i^T + (s_i/2) c_i (-x_i)(-x_i)^T = \sum_{x_i \in Y} s_i c_i x_i x_i^T$$

45

which we know is a good approximation to the identity. Thus the vectors in $Y \cup -Y$ with weights $b_i = s_i c_i / 2$ on x_i and $-x_i$ constitute a $(1 + \epsilon)$-approximate John's decomposition with only $32n/\epsilon^2$ points. Substituting this fact in place of Lemma 3.1 [38] in the proof of Theorem 1.1 [38] gives the promised result. □

When the body K is not symmetric, there is no immediate way to guarantee the mean zero condition. If we simply recenter the vectors produced by Theorem 5.2 to have mean zero by adding $u = -\frac{\sum_i b_i x_i}{\sum_i b_i}$ to each x_i, then the corresponding inertia matrix is

$$\sum_i b_i (x_i + u)(x_i + u)^T = \sum_i b_i x_i x_i^T - \left(\sum_i b_i \right) u u^T \tag{5.3}$$

which no longer approximates the identity (indeed, it can have negative eigenvalues) if $\| \left(\sum_i b_i \right) u u^T \|$ is large. This is the issue that we address here. In Section 5.2, we prove a variant of Theorem 5.2 which allows us to obtain very good control on the mean u at the cost of having a worse (at best, factor 4) approximation of the inertia matrix to the identity. In Section 5.3, we show that this is still sufficient to carry out Rudelson's construction of the approximating body H. The end result is the following theorem.

Theorem 5.4. *For every $\epsilon > 0$ the following is true for n sufficiently large. If K is a convex body in $\mathbb{R}^n$, then there is a convex body H such that*

$$H \subset K \subset (\sqrt{5} + \epsilon) H$$

has at most $O_\epsilon(n)$ contact points with its John Ellipsoid.

5.2 Approximate John's Decompositions

In this section we will prove the following theorem.

Theorem 5.5. *Suppose we are given a John's decomposition of the identity, i.e., unit vectors $x_1, \ldots, x_m \in \mathbb{R}^n$ with nonnegative scalars c_i such that*

$$\sum_i c_i x_i = 0 \tag{5.4}$$

$$\sum_i c_i x_i x_i^T = I. \tag{5.5}$$

Then for every $\epsilon > 0$ there are scalars b_i, at most $O_\epsilon(n)$ nonzero, and a vector u such that

$$I \preceq \sum_i b_i (x_i + u)(x_i + u)^T \preceq (4 + \epsilon) I$$

$$\sum_i b_i (x_i + u) = 0$$

46

$$\left(\sum_i b_i \right) \|u\|^2 \leq \epsilon.$$

We remark that the requirement (5.4) is necessary to allow a useful bound on u, since otherwise we can take $x_i = e_i$ with $c_i = 1$ for the canonical basis vectors $\{e_i\}_{i \leq n}$ and it is easily seen that

$$\left(\sum_i b_i \right) \|u\|^2 \geq \min b_i = \lambda_{min} \left(\sum_i b_i x_i x_i^T \right)$$

for every choice of scalars b_i, which is worthless considering (5.3).

5.2.1   An Outline of The Proof

As in the proof of Theorem 5.2, we will build the approximate John's decomposition $(b_i, x_i + u)$ by an iterative process which adds one vector at a time. At any step of the process, let

$$A = \sum_j b_j x_j x_j^T$$

denote the inertia matrix of the vectors that have already been added (i.e., the b_i's that have been set to some nonzero value), and let

$$z = \sum_j b_j x_j$$

denote their weighted sum. Initially both $A = 0$ and $z = 0$.

Barrier Functions. As in the proof of Theorem 3.1, The choice of vector to add in each step will be guided by two 'barrier' potential functions which we will use to maintain control on the eigenvalues of A. For real numbers $u, l \in \mathbb{R}$, which we will call the upper and lower barrier respectively, define:

$$\Phi^u(A) \overset{\text{def}}{=} \mathrm{Tr}(uI - A)^{-1} = \sum_i \frac{1}{u - \lambda_i} \quad \text{(Upper potential)}.$$

$$\Phi_l(A) \overset{\text{def}}{=} \mathrm{Tr}(A - lI)^{-1} = \sum_i \frac{1}{\lambda_i - l} \quad \text{(Lower potential)},$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A.

As long as $A \prec uI$ and $A \succ lI$ (i.e., $\lambda_{\max}(A) < u$ and $\lambda_{\min}(A) > l$), these potential functions measure how far the eigenvalues of A are from the barriers u and l. In particular, they blow up as any eigenvalue approaches a barrier, since then $uI - A$ (or $A - lI$) approaches a singular matrix. Thus if $\Phi_l(A)$ and $\Phi^u(A)$ are appropriately bounded, then we can conclude that the eigenvalues of A are 'well–behaved'. For a more thorough discussion of where these barrier functions come from and why they work, we refer the reader to Section 3.1.1.

Invariants. We will maintain three invariants throughout the process. Note that u, l, A and z vary from step to step, while P_U, P_L, and ϵ remain fixed.

- The eigenvalues of A lie strictly between l and u:

$$lI \prec A \prec uI. \tag{5.6}$$

- Both the upper and lower potentials are bounded by some fixed values P_L and P_U:

$$\Phi_l(A) \leq P_L \qquad \Phi^u(A) \leq P_U. \tag{5.7}$$

- The running sum z is appropriately small:

$$\|z\|^2 \leq \epsilon \cdot \operatorname{Tr}(A) \tag{5.8}$$

Initialization. At the beginning of the process we have $A = 0$ and $z = 0$, and the barriers at initial values

$$l = l_0 = -1 \quad \text{and} \quad u = u_0 = 1. \tag{5.9}$$

It is easy to see that (5.6), (5.7), and (5.8) all hold with $P_U = P_L = n$ at this point.
Maintenance. The process will evolve in steps. Each step will consist of adding two vectors, tv and rw, where $t, r \geq 0$ and $v, w \in \{x_i\}_{i \leq m}$. We will call these the primary vector and the fix vector, respectively.

The primary vector will allow us to move the upper and lower barriers forward by fixed amounts $\delta_l > 0$ and $\delta_u > 0$ while maintaining the invariants (5.6) and (5.7); in particular, we will choose it to satisfy:

$$\Phi_{l+\delta_l}(A+tvv^T) \leq \Phi_l(A) \qquad \Phi^{u+\delta_u}(A+tvv^T) \leq \Phi^u(A) \qquad (l+\delta_l)I \prec A+tvv^T \prec (u+\delta_u)I. \tag{5.10}$$

The fix will correct any undesirable impact that the primary has on the sum; specifically, we will choose rw in a way that guarantees

$$\langle z, tv + rw \rangle \leq 0 \tag{5.11}$$

where z is the sum at the end of the previous step. Thus the net increase in the length of the sum in any step is given by

$$\|z + (tv + rw)\|^2 = \|z\|^2 + 2\langle z, tv + rw \rangle + \|tv + rw\|^2 \leq \|z\|^2 + (t + r)^2, \tag{5.12}$$

since v and w are unit vectors. The corresponding increase in the trace is simply $t + r$, and so if we guarantee in addition that the steps are sufficiently small:

$$t + r \leq \epsilon \tag{5.13}$$

then the invariant (5.8) can be maintained by induction as follows:

$$\frac{\|z + (tv + rw)\|^2}{\text{Tr}(A + tvv^T + rww^T)} \leq \frac{\|z\|^2 + (t + r)^2}{\text{Tr}(A) + (t + r)} \qquad \text{by (5.12)}$$

$$\leq \max\left\{\frac{\|z\|^2}{\text{Tr}(A)}, \frac{(t + r)^2}{t + r}\right\}$$

$$\leq \epsilon \qquad \text{by (5.13)}.$$

However, we need to make sure that adding the fix vector does not cause us to violate (5.6) or (5.7). To do this, the addition of rw will be accompanied by an appropriately large shift of $\delta_u^f > 0$ in the upper barrier. In particular, we will make sure that on top of satisfying (5.11), rw also satisfies

$$\Phi^{u + \delta_u + \delta_u^f}(A + tvv^T + rww^T) \leq \Phi^{u + \delta_u}(A + tvv^T) \quad \text{and} \quad A + tvv^T + rww^T \prec (u + \delta_u + \delta_u^f)I. \tag{5.14}$$

Since $A + tvv^T + rww^T \succ A + tvv^T$, there is no need to shift the lower barrier, and the analogous bound for the lower potential is immediate:

$$\Phi_{l + \delta_l + 0}(A + tvv^T + rww^T) \leq \Phi_{l + \delta_l}(A + tvv^T) \quad \text{with} \quad A + tvv^T + rww^T \succ (l + \delta_l)I.$$

Together with (5.10), these inequalities guarantee that

$$\Phi_{l + \delta_l}(A + tvv^T + rww^T) \leq \Phi_l(A) \leq P_L$$

and

$$\Phi^{u + \delta_u + \delta_u^f}(A + tvv^T + rww^T) \leq \Phi^u(A) \leq P_U,$$

thus maintaining both (5.6) and (5.7), as desired.

To summarize what has occurred during the step: we have added two vectors tv and rw and shifted l and u forward by δ_l and $\delta_l + \delta_u^f$, respectively, in a manner that our three invariants continue to hold. To show that such a step can actually be taken, we need to prove that as long as the invariants are maintained there must exist scalars $t, r \geq 0$ and vectors $v, w \in \{x_i\}_{i \leq m}$ which satisfy all of the conditions (5.10), (5.11), (5.13), and (5.14). We will do this in Lemma 5.6.

Termination. After s steps of the process, we have

$$(-1 + s\delta_l)I \prec A \prec (1 + s(\delta_u + \delta_u^f))I \qquad \text{by (5.6)}.$$

If we take s sufficiently large, then we can make the ratio $\lambda_{\max}(A)/\lambda_{\min}(A)$ arbitrarily close to $\frac{\delta_u + \delta_u^f}{\delta_l}$. In the actual proof, which we will present shortly, we will show that the process can be realized with parameters $\delta_l, \delta_u, \delta_u^f$ for which this ratio can be made arbitrarily close to 4 in $s = O(n)$ steps.

As for the mean, we will set $u = -\frac{\sum_j b_j x_j}{\sum_j b_j} = -\frac{z}{\mathrm{Tr}(A)}$, so that immediately

$$\left(\sum_j b_j\right) \|u\|^2 = \frac{\|z\|^2}{\mathrm{Tr}(A)} \le \epsilon \tag{5.15}$$

as promised.

5.2.2 Realizing the Proof

To complete the proof, we will identify a range of parameters $\delta_l, \delta_u, \delta_u^f$ for which the above process can actually be sustained.

Lemma 5.6 (One Step). *Suppose $(c_i, x_i)_{i \le m}$ is a John's decomposition and z is any vector. Let A be a matrix satisfying the invariants (5.6) and (5.7). If*

$$\frac{1}{\delta_u} + \frac{1}{\delta_u^f} + 2P_U + P_L + \frac{4n}{\epsilon} \le \frac{1}{\delta_l} \tag{5.16}$$

Then there are scalars $t, r \ge 0$ and vectors $v, w \in \{x_i\}$ which satisfy (5.10), (5.11), (5.13), and (5.14).

To this end, we recall the following lemmas from Chapter 3, which characterize how much of a vector one can add to a matrix without increasing the upper and lower potentials.

Lemma 5.7 (Upper Barrier Shift, copy of Lemma 3.4). *Suppose $A \prec uI$ and $\delta_u > 0$. Then there is a positive definite matrix $\mathbb{U} = \mathbb{U}(A, u, \delta_u)$ so that if v is any vector which satisfies*

$$v^T \mathbb{U} v \ge \frac{1}{t}$$

then

$$\Phi^{u+\delta_U}(A + tvv^T) \le \Phi^u(A) \quad \text{and} \quad \lambda_{\max}(A + tvv^T) < u + \delta_U.$$

That is, if we add t times vv^T to A and shift the upper barrier by δ_U, then we do not increase the upper potential.

Lemma 5.8 (Lower Barrier Shift, copy of Lemma 3.5). *Suppose $A \succ lI$, $\delta_l > 0$, and $\Phi_l(A) < 1/\delta_l$. Then there is a matrix $\mathbb{L} = \mathbb{L}(A, l, \delta_l)$ so that if v is any vector which satisfies*

$$v^T \mathbb{L} v \le \frac{1}{t}$$

then

$$\Phi_{l+\delta_L}(A + tvv^T) \le \Phi_l(A) \quad \text{and} \quad \lambda_{\min}(A + tvv^T) > l + \delta_L.$$

That is, if we add t times vv^T to A and shift the lower barrier by δ_L, then we do not increase the lower potential.

We will prove that desirable vectors exist by taking averages of the quantities $\mathbf{v}^T\mathbb{U}\mathbf{v}$ and $\mathbf{v}^T\mathbb{L}\mathbf{v}$ over our contact points $\{x_i\}$ with weights $\{c_i\}$. Since

$$\sum_i c_i x_i^T \mathbb{U} x_i = \mathbb{U} \bullet \left(\sum_i c_i x_i x_i^T \right)$$

$$= \mathbb{U} \bullet I \qquad\qquad \text{by (5.2)}$$

$$= \text{Tr}(\mathbb{U})$$

(and similarly for $\mathbb{L}$), it will be useful to recall bounds on $\text{Tr}(\mathbb{U})$ and $\text{Tr}(\mathbb{L})$ from Chapter 3. Remarkably, these bounds do not depend on the matrix A or on u, l at all, but only on the shifts δ_u and δ_l and on the potentials.

Lemma 5.9 (Traces of $\mathbb{L}$ and $\mathbb{U}$). *If $lI \prec A \prec uL$ with $\Phi_l(A) < P_L$ and $\Phi^u(A) < P_U$ then*

$$\text{Tr}(\mathbb{U}) \leq \frac{1}{\delta_u} + P_U$$

and

$$\text{Tr}(\mathbb{L}) \geq \frac{1}{\delta_l} - P_L.$$

We are now in a position to prove Lemma 5.6.

Proof of Lemma 5.6. Let $\mathbb{L} = \mathbb{L}(A, l, \delta_l)$, $\mathbb{U} = \mathbb{U}(A, u, \delta_u)$, $\mathbb{U}^f = \mathbb{U}(A, u + \delta_u, \delta_u^f)$ be the matrices produced by lemmas 5.7 and 5.8.

Let us focus on the primary vector first. By Lemmas 5.7 and 5.8, we can add $t\mathbf{v}$ without increasing potentials if

$$v^T \mathbb{U} v \leq 1/t \leq v^T \mathbb{L} v.$$

In fact, we will insist on v for which

$$v^T \mathbb{U} v + 2/\epsilon \leq 1/t \leq v^T \mathbb{L} v$$

as this will ensure that we can take $t \leq \epsilon/2$.

Let $D(v) = v^T \mathbb{L} v - v^T \mathbb{U} v - 2/\epsilon$ and call $\mathcal{F} = \{x_i : D(x_i) \geq 0\}$ the set of *feasible* vectors. Let $\mathcal{P} = \{x_i : \langle x_i, z \rangle > 0\}$ be the set of vectors with positive inner product with z, and let $\mathcal{N} = \{x_i : \langle x_i, z \rangle \leq 0\}$ be the vectors in the complementary halfspace.

We will always add as little of a primary vector can, so we can assume that we take $1/t = v^T \mathbb{L} v$ whenever $v \in \mathcal{F}$. Here is the rule for choosing which v to add: choose the feasible v for which $t\langle v, z \rangle$ is minimized. If this quantity is negative then there is no need for a fix vector, and taking $w = 0$ we are done. Otherwise we know that $\mathcal{F} \subset \mathcal{P}$ and

$$\frac{\langle v, z \rangle}{\alpha} \geq \frac{1}{t} = v^T L v \quad \forall v \in \mathcal{F}. \tag{5.17}$$

where $\alpha = \min\{t\langle v, z \rangle : v \in \mathcal{F}\}$. Taking a sum, we notice that

$$\sum_{\mathcal{P}} c_i \frac{\langle x_i, z \rangle}{\alpha} \geq \sum_{\mathcal{F}} c_i \frac{\langle x_i, z \rangle}{\alpha} \qquad \text{since } \mathcal{F} \subset \mathcal{P}$$

$$\geq \sum_{\mathcal{F}} c_i x_i^T \mathbb{L} x_i \qquad \text{by (5.17)}$$

$$\geq \sum_{\mathcal{F}} c_i D(x_i) \qquad \text{since } \mathbb{U} \succeq 0 \text{ implies that } D(x_i) \leq x_i^T \mathbb{L} x_i$$

$$\geq \sum_i c_i D(x_i) \qquad \text{since } D < 0 \text{ outside } \mathcal{F}$$

However, since $\sum_i c_i x_i = 0$ this implies that

$$\sum_{\mathcal{N}} c_i \frac{\langle x_i, -z \rangle}{\alpha} = \sum_{\mathcal{P}} c_i \frac{\langle x_i, z \rangle}{\alpha} \geq \sum_i c_i D(x_i). \tag{5.18}$$

We will use (5.18) to show that a suitable fix vector w exists. We are interested in finding a $w \in \{x_i\}$ and $r \geq 0$ for which

$$r\langle w, -z \rangle \geq \alpha \qquad \text{(sufficient to reverse } \alpha = t\langle v, z\rangle\text{)—for (5.11)}$$
$$w^T \mathbb{U}^f w + 2/\epsilon \leq 1/r \quad \text{(upper barrier feasible with shift } \delta_u^f \text{ and } r \leq \epsilon/2\text{)—for (5.13,5.14)}$$

Thus it suffices to find a w for which

$$w^T \mathbb{U}^f w + 2/\epsilon \leq \frac{\langle w, -z \rangle}{\alpha},$$

and then we can squeeze $1/r$ in between. Taking a weighted sum over all vectors of interest, it will be sufficient to show that

$$\sum_{\mathcal{N}} c_i x_i^T \mathbb{U}^f x_i + 2c_i/\epsilon \leq \sum_{\mathcal{N}} c_i \frac{\langle x_i, -z \rangle}{\alpha}.$$

For the left hand side we use the crude estimate

$$\sum_{\mathcal{N}} c_i x_i^T \mathbb{U}^f x_i + 2c_i/\epsilon \leq \sum_{\mathcal{N} \cup \mathcal{P}} c_i x_i^T \mathbb{U}^f x_i + 2c_i/\epsilon = \text{Tr}(\mathbb{U}^f) + 2n/\epsilon$$

52

and for the right hand side we consider that

$$\sum_{\mathcal{N}} c_i \frac{\langle x_i, -z \rangle}{\alpha} \geq \sum_i c_i D(x_i) \qquad\qquad \text{by (5.18)}$$

$$= \sum_i c_i x_i^T \mathbb{L} x_i - c_i x_i^T \mathbb{U} x_i - 2c_i/\epsilon$$

$$= \text{Tr}(\mathbb{L}) - \text{Tr}(\mathbb{U}) - 2n/\epsilon \qquad\qquad \text{since } \sum_i c_i x_i x_i^T = I$$

Thus it will be enough to have

$$\text{Tr}(\mathbb{U}^f) + 2n/\epsilon \leq \text{Tr}(\mathbb{L}) - \text{Tr}(\mathbb{U}) - 2n/\epsilon$$

which follows from our hypothesis (5.16) and Lemma 5.9. $\qquad\qquad\square$

Proof of Theorem 5.5. If we start with $l_0 = -1$ and $u_0 = 1$ then we can take $P_L = P_U = n$. If we set $\delta_u = \delta_u^f = (2 + \epsilon)\delta_l$, then (5.16) reduces to

$$\frac{2}{(2+\epsilon)\delta_l} + 3n + \frac{4n}{\epsilon} \leq \frac{1}{\delta_l},$$

which it is easy to check is satisfied whenever

$$\delta_l \leq \frac{\epsilon^2}{10n}.$$

At the end of s steps we have

$$\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} \leq \frac{1 + s \cdot 2(2+\epsilon)\delta_l}{-1 + s\delta_l}$$

which can be made arbitrarily at most $4 + 3\epsilon$ by taking $s \geq 100n/\epsilon^3$ steps. By (5.3) and (5.15), the term $\text{Tr}(A)uu^T$ only affects this by at most ϵ. $\qquad\square$

5.3   Construction of the Approximating Body

The contents of this section are very similar to Rudelson [38] Section 4, but the calculations are a bit more subtle because we only have a 4—approximate John's decomposition as opposed to a $(1 + \epsilon)$-approximate one.

The main result is a generic procedure for turning approximate John's decompositions into approximating bodies:

Lemma 5.10. *Suppose K has contact points $(c_i, x_i)_{i \leq m}$ and $(b_i, y_i = x_i + u)_{i \leq s}$ are the*

vectors produced by Theorem 5.5, with

$$A = \sum_i b_i y_i y_i^T.$$

Then for $\epsilon > 0$ and n sufficiently large, there is a body H with at most s contact points and

$$d(H, K) \leq (1 + \epsilon) \left(\kappa(A) \left(1 + \frac{(\sqrt{\kappa(A)} - 1)^2}{4} \right) \right)^{1/2}. \tag{5.19}$$

This immediately yields a proof of Theorem 5.4:

Proof of Theorem 5.4. Since the condition number $\kappa(A)$ guaranteed by Theorem 5.5 can be made arbitrarily close to 4, the number in (5.19) can be made arbitrarily close to

$$\left(4 \left(1 + \frac{1}{4} \right) \right)^{1/2} = \sqrt{5}$$

as desired. □

Let $(b_i, y_i = x_i + u)_{i \leq s}$ be the approximate John's decomposition guaranteed by Theorem 5.5, with

$$\sum_i b_i y_i = 0$$

and

$$\sum_i b_i y_i y_i^T = A.$$

There are two problems with this: the y_i are not unit vectors, and their moment ellipsoid $\mathcal{E} = A^{1/2} B_2^n$ is not the sphere. We will adjust the vectors in a manner that fixes both these problems, to obtain an exact John's decomposition $(\hat{a}_i, \hat{u}_i)_{i \leq s}$.

Add a small vector v, to be determined later, to each y_i to obtain vectors

$$\hat{y}_i = y_i + v$$

with inertia matrix

$$A_v = \sum_i b_i \hat{y}_i \hat{y}_i^T = \sum_i b_i y_i y_i^T + \sum_i b_i v v^T = A + \text{Tr}(A) v v^T. \tag{5.20}$$

(we will use $\hat{\ }$ to denote vectors that depend on v.) Let $R_v = A_v^{1/2}$ and let $\mathcal{E}_v = R_v B_2^n$ be the corresponding moment ellipsoid. If we rescale each $\hat{y}_i$ to lie on $\mathcal{E}_v$:

$$\hat{z}_i = \frac{\hat{y}_i}{\|\hat{y}_i\|_{\mathcal{E}_v}} \qquad \text{where } \|\hat{y}_i\|_{\mathcal{E}_v} = \|R_v^{-1} \hat{y}_i\|. \tag{5.21}$$

and then apply the inverse transformation R_v^{-1} which maps $\mathcal{E}_v$ to B_2^n, then we obtain

unit vectors

$$\hat{u}_i = R_v^{-1}\hat{z}_i.$$

Moreover, if these are given weights

$$\hat{a}_i = b_i\|\hat{y}_i\|_{\mathcal{E}v}^2$$

then we have an exact decomposition of the identity since

$$\sum_i \hat{a}_i \hat{u}_i \hat{u}_i^T = R_v^{-1}\left(\sum_i b_i\|\hat{y}_i\|_{\mathcal{E}v}^2 \frac{\hat{y}_i\hat{y}_i^T}{\|\hat{y}_i\|_{\mathcal{E}v}^2}\right)R_v^{-1} = R_v^{-1}A_v R_v^{-1} = I.$$

In the following lemma, we show that there must exist a small v for which the weighted sum

$$\sum_i \hat{a}_i \hat{u}_i = \sum_i b_i\|\hat{y}_i\|_{\mathcal{E}v}^2 R_v^{-1}\frac{\hat{y}_i}{\|\hat{y}_i\|_{\mathcal{E}v}} = R_v^{-1}\left(\sum_i b_i\|\hat{y}_i\|_{\mathcal{E}v}\hat{y}_i\right)$$

is equal to zero. This will complete the construction of $(\hat{a}_i, \hat{u}_i)_{i\leq s}$.

Lemma 5.11. *Let b_i, y_i, A, etc. be as above and let $\epsilon > 0$, and let n be sufficiently large. Then there is a vector v with*

$$\|v\| \leq v_A \overset{\text{def}}{=} \frac{1+\epsilon}{2}\left(\sqrt{\kappa(A)} - 1\right)\sqrt{\frac{\|A\|}{\text{Tr}(A)}} \tag{5.22}$$

for which

$$\sum_i b_i\|\hat{y}_i\|_{\mathcal{E}v}\hat{y}_i = 0.$$

Proof. We need to find a v for which

$$\sum_i b_i\sqrt{(y_i + v)^T A_v^{-1}(y_i + v)}(y_i + v) = 0.$$

As in [38], we will do this using the Brouwer fixed point theorem. In particular it will suffice to show that the function

$$F(v) = -\frac{\sum_i b_i\beta_i^{(v)}y_i}{\sum_i b_i\beta_i^{(v)}} \qquad \text{where } \beta_i^{(v)} = \sqrt{(y_i + v)^T(A + vv^T)^{-1}(y_i + v)}$$

$$= -\frac{\sum_i b_i(\beta_i^{(v)} - \mu)y_i}{\sum_i b_i\beta_i^{(v)}} \qquad \text{for any } \mu \in \mathbb{R} \text{ since } \sum_i b_iy_i = 0$$

maps $v_A B_2^n$ to itself. This can be demonstrated as follows:

$$\|F(v)\| = \max_{\|w\|=1} \frac{\sum_i b_i(\beta_i^{(v)} - \mu)\langle y_i, w\rangle}{\sum_i b_i \beta_i^{(v)}}$$

$$\leq \frac{(\sum_i b_i(\beta_i^{(v)} - \mu)^2)^{1/2}}{\sum_i b_i \beta_i^{(v)}} \cdot \max_{\|w\|=1} \left(\sum_i b_i\langle y_i, w\rangle^2\right)^{1/2}$$

by Cauchy–Schwarz

$$= \frac{(\sum_i b_i(\beta_i^{(v)} - \mu)^2)^{1/2}}{\sum_i b_i \beta_i^{(v)}} \cdot \|A\|^{1/2}$$

$$\leq \frac{1}{1 - O(\mathrm{Tr}(A)^{-1/2})} \frac{(\sum_i b_i(\beta_i^{(v)} - \mu)^2)^{1/2}}{\sum_i b_i \beta_i^{(0)}} \cdot \|A\|^{1/2}$$

by Lemma 5.12 and $\beta_i^{(v)} = \Omega(1)$

$$\leq \frac{1}{1 - O(\mathrm{Tr}(A)^{-1/2})} \frac{(\sum_i b_i(\beta_i^{(v)} - \mu)^2)^{1/2}}{\sum_i b_i} \cdot \|A\|^{1/2} \cdot \|A\|^{1/2}$$

since $\min_i \beta_i^{(0)} \geq \|A\|^{-1}$

$$\leq \frac{1}{1 - O(\mathrm{Tr}(A)^{-1/2})} \frac{(\sum_i b_i(\beta_i^{(0)} - \mu)^2)^{1/2} + O(\mathrm{Tr}(A)^{1/4})}{\mathrm{Tr}(A)} \cdot \|A\|$$

by Lemma 5.13 and $\sqrt{a + b} \leq \sqrt{a} + \sqrt{b}$

$$\leq \frac{(1 + \epsilon)(\sum_i b_i(\beta_i^{(0)} - \mu)^2)^{1/2}}{\mathrm{Tr}(A)} \cdot \|A\|$$

for any $\epsilon > 0$ and large n, since $\mathrm{Tr}(A) = \Omega(n)$

$$\leq \frac{(1 + \epsilon)(\sum_i b_i)^{1/2} \max_i |\beta_i^{(0)} - \mu|}{\mathrm{Tr}(A)} \cdot \|A\|$$

$$= (1 + \epsilon)\frac{|\max_i \beta_i^{(0)} - \min_i \beta_i^{(0)}|}{2} \frac{\|A\|}{\mathrm{Tr}(A)^{1/2}}$$

setting $\mu = (\max_i \beta_i^{(0)} - \min_i \beta_i^{(0)})/2$

$$\leq \frac{1 + \epsilon}{2} \left(\|A^{-1}\|^{1/2} - \frac{1}{\|A\|^{1/2}}\right) \frac{\|A\|}{\mathrm{Tr}(A)^{1/2}}$$

$$= \frac{1 + \epsilon}{2} \left(\|A^{-1}\|^{1/2}\|A\|^{1/2} - 1\right) \frac{\|A\|^{1/2}}{\mathrm{Tr}(A)^{1/2}},$$

as desired. □

Lemma 5.12. *If* $\|v\| = O(\mathrm{Tr}(A)^{-1/2})$ *then*

$$\sum_i b_i \beta_i^{(v)} \geq \sum_i b_i \beta_i^{(0)} - O(\mathrm{Tr}(A)^{1/2}).$$

Proof. We can lowerbound the individual terms as

$$\begin{aligned}
\beta_i^{(v)} &= \|R_v^{-1}(y_i + v)\| \\
&\geq \|R_v^{-1}y_i\| - \|R_v^{-1}v\| \\
&= \left(y_i^T(A + vv^T)^{-1}y_i\right)^{1/2} - \|R_v^{-1}v\| \\
&= \left(y_i^T\left(A^{-1} - \frac{A^{-1}vv^TA^{-1}}{1 + v^TA^{-1}v}\right)y_i\right)^{1/2} - \|R_v^{-1}v\| \\
&\quad\quad\quad \text{by Sherman–Morrison} \\
&\geq \sqrt{y_i^TA^{-1}y_i} - \left(y_i^T\left(\frac{A^{-1}vv^TA^{-1}}{1 + v^TA^{-1}v}\right)y_i\right)^{1/2} - \|R_v^{-1}v\| \\
&\quad\quad\quad \text{since } \sqrt{a - b} \geq \sqrt{a} - \sqrt{b}.
\end{aligned}$$

Taking a sum, we observe the difference of interest is bounded by

$$\sum_i b_i\beta_i^{(0)} - \sum_i b_i\beta_i^{(v)} \leq \sum_i b_i\left(y_i^T\left(\frac{A^{-1}vv^TA^{-1}}{1 + v^TA^{-1}v}\right)y_i\right)^{1/2} + \sum_i b_i\|R_v^{-1}v\|.$$

The first sum is handled by Cauchy Schwarz:

$$\begin{aligned}
\sum_i b_i\left(y_i^T\left(\frac{A^{-1}vv^TA^{-1}}{1 + v^TA^{-1}v}\right)y_i\right)^{1/2} &\leq \left(\sum_i b_i\right)^{1/2}\left(\sum_i b_iy_i^T\left(\frac{A^{-1}vv^TA^{-1}}{1 + v^TA^{-1}v}\right)y_i\right)^{1/2} \\
&= \mathrm{Tr}(A)^{1/2}\left(\frac{\mathrm{Tr}(AA^{-1}vv^TA^{-1})}{1 + v^TA^{-1}v}\right)^{1/2} \\
&\quad\quad\quad \text{since } \sum_i b_iy_iy_i^T = A \\
&< \mathrm{Tr}(A)^{1/2}.
\end{aligned}$$

For the second, we observe crudely that

$$\sum_i b_i\|R_v^{-1}v\| \leq \mathrm{Tr}(A)\|R_v^{-1}\|\|v\| = O(\mathrm{Tr}(A)^{1/2})$$

since $\|R_v^{-1}\| = O(1)$. $\qquad\square$

Lemma 5.13. *If* $\|v\| = O(\mathrm{Tr}(A)^{-1/2})$ *then*

$$\sum_i b_i(\beta_i^{(v)} - \mu)^2 \leq \sum_i b_i(\beta_i^{(0)} - \mu)^2 + O(\mathrm{Tr}(A)^{-1/2}).$$

Proof. Write

$$\sum_i b_i(\beta_i^{(v)} - \mu)^2 = \sum_i b_i\beta_i^{(v)2} - 2b_i\beta_i^{(v)} + b_i\mu^2.$$

Then $\beta_i^{(v)2} \leq \beta_i^{(0)2}$ by $A + vv^T \succeq A$, and

$$-2\sum_i b_i \beta_i^{(v)} \leq -2\sum_i b_i \beta_i^{(0)} + O(2\mathrm{Tr}(A)^{1/2}) \quad \text{by Lemma 5.12,}$$

as desired. □

Proof of Lemma 5.10. We are now in a position to construct the body H promised in Theorem 5.4. Let K be a convex body with B_2^n as its John ellipsoid and contact points $(c_i, x_i)_{i \leq m}$. Use Theorem 5.5 to obtain a subsequence $(b_i, y_i)_{i \leq s}$ with only $O(n)$ points Let $v, A_v, \{\hat{z}_i\}_{i \leq s}$, and $(\hat{a}_i, \hat{u}_i)_{i \leq s}$ be as above. Let $\theta_{max} = \max_i \|\hat{z}_i\|_2$ and $\theta_{min} = \min_i \|\hat{z}_i\|_2$ and let $\epsilon > 0$, and consider the body

$$L = \mathrm{conv}\left(\frac{\theta_{min}}{1+\epsilon}K, \hat{z}_1, \ldots, \hat{z}_s\right). \tag{5.23}$$

We will show that

$$\frac{\theta_{min}}{1+\epsilon}K \subset L \subset (1+\epsilon)\theta_{max}K.$$

The first containment is obvious; for the second, observe that each $\hat{z}_i \in (1+\epsilon)\theta_{max}K$ for sufficiently large n since

$$\begin{aligned}
\|\hat{z}_i\|_K &= \frac{\|\hat{y}_i\|_K}{\|\hat{y}_i\|_{\mathcal{E}_v}} = \frac{\|x_i + u + v\|_K}{\|\hat{y}_i\|_{\mathcal{E}_v}} \\
&\leq (1+\epsilon)\frac{\|x_i + u + v\|_2}{\|\hat{y}_i\|_{\mathcal{E}_v}} \qquad \text{since } \|x_i\|_2 = \|x_i\|_K \text{ and } \|u\|, \|v\| \leq O(n^{-1/2}) \\
&= (1+\epsilon)\|\hat{z}_i\|_2 \\
&\leq (1+\epsilon)\theta_{max}.
\end{aligned}$$

It is also clear that $\hat{z}_i$ are the only contact points of L with $\mathcal{E}_v$ since all $\hat{z}_i \notin \frac{\theta_{min}}{1+\epsilon}B_2^n$. It remains to show that $\mathcal{E}_v$ is the John Ellipsoid of L. Applying R_v^{-1}, we see that $R_v^{-1}L$ has contact points $\hat{u}_i = R_v^{-1}\hat{z}_i$ with B_2^n, and we have already shown that these satisfy the conditions of John's theorem with weights $\hat{a}_i$.

The distance between L and K is now at most

$$(1+\epsilon)^2 \frac{\theta_{max}}{\theta_{min}} \leq (1+\epsilon)^2 \frac{\max_i \|\hat{y}_i\|_{\mathcal{E}_v}}{\min_i \|\hat{y}_i\|_{\mathcal{E}_v}} \leq (1+\epsilon)^3 \sqrt{\kappa(A_v)}$$

since $\|\hat{y}_i\|_{\mathcal{E}_v} \leq (1+\epsilon)\|x_i\|_{\mathcal{E}_v}$ for large n.

Combining (5.20) and (5.22) gives the required bound (5.19). □

Chapter 6

Restricted Invertibility

6.1 Introduction

In this chapter we study the following well-known theorem of Bourgain and Tzafriri.

Theorem 6.1 (Restricted Invertibility [14]). *There are universal constants $c, d > 0$, such that whenever B is an $n \times n$ matrix with unit length columns, one can find a subset $S \subset [n]$ of cardinality*

$$|S| \geq cn/\|B\|_2^2$$

for which

$$\sigma_{min}(B_S) \geq d, \tag{6.1}$$

where B_S is the submatrix of B with columns in S.

This theorem has had significant applications in the local theory of Banach spaces and in the study of convex bodies in high dimensions. It is also considered a step towards the resolution of the famous Kadison–Singer conjecture, which asks if there exists a partition of $[n]$ into a constant number of subsets $S_1, \ldots, S_k$ for which (6.1) holds. Recently, the theorem has attracted attention in numerical analysis due to its connection with the column subset selection problem, which seeks to select a 'representative' subset of columns from a given matrix. In particular, Tropp [47] has developed a randomized polynomial time algorithm which finds the subset S efficiently.

Bourgain and Tzafriri's proof of Theorem 6.1 uses probabilistic and functional analytic techniques and is non-constructive. In the original paper the theorem was shown to hold for $c = d \sim \frac{1}{10^{72}}$. Later on [15], the same authors proved it for $c = c(\epsilon) = c'\epsilon^2$ and $d = (1 + \epsilon)^{-1}$ for every $0 < \epsilon < 1$, where c' is a universal (tiny) constant. They were interested in the case when ϵ is small; the quadratic dependence of $c(\epsilon)$ on ϵ was shown to be necessary in [11]. In another regime, modern methods can be used to obtain the constants $c = 1/128$ and $d = 1/8\sqrt{2\pi}$ [16, 47].

Here we present a short proof that uses only basic linear algebra, achieves much better constants, and contains a deterministic algorithm for finding S. Our method of proof involves building the set S iteratively using a 'barrier' potential function, similar

to that in the proof of Theorem 3.1 in Chapter 3. The main conceptual differences between this proof and the earlier one are:

1. Here we use only one barrier instead of two, since we seek only a lower bound on the eigenvalues of vectors in S, the set that we construct.

2. The additional freedom arising from having only one barrier allows us to guarantee that the weights s_i are either 0 or 1 (i.e., a vector is either inside S or not) and that no vector is chosen twice.

Specifically, we prove the following generalization of Theorem 6.1[1]

Theorem 6.2. *Suppose* $v_1, \ldots v_m \in \mathbb{R}^n$, $\sum_i v_i v_i^T = I$, *and* $0 < \epsilon < 1$. *Let* $L : \ell_2^n \to \ell_2^n$ *be a linear operator. Then there is a subset* $S \subset [m]$ *of size* $|S| \geq \epsilon^2 \frac{\|L\|_F^2}{\|L\|_2^2}$ *for which* $\{Lv_i\}_i$ *is linearly independent and*

$$\lambda_{\min} \left(\sum_{i \in S} Lv_i(Lv_i)^T \right) > \frac{(1-\epsilon)^2 \|L\|_F^2}{m},$$

where $\lambda_{\min}$ *is computed on* $\mathrm{span}\{Lv_i\}_{i \in S}$.

This form of generalization was introduced by Vershynin [49] in his application to the study of contact points of convex bodies via John's decompositions of the identity. It says that given any such decomposition and any $L : \ell_2^n \to \ell_2^n$, there is a part of the decomposition on which L is well-invertible whose size is proportional to the stable rank $\frac{\|L\|_F^2}{\|L\|_2^2}$.

The original form of Bourgain and Tzafriri's theorem follows quickly from Theorem 6.2 with constants
$$c(\epsilon) = \epsilon^2 \quad \text{and} \quad d(\epsilon) = (1-\epsilon)^2$$

by taking $\{v_i\}$ from the standard basis $\{e_i\}_{i \leq n}$ and assuming $\|Le_i\| = 1$. This dominates previous bounds in all regimes, for ϵ small and large.

6.2   Proof of the Theorem

We will build the matrix $A = \sum_{i \in S} (Lv_i)(Lv_i)^T$ by an iterative process that adds one vector to S in each step. The process will be guided by the potential function

$$\Phi_b(A) = \sum_i (Lv_i)^T (A - bI)^{-1} (Lv_i)$$

$$= \mathrm{Tr} \left[L^T (A - bI)^{-1} L \right] \qquad \text{since} \sum_i v_i v_i^T = I,$$

---

[1]This statement is identical to Chapter 1 Theorem 1.10 if we take B to be the $n \times m$ matrix with columns Lv_i. We write it differently here in order to maintain consistency with the presentation in the literature [49, 14].

where the *barrier* b is a real number that varies from step to step.

Initially $A = 0$, the barrier is at $b = b_0 > 0$, and the potential is

$$\Phi_{b_0}(0) = \text{Tr}\left[L^T(0 - b_0 I)^{-1} L\right] = -\text{Tr}\left[L^T L\right] / b_0 = -\frac{\|L\|_F^2}{b_0}.$$

Each step of the process involves adding some rank-one matrix ww^T to A where $w \in \{Lv_i\}_{i \leq m}$ (if $w = Lv_j$ then this corresponds to adding j to S) and shifting the barrier towards zero by some fixed amount $\delta > 0$, without increasing the potential. Specifically, we want

$$\Phi_{b-\delta}(A + ww^T) \leq \Phi_b(A).$$

We will maintain the invariant that after k vectors have been added, A has exactly k nonzero eigenvalues, all greater than b. Keeping the potential small (in fact, sufficiently negative) will ensure that there is a suitable vector to add at each step.

In any step of the process, we are only interested in vectors w which add a new nonzero eigenvalue that is greater than $b' = b - \delta$. These are identified in the following lemma, where the notation $A \succeq B$ means that $A - B$ is positive semidefinite.

Lemma 6.3. *Suppose $A \succeq 0$ has k nonzero eigenvalues, all greater than $b' > 0$. If $w \neq 0$ and*

$$w^T(A - b'I)^{-1}w < -1 \tag{6.2}$$

then $A + ww^T$ has $k + 1$ nonzero eigenvalues greater than b'.

Proof. Let $\lambda_1 \geq \cdots \geq \lambda_k$ be the nonzero eigenvalues of A, and let $\lambda'_1 \geq \cdots \geq \lambda'_{k+1}$ be the $k + 1$ largest eigenvalues of $A + ww^T$. As the latter matrix is obtained from A by the addition of a rank one positive semi-definite matrix, their eigenvalues interlace:

$$\lambda'_1 \geq \lambda_1 \geq \lambda'_2 \geq \cdots \geq \lambda_k \geq \lambda'_{k+1}.$$

Consider the quantity

$$\text{Tr}\left[(A - b'I)^{-1}\right] = \sum_{i \leq k} \frac{1}{\lambda_i - b'} + \sum_{i > k} \frac{1}{0 - b'},$$

where we have written the positive and negative terms in the sum separately. By the Sherman–Morisson formula,

$$\text{Tr}\left[(A + ww^T - b'I)^{-1}\right] - \text{Tr}\left[(A - b'I)^{-1}\right] = -\frac{w^T(A - b'I)^{-2}w}{1 + w^T(A - b'I)^{-1}w}. \tag{6.3}$$

Since $w^T(A - b'I)^{-1}w < -1$, the denominator in the right-hand term is negative. The numerator is positive since $A - b'I$ is non-singular and $(A - b'I)^{-2} \succeq 0$. So, the right-hand side of (6.3) is positive.

On the other hand, a direct evaluation of this difference yields

$$0 < \operatorname{Tr}\left[(A + ww^T - b'I)^{-1}\right] - \operatorname{Tr}\left[(A - b'I)^{-1}\right]$$

$$= \frac{1}{\lambda'_{k+1} - b'} - \frac{1}{0 - b'} + \sum_{i=1}^{k} \frac{1}{\lambda'_i - b'} - \sum_{i=1}^{k} \frac{1}{\lambda_i - b'}$$

$$= \frac{1}{\lambda'_{k+1} - b'} + \frac{1}{b}.$$

As $\lambda'_{k+1} \geq 0$, this is only possible if $\lambda'_{k+1} > b'$, as desired. $\qquad\square$

The updated potential after one step, as the barrier moves from b to $b' = b - \delta$, can be calculated using the Sherman–Morisson formula:

$$\Phi_{b'}(A + ww^T) = \operatorname{Tr}\left[L^T(A - b'I + ww^T)^{-1}L\right]$$

$$= \operatorname{Tr}\left[L^T(A - b'I)^{-1}L\right] - \frac{\operatorname{Tr}\left[L^T(A - b'I)^{-1}ww^T(A - b'I)^{-1}L\right]}{1 + w^T(A - b'I)^{-1}w}$$

$$= \operatorname{Tr}\left[L^T(A - b'I)^{-1}L\right] - \frac{w^T(A - b'I)^{-1}LL^T(A - b'I)^{-1}w}{1 + w^T(A - b'I)^{-1}w}$$

$$= \Phi_{b'}(A) - \frac{w^T(A - b'I)^{-1}LL^T(A - b'I)^{-1}w}{1 + w^T(A - b'I)^{-1}w}.$$

To prevent an increase in potential, we want choose a w such that

$$\Phi_{b'}(A) - \frac{w^T(A - b'I)^{-1}LL^T(A - b'I)^{-1}w}{1 + w^T(A - b'I)^{-1}w} \leq \Phi_b(A). \tag{6.4}$$

We can now determine how small we need the potential to be in order to guarantee that a suitable w, which will allow us to keep on going, always exists.

Lemma 6.4. *Suppose $A_{n \times n}$ has k nonzero eigenvalues, all of which are greater than b, and let Z be the orthogonal projection onto the kernel of A. If*

$$\Phi_b(A) \leq -m - \frac{\|L\|_2^2}{\delta} \tag{6.5}$$

and

$$0 < \delta < b \leq \delta \frac{\|LZ\|_F^2}{\|L\|_2^2} \tag{6.6}$$

then there exists a vector $w \in \{Lv_i\}_{i \leq m}$ for which $A + ww^T$ has $k + 1$ nonzero eigenvalues greater than $b' = b - \delta$ and $\Phi_{b'}(A + ww^T) \leq \Phi_b(A)$.

Proof. [2] The vectors satisfying *both* of the inequalities (6.2) and (6.4) are precisely

---

[2]We would like to thank Pete Casazza for pointing out an important mistake in an earlier version of this proof.

those w for which

$$w^T(A - b'I)^{-1}LL^T(A - b'I)^{-1}w$$
$$\leq (\Phi_b(A) - \Phi_{b'}(A)) \cdot (-1 - w^T(A - b'I)^{-1}w).$$

We can show that such a w exists by taking the sum over all $w \in \{Lv_i\}_{i \leq m}$ and ensuring that the inequality holds in the sum, i.e., that

$$\text{Tr}\left[L^T(A - b'I)^{-1}LL^T(A - b'I)^{-1}L\right]$$
$$\leq (\Phi_b(A) - \Phi_{b'}(A)) \cdot (-m - \text{Tr}\left[L^T(A - b'I)^{-1}L\right]). \tag{6.7}$$

Let $\Delta_b := \Phi_b(A) - \Phi_{b'}(A)$. From the assumption $\Phi_b(A) \leq -m - \frac{\|L\|_2^2}{\delta}$ we immediately have

$$\text{Tr}\left[L^T(A - b'I)^{-1}L\right] = \Phi_b(A) - \Delta_b \leq -m - \frac{\|L\|_2^2}{\delta} - \Delta_b$$

and so (6.7) will follow from

$$\text{Tr}\left[L^T(A - b'I)^{-1}LL^T(A - b'I)^{-1}L\right] \leq \Delta_b \cdot \left(\frac{\|L\|_2^2}{\delta} + \Delta_b\right). \tag{6.8}$$

Noting that $LL^T \preceq \|L\|_2^2 I$, we can bound the left hand side as

$$\text{Tr}\left[L^T(A - b'I)^{-1}LL^T(A - b'I)^{-1}L\right] \leq \|L\|_2^2\text{Tr}\left[L^T(A - b'I)^{-2}L\right]. \tag{6.9}$$

Let P be the projection onto the image of A and let Z be the projection onto its kernel, so that $P + Z = I$. Let $\Phi_{b'}^P(A) = \text{Tr}\left[L^TP(A - b'I)^{-1}PL\right]$ and $\Phi_{b'}^Z(A) = \text{Tr}\left[L^TZ(A - b'I)^{-1}ZL\right]$ be the potentials computed on these subspaces. Since P, Z, A, $(A - b'I)^{-1}$, and $(A - b'I)^{-2}$ are mutually diagonalizable, we can write

$$\Phi_{b'}(A) = \Phi_{b'}^P(A) + \Phi_{b'}^Z(A), \qquad \Delta_b = \Delta_b^P + \Delta_b^Z, \qquad \text{and}$$

$$\text{Tr}\left[L^T(A - b'I)^{-2}L\right] = \text{Tr}\left[L^TP(A - b'I)^{-2}PL\right] + \text{Tr}\left[L^TZ(A - b'I)^{-2}ZL\right].$$

As $P(A - b'I)^{-1}P \succeq 0$ and $P(A - bI)^{-1}P \succeq 0$, it is easy to check that

$$(b - b')P(A - b'I)^{-2}P \preceq P(A - bI)^{-1}P - P(A - b'I)^{-1}P$$

which immediately gives

$$\|L\|_2^2\text{Tr}\left[L^TP(A - b'I)^{-2}PL\right] \leq \Delta_b^P \frac{\|L\|_2^2}{\delta}. \tag{6.10}$$

Thus, by (6.8), (6.9), and (6.10), we are done if we can show that

$$\|L\|_2^2\text{Tr}\left[L^TZ(A - b'I)^{-2}ZL\right] \leq (\Delta_b^P + \Delta_b^Z) \cdot \left(\frac{\|L\|_2^2}{\delta} + \Delta_b\right) - \Delta_b^P\frac{\|L\|_2^2}{\delta}.$$

Taking into account that $\Delta_b^P, \Delta_b^Z \geq 0$, this is implied by the statement

$$\|L\|_2^2 \mathrm{Tr}\left[L^T Z(A - b'I)^{-2} ZL\right] \leq \Delta_b^Z \cdot \left(\frac{\|L\|_2^2}{\delta} + \Delta_b^Z\right). \tag{6.11}$$

We now compute $\mathrm{Tr}\left[L^T Z(A - b'I)^{-2} ZL\right] = \frac{\|LZ\|_F^2}{b'^2}$ and

$$\Delta_b^Z = \mathrm{Tr}\left[L^T Z((A - bI)^{-1} - (A - b'I)^{-1}))ZL\right] = \delta \frac{\|LZ\|_F^2}{bb'}$$

which upon substituting and rearranging reduces (6.11) to

$$\|L\|_2^2 \leq \frac{\delta\|LZ\|_F^2}{b}$$

which we have assumed in (6.6). $\qquad\square$

Proof of Theorem 6.2. The requirements (6.5) and (6.6) of Lemma 6.4 are satisfied at the beginning of the process if we start with $b = b_0$ and δ satisfying

$$\Phi_{b_0}(0) = -\frac{\|L\|_F^2}{b_0} \leq -m - \frac{\|L\|_2^2}{\delta}$$

and

$$b_0 \leq \delta\frac{\|L\|_F^2}{\|L\|_2^2} \quad \text{since initially } A = 0 \text{ and } Z = \mathrm{Proj}_{\ker(A)} = I.$$

Both of these conditions are met if we take $0 < \epsilon < 1$ and

$$b_0 = \frac{(1 - \epsilon)\|L\|_F^2}{m} \qquad \text{and} \qquad \delta = \frac{(1 - \epsilon)\|L\|_2^2}{\epsilon m}.$$

They are maintained at every step of the process by Lemma 6.4 and the fact that the Frobenius norm $\|LZ\|_F^2$ decreases by at most $\|L\|_2^2$ in one step. Taking $\epsilon^2 \frac{\|L\|_F^2}{\|L\|_2^2}$ steps leaves the barrier at

$$\frac{(1 - \epsilon)\|L\|_F^2}{m} - \epsilon^2(1 - \epsilon)\frac{\|L\|_F^2}{\epsilon m}$$

which is the desired bound. $\qquad\square$

Chapter 7

The Kadison–Singer Conjecture

We conclude by drawing a connection between the main results of this thesis and an outstanding open problem in mathematics, the Kadison–Singer conjecture. This conjecture, which dates back to 1959, is equivalent to the well-known Paving Conjecture [4, 17] as well as to a stronger form of the restricted invertibility theorem in Chapter 6. The following formulation is due to Nik Weaver [51].

Conjecture 7.1. There are universal constants $\epsilon > 0, \delta > 0$, and $r \in \mathbb{N}$ for which the following statement holds. If $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ satisfy $\|\mathbf{v}_i\| \leq \delta$ for all i and

$$\sum_{i \leq m} \mathbf{v}_i \mathbf{v}_i^T = I,$$

then there is a partition $X_1, \ldots X_r$ of $\{1, \ldots, m\}$ for which

$$\left\| \sum_{i \in X_j} \mathbf{v}_i \mathbf{v}_i^T \right\| \leq 1 - \epsilon$$

for every $j = 1, \ldots, r$.

Suppose we had an 'unweighted' version of Theorem 3.2 which, assuming $\|\mathbf{v}_i\| \leq \delta$, guaranteed that the scalars s_i were all either 0 or some constant $\beta > 0$, and gave a constant approximation factor $\kappa < \beta$. Then we would have

$$I \preceq \beta \sum_{i \in S} \mathbf{v}_i \mathbf{v}_i^T \preceq \kappa \cdot I,$$

for $S = \{i : s_i \neq 0\}$, yielding a proof of Conjecture 7.1 with $r = 2$ and $\epsilon = \min\{1 - \frac{\kappa}{\beta}, \frac{1}{\beta}\}$ since

$$\left\| \sum_{i \in S} \mathbf{v}_i \mathbf{v}_i^T \right\| \leq \frac{\kappa}{\beta} \leq 1 - \epsilon \quad \text{and}$$

$$\left\| \sum_{i \in \overline{S}} \mathbf{v}_i \mathbf{v}_i^T \right\| = 1 - \lambda_{\min} \left(\sum_{i \in S} \mathbf{v}_i \mathbf{v}_i^T \right) \leq 1 - \frac{1}{\beta} \leq 1 - \epsilon.$$

Similarly, a strengthening of Theorem 6.2 which *partitions* $[m]$ into a constant number of sets σ_i also implies an affirmative resolution to the conjecture. Thus, Kadison–Singer is the common conjectured strengthening of the two main results of this thesis, and indeed they can be viewed as steps towards its resolution.

As a special case, a proof of the conjecture would also imply the existence of *unweighted* sparsifiers for the complete graph and other (sufficiently dense) edge-transitive graphs. It is also worth noting that the $\|\mathbf{v}_i\| \leq \delta$ condition when applied to vectors arising from a graph simply means that the effective resistances of all edges are bounded; thus, we would be able to conclude that any graph with sufficiently small resistances can be split into two graphs that approximate it spectrally.

Bibliography

[1] ACHLIOPTAS, D. Database-friendly random projections. In *PODS '01: Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (New York, NY, USA, 2001), ACM Press, pp. 274–281.

[2] ACHLIOPTAS, D., AND MCSHERRY, F. Fast computation of low rank matrix approximations. In *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing* (New York, NY, USA, 2001), ACM, pp. 611–618.

[3] AHLSWEDE, R., AND WINTER, A. Strong converse for identification via quantum channels. *Information Theory, IEEE Transactions on 48*, 3 (mar 2002), 569 –579.

[4] C. A. AKEMANN AND J. ANDERSON. Lyapunov theorems for operator algebras. *Mem. Amer. Math. Soc.*, 94, 1991.

[5] ALON, N., AND CHUNG, F. Explicit construction of linear sized tolerant networks. *Discrete Mathematics 72* (1988), 15–19.

[6] ARORA, S., HAZAN, E., AND KALE, S. A fast random sampling algorithm for sparsifying matrices. In *APPROX-RANDOM* (2006), pp. 272–279.

[7] BALL, K. An elementary introduction to modern convex geometry. In *in Flavors of Geometry* (1997), Univ. Press, pp. 1–58.

[8] BATSON, J. D., SPIELMAN, D. A., AND SRIVASTAVA, N. Twice-Ramanujan sparsifiers. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing* (New York, NY, USA, 2009), ACM, pp. 255–262.

[9] BENCZÚR, A. A., AND KARGER, D. R. Approximating s-t minimum cuts in $O(n^2)$ time. In *Proceedings of The Twenty-Eighth Annual ACM Symposium On The Theory Of Computing (STOC '96)* (New York, USA, May 1996), ACM Press, pp. 47–55.

[10] BENCZÚR, A. A., AND KARGER, D. R. Approximating s-t minimum cuts in $\tilde{O}(n^2)$ time. In *STOC* (1996), pp. 47–55.

[11] BERMAN, K., HALPERN, H., KAFTAL, V., AND WEISS, G. Matrix norm inequalities and the relative dixmier property. *Integral Equations and Operator Theory 11* (1988), 28–48.

[12] Bilu, Y., and Linial, N. Constructing expander graphs by 2–lifts and discrepancy vs. spectral gap. In *FOCS* (2004), pp. 404–412.

[13] B. Bollobas. *Modern Graph Theory*. Springer, July 1998.

[14] Bourgain, J., and Tzafriri, L. Invertibility of ŚlargeŠ submatrices with applications to the geometry of banach spaces and harmonic analysis. *Israel Journal of Mathematics 57* (1987), 137–224.

[15] Bourgain, J., and Tzafriri, L. On a problem of Kadison and Singer. *J. Reine Angew. Math. 420* (1991), 1–43.

[16] Casazza, P., and Tremain, J. Revisiting the Bourgain-Tzafriri Restricted Invertibility Theorem. *Operators and Matrices 3* (2009), 97–110.

[17] Peter G. Casazza and Janet C. Tremain. The Kadison–Singer problem in mathematics and engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2032–2039, 2006.

[18] Chew, P. There is a planar graph almost as good as the complete graph. In *SCG '86: Proceedings of the second annual symposium on Computational geometry* (1986), ACM, pp. 169–177.

[19] Chung, F. R. K. *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics. American Mathematical Society, 1997.

[20] H. Dette and W. J. Studden. Some new asymptotic properties for the zeros of Jacobi, Laguerre, and Hermite polynomials. *Constructive Approximation*, 11(2):227–238, 1995.

[21] P. Drineas and R. Kannan. Fast monte-carlo algorithms for approximate matrix multiplication. In *FOCS '01*, pages 452–459, 2001.

[22] P. Drineas and R. Kannan. Pass efficient algorithms for approximating large matrices. In *SODA '03*, pages 223–232, 2003.

[23] Frieze, A., Kannan, R., and Vempala, S. Fast monte-carlo algorithms for finding low-rank approximations. *J. ACM 51*, 6 (2004), 1025–1041.

[24] Giannopoulos, A., and Milman, V. Extremal problems and isotropic positions of convex bodies.

[25] Godsil, C., and Royle, G. *Algebraic Graph Theory*. Graduate Texts in Mathematics. Springer, 2001.

[26] Goldberg, A. V., and Tarjan, R. E. A new approach to the maximum flow problem. In *STOC '86: Proceedings of the eighteenth annual ACM symposium on Theory of computing* (New York, NY, USA, 1986), ACM, pp. 136–146.

[27] GOLUB, G. H., AND VAN LOAN, C. F. *Matrix Computations, 3rd. Edition.* The Johns Hopkins University Press, Baltimore, MD, 1996.

[28] GRUBER, P. M. Minimal ellipsoids and their duals. *Rendiconti del Circolo Matematico di Palermo 37*, 1 (1988), 35–64.

[29] GUATTERY, S., AND MILLER, G. L. Graph embeddings and laplacian eigenvalues. *SIAM J. Matrix Anal. Appl. 21*, 3 (2000), 703–723.

[30] HOORY, S., LINIAL, N., AND WIGDERSON, A. Expander graphs and their applications. *Bulletin of the American Mathematical Society 43*, 4 (2006), 439–561.

[31] W. JOHNSON AND J. LINDENSTRAUSS. Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math.*, 26:189–206, 1984.

[32] KHANDEKAR, R., RAO, S., AND VAZIRANI, U. Graph partitioning using single commodity flows. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing* (New York, NY, USA, 2006), ACM, pp. 385–390.

[33] LUBOTZKY, A., PHILLIPS, R., AND SARNAK, P. Ramanujan graphs. *Combinatorica 8*, 3 (1988), 261–277.

[34] G. LUGOSI. Concentration-of-measure inequalities, 2003. Available at http://www.econ.upf.edu/~lugosi/anu.ps.

[35] MARGULIS, G. A. Explicit group theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problems of Information Transmission 24*, 1 (July 1988), 39–46.

[36] NILLI, A. On the second eigenvalue of a graph. *Discrete Mathematics 91*, 2 (1991), 207–210.

[37] REINGOLD, O., VADHAN, S., AND WIGDERSON, A. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. *Foundations of Computer Science, Annual IEEE Symposium on 0* (2000), 3.

[38] RUDELSON, M. Contact points of convex bodies. *Israel J. Math 101*, 1 (1997), 92–124.

[39] RUDELSON, M. Random vectors in the isotropic position. *J. of Functional Analysis 163*, 1 (1999), 60–72.

[40] RUDELSON, M., AND VERSHYNIN, R. Sampling from large matrices: An approach through geometric functional analysis. *J. ACM 54*, 4 (2007), 21.

[41] SPIELMAN, D. A., AND SRIVASTAVA, N. Graph sparsification by effective resistances. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing* (2008), pp. 563–568. Full version available at http://arXiv.org/abs/0803.0929.

[42] SPIELMAN, D. A., AND TENG, S.-H. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM Symposium on Theory of Computing (STOC-04)* (2004), pp. 81–90.

[43] SPIELMAN, D. A., AND TENG, S.-H. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM Symposium on Theory of Computing (STOC-04)* (New York, June 13–15 2004), ACM Press, pp. 81–90. Full version available at `http://arxiv.org/abs/cs.DS/0310051`.

[44] SPIELMAN, D. A., AND TENG, S.-H. Nearly-linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. Available at `http://www.arxiv.org/abs/cs.NA/0607105`, 2006.

[45] SPIELMAN, D. A., AND TENG, S.-H. Spectral sparsification of graphs. *CoRR abs/0808.4134* (2008). Available at `http://arxiv.org/abs/0808.4134`.

[46] THORUP, M., AND ZWICK, U. Approximate distance oracles. pp. 183–192.

[47] TROPP, J. A. Column subset selection, matrix factorization, and eigenvalue optimization. In *SODA '09: Proceedings of the Nineteenth Annual ACM -SIAM Symposium on Discrete Algorithms* (Philadelphia, PA, USA, 2009), Society for Industrial and Applied Mathematics, pp. 978–986.

[48] TROPP, J. A. Topics in Sparse Approximation. Ph.D. dissertation, Computational and Applied Mathematics, Univ. Texas at Austin, Aug. 2004.

[49] VERSHYNIN, R. John's decompositions: Selecting a large part. *Israel Journal of Mathematics 122* (2001), 253–277.

[50] VERSHYNIN, R. A note on sums of independent random matrices after ahlswede-winter. Available at http://www-personal.umich.edu/ romanv/teaching/reading-group/ahlswede-winter.pdf, 2009.

[51] NIK WEAVER. The Kadison-Singer problem in discrepancy theory. *Discrete Mathematics*, 278(1–3):227 – 239, 2004.