SIAM REVIEW
Vol. 56, No. 2, pp. 315–334

Twice-Ramanujan Sparsifiers*

Joshua Batson†
Daniel A. Spielman‡
Nikhil Srivastava§

Abstract. A sparsifier of a graph is a sparse graph that approximates it. A spectral sparsifier is one that approximates it spectrally, which means that their Laplacian matrices have similar quadratic forms. We prove that every graph has a spectral sparsifier with a number of edges linear in its number of vertices. In particular, we prove that for every $\epsilon \in (0,1)$ and every undirected, weighted graph $G = (V, E, w)$ on n vertices, there exists a weighted graph $H = (V, F, \tilde{w})$ with at most $\lceil (n-1)/\epsilon^2 \rceil$ edges such that for every $x \in \mathbb{R}^V$, $(1-\epsilon)^2 \cdot x^T L_G x \leq x^T L_H x \leq (1+\epsilon)^2 \cdot x^T L_G x$, where L_G and L_H are the Laplacian matrices of G and H, respectively. We give an elementary deterministic polynomial time algorithm for constructing H. This result is a special case of a significantly more general theorem which provides sparse approximations of general positive semidefinite matrices: given any real matrix $B_{n\times m}$ and $\epsilon \in (0,1)$, there is a nonnegative diagonal matrix $S_{m\times m}$ with at most $\lceil n/\epsilon^2 \rceil$ nonzero entries such that $(1-\epsilon)^2 BB^T \preceq BSB^T \preceq (1+\epsilon)^2 BB^T$.

Key words. graph sparsification, spectral graph theory, spectral sparsification, rank-one updates, barrier method, paving conjecture, sparse quadratic forms

AMS subject classifications. Primary, 68R10; Secondary, 15B48

DOI. 10.1137/130949117

1. Introduction.

A sparsifier of a weighted graph G is a sparse graph H that is similar to G in some useful way. Many notions of similarity have been considered. For example, graph spanners [30, 9] have the property that the distance between every pair of vertices in H is approximately the same as in G. Benczúr and Karger's [4] cut sparsifiers have the property that the weight of the boundary of every set of vertices is approximately the same in H as in G.

In this paper, we consider a strengthening of cut sparsification: spectral sparsification. Following Spielman and Teng [35, 37], we say that H is a κ-approximation of a weighted, undirected graph $G = (V, E, w)$ if there exist $b \geq a > 0$ with $b/a \leq \kappa$ such that for all $x \in \mathbb{R}^V$,

$$a \cdot x^T L_G x \leq x^T L_H x \leq b \cdot x^T L_G x, \tag{1}$$

where L_G and L_H are the Laplacian matrices of G and H. In the language of numerical

*Published electronically May 8, 2014. This paper originally appeared in *SIAM Journal on Computing*, Volume 41, Number 6, 2012, pages 1704–1721. This material is based upon work supported by the National Science Foundation under grant CCF-0634957. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

http://www.siam.org/journals/sirev/56-2/94911.html

†Department of Mathematics, MIT, Cambridge, MA 02139 (joshuab@mac.com). Work on this paper was performed while this author was at Yale College.

‡Program in Applied Mathematics and Department of Computer Science, Yale University, New Haven, CT 06520 (spielman@cs.yale.edu).

§Microsoft Research, Bangalore, KA 560025, India (niksri@microsoft.com).

linear algebra, condition (1) simply says that the relative condition number of L_G and L_H is at most κ.

We recall that the Laplacian matrix of a weighted undirected graph $G = (V, E, w)$ may be defined as a sum of outer products over its edges:

$$L_G = \sum_{(u,v)\in E} w_{u,v}(\chi_u - \chi_v)(\chi_u - \chi_v)^T,$$

where $w_{u,v}$ is the weight of edge (u, v) and $\chi_u \in \mathbb{R}^v$ is the characteristic vector of vertex u (with a 1 on coordinate u and zeros elsewhere). Thus, its quadratic form is given by

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

This has a clear combinatorial meaning when x is the characteristic vector of a subset $S \subset V$ of the vertices (i.e., $x_u = 1$ if $u \in S$, and $x_u = 0$ otherwise). In this case, $x^T L_G x$ is the sum of the weights of edges with exactly one endpoint in S:

$$x^T L_G x = \sum_{(u,v)\in E, u\in S, v\notin S} w_{u,v}.$$

Benczur and Karger's cut sparsifiers are designed to approximately preserve exactly these quantities. Extending the approximation guarantee to all vectors $x \in \mathbb{R}^V$ as prescribed in (1) approximately preserves many additional properties, most notably the eigenvalues of L_G and the effective resistances between pairs of vertices in G when it is viewed as a resistive circuit. The reason is that all of these quantities have a variational characterization in terms of $x^T L_G x$, which is the natural notion of energy for functions x on the vertices of G.

Our main result is that every undirected weighted graph on n vertices has a $(\frac{1+\epsilon}{1-\epsilon})^2$-approximation with at most n/ϵ^2 edges, for every $\epsilon \in (0, 1)$. We henceforth use the standard notation $A \preceq B$ to indicate that $x^T Ax \leq x^T Bx$ for every vector x, or, equivalently, that $B - A$ is positive semidefinite.

THEOREM 1.1. *For every $\epsilon \in (0, 1)$, every undirected weighted graph $G = (V, E, w)$ on n vertices contains a weighted subgraph $H = (V, F, \tilde{w})$ with $\lceil (n-1)/\epsilon^2 \rceil$ edges that satisfies*

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

Our proof provides a deterministic greedy algorithm for computing the graphs H in time $O(n^3 m/\epsilon^2)$. 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 [37].

In section 4 we show that the quantitative bounds achieved by Theorem 1.1 are within a factor of four of optimal. The examples that demonstrate this near-optimality are simply the complete graphs K_n. The more restricted problem of finding *unweighted, regular* approximations of complete graphs is better understood, and the optimal graphs for this problem are called *Ramanujan graphs* [24, 26]. The title of this paper reflects the fact that our sparsifiers are at most twice as dense as the Ramanujan graphs achieving the same approximation quality.

We derive Theorem 1.1 from a more general statement, which says that every sum of outer products of vectors can be spectrally approximated by a sparse sum which has almost the same quadratic form.

THEOREM 1.2. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ be vectors in $\mathbb{R}^n$ with $\sum_{i \le m} \mathbf{v}_i \mathbf{v}_i^T = M$. For every $\epsilon \in (0,1)$, there exist scalars $s_i \ge 0$ with $|\{i : s_i \neq 0\}| \le \lceil \mathrm{rank}(M)/\epsilon^2 \rceil$ so that*

$$(1-\epsilon)^2 M \preceq \sum_{i \le m} s_i \mathbf{v}_i \mathbf{v}_i^T \preceq (1+\epsilon)^2 M.$$

This is seen to be equivalent to the statement written in the abstract by considering the $n \times m$ matrix B with columns $\mathbf{v}_1, \ldots, \mathbf{v}_m$, which satisfies $BB^T = M$, and taking S to be a diagonal matrix containing the scalars $s_1, \ldots, s_m \ge 0$ provided by the theorem.

1.1. Related Work. Spielman and Teng [35, 37] introduced the notion of sparsification that we consider and proved that $(1+\epsilon)$-approximations with $\widetilde{O}\left(n/\epsilon^2\right)$ edges could be constructed in $\widetilde{O}\left(m\right)$ time, where we write $\widetilde{O}\left(m\right)$ to indicate $O(m \log^c m)$ for some constant c. They used these sparsifiers to obtain a nearly linear time algorithm for solving systems of linear equations in symmetric, diagonally dominant matrices [35, 36].

Spielman and Teng were inspired by the notion of sparsification introduced by Benczúr and Karger [4] for cut problems, which only requires inequality (1) to hold for all $x \in \{0,1\}^V$. Benczúr and Karger showed how to construct graphs H meeting this guarantee with $O\left(n \log n/\epsilon^2\right)$ edges in $O\left(m \log^3 n\right)$ time, for $b/a = (1+\epsilon)/(1-\epsilon)$.

Spielman and Srivastava [34] proved the existence of spectral sparsifiers with $O\left(n \log n/\epsilon^2\right)$ edges and showed how to construct them in $\widetilde{O}\left(m\right)$ time. They conjectured that it should be possible to find such sparsifiers with only $O\left(n/\epsilon^2\right)$ edges. We affirmatively resolve this conjecture.

Partial progress was made toward the conjecture by Goyal, Rademacher, and Vempala [15], who showed how to find graphs H with only $2n$ edges that $O\left(\log n\right)$-approximate bounded degree graphs. Kapralov and Panigrahy [17] have shown how to use graph spanners to construct spectral sparsifiers. Faster algorithms for finding spectral sparsifiers have been discovered by Koutis, Levin, and Peng [21]. Algorithms that work in the streaming model have been developed by Goel, Kapralov, and Post and by Kelner and Levin [18].

We remark that all of these constructions were randomized. Ours is the first deterministic algorithm to achieve the guarantees of any of these papers.

The results we present in this paper have been improved in many ways. The main linear-algebraic result (Theorem 1.2) was extended from a theorem about sums of rank-one matrices to sums of arbitrary positive semidefinite matrices by de Carli Silva, Harvey, and Sato [7]. Kolla et al. [20] extended our algorithm to prove that every graph has a very good ultrasparsifier, that is, an approximation by a graph with $(1+\beta)n$ edges for β close to zero. These ultrasparsifiers provided evidence for the existence of very fast algorithms for solving systems of linear equations in symmetric, diagonally dominant matrices. They were also the inspiration for the discovery of such algorithms by Koutis, Miller, and Peng [22, 23].

Since many objects besides graphs can be faithfully represented as quadratic forms of sums of outer products of vectors, Theorem 1.2 implies useful sparse approximation results in several areas of mathematics. In particular, it was an important ingredient in recent significant progress in the following problems in geometry and analysis:

dimension reduction for finite point sets in in ℓ_1 (see Newman and Rabinovich [28]); isomorphic embedding of subspaces of L_p into ℓ_p for even p (see Schechtman [32]); approximation of arbitrary symmetric convex bodies by polytopes with few vertices (see Barvinok [3]); and existence of "low complexity" extension operators in Sobolev spaces (see Fefferman, Israel, and Luli [12]). For a more detailed exposition of these applications, we refer the reader to [27].

The techniques used to prove Theorem 1.2 have also proved to be very useful. In particular, the barrier function argument used in section 3.2 has been used to obtain an improvement [33] of Bourgain and Tzafriri's restricted invertibility theorem [5], give better algorithms for the column subset selection problem in numerical linear algebra [6], and obtain better concentration inequalities for the extreme singular values of a large class of random matrices with independent rows [38]. It was a key element of the recent solution of the Kadison–Singer problem [25].

2. Preliminaries.

2.1. Pseudoinverses and Square Roots. Suppose we are given a rank-r symmetric matrix M with eigendecomposition

$$M = \sum_{i=1}^{r} \lambda_i u_i u_i^T,$$

where $\lambda_1, \ldots, \lambda_r$ are the nonzero eigenvalues of M and $u_1, \ldots, u_r$ are a corresponding set of orthonormal eigenvectors. The *Moore–Penrose pseudoinverse* of M is then defined as

$$M^+ = \sum_{i=1}^{r} \frac{1}{\lambda_i} u_i u_i^T.$$

Notice that $\ker(M) = \ker(M^+)$ and that

$$MM^+ = M^+M = \sum_{i=1}^{r} u_i u_i^T,$$

which is simply the projection onto range(M).

If $M \succeq 0$, then each of the eigenvalues λ_i is nonnegative and has a square root. In this case, we may define the square root of M as

$$M^{\frac{1}{2}} = \sum_{i=1}^{r} \sqrt{\lambda_i} u_i u_i^T,$$

which is also positive semidefinite.

2.2. Formulas for Rank-One Updates. We use the following well-known theorem from linear algebra, which describes the behavior of the inverse of a matrix under rank-one updates (see [14, section 2.1.3]).

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*

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

3. The Main Result. At the heart of this work is the following purely linear algebraic theorem, which says that every decomposition of the identity into a sum of outer products has a sparse approximation. Such decompositions appear in many areas of mathematics and are also called isotropic sets, tight frames, and John's decompositions (when their mean is zero).

THEOREM 3.1. *Suppose $d>1$ and $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m$ are vectors in $\mathbb{R}^n$ with*

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

then there exist scalars $s_i\ge 0$ with $|\{i : s_i\neq 0\}|\le\lceil dn\rceil$ such that

$$\left(1-\frac{1}{\sqrt{d}}\right)^2 I \preceq \sum_{i\le m} s_i\mathbf{v}_i\mathbf{v}_i^T \preceq \left(1+\frac{1}{\sqrt{d}}\right)^2 I. \tag{3}$$

By taking $d=1/\epsilon^2$, we see that Theorem 3.1 is a special case of Theorem 1.2 with $M=I$. However, Theorem 3.1 can be used to easily prove Theorem 1.2 by a simple rescaling procedure as follows.

Proof of Theorem 1.2. Let $d=1/\epsilon^2$. Suppose we are given $\mathbf{v}_1,\ldots,\mathbf{v}_m\in\mathbb{R}^n$ with $\sum_{i\le m}\mathbf{v}_i\mathbf{v}_i^T = M$ and $\operatorname{rank}(M)=r$. Since $M\succeq 0$, its pseudoinverse $M^+\succeq 0$ has a square root $(M^+)^{\frac12}$. Consider the vectors $\mathbf{w}_i := (M^+)^{\frac12}\mathbf{v}_i$ for $i=1,\ldots,m$, and observe that

$$\sum_{i=1}^m \mathbf{w}_i\mathbf{w}_i^T = P,$$

where P is the rank-r projection onto the range of M. Since P is the identity on a copy of $\mathbb{R}^r$ and the condition (3) is invariant under change of basis, we may apply Theorem 3.1 to obtain scalars $s_i\ge 0$, at most $\lceil dr\rceil = \lceil r/\epsilon^2\rceil$ of which are nonzero, with

$$(1-\epsilon)^2 x^T x \le x^T\left(\sum_{i=1}^m s_i\mathbf{w}_i\mathbf{w}_i^T\right)x \le (1+\epsilon)^2 x^T x \qquad \forall x\in\operatorname{range}(P).$$

Since $\operatorname{range}(P)=\operatorname{range}(M^{\frac12})$, quantifying over all $x\in\operatorname{range}(P)$ is the same as quantifying over all $M^{\frac12}y$ with $y\in\mathbb{R}^n$, so the above is equivalent to

$$(1-\epsilon)^2 y^T M y \le y^T M^{\frac12}\left(\sum_{i=1}^m s_i(M^+)^{\frac12}\mathbf{v}_i\mathbf{v}_i^T(M^+)^{\frac12}\right)M^{\frac12}y \le (1+\epsilon)^2 y^T M y \qquad \forall y\in\mathbb{R}^n.$$

To complete the proof, observe that $M^{\frac12}(M^+)^{\frac12}=P$ and that $\mathbf{v}_i\in\operatorname{range}(M)$, so $P\mathbf{v}_i=\mathbf{v}_i$. ∎

The sparsification result for graphs follows immediately as a special case of Theorem 1.2.

Proof of Theorem 1.1. Fix $\epsilon \in (0,1)$. Assume without loss of generality that G is connected, and let

$$L_G = \sum_{(u,v)\in E} w_{u,v}(\chi_u - \chi_v)(\chi_u - \chi_v)^T$$

be its Laplacian. By inspecting the quadratic form $x^T L_G x$, it is easy to see that the nullspace of L_G is simply the span of the all-ones vector, so $\operatorname{rank}(L_G) = n-1$. Applying Theorem 1.2 yields scalars $s_{u,v} \geq 0$ with

$$(1-\epsilon)^2 L_G \preceq \sum_{(u,v)\in E} s_{u,v} w_{u,v}(\chi_u - \chi_v)(\chi_u - \chi_v)^T \preceq (1+\epsilon)^2 L_G.$$

The matrix in the middle is just the Laplacian matrix of the weighted subgraph H of G with weights $\tilde{w}_{u,v} = s_{u,v} w_{u,v}$, at most $\lceil (n-1)/\epsilon^2 \rceil$ of which are nonzero by Theorem 1.2, as desired. □

It is worth mentioning that the above reduction from spectral approximations of graphs to spectral approximations of the identity is essentially the same as the one in [34]. In that paper, the authors consider the symmetric projection matrix $\Pi = B L_G^+ B^T$ instead of the identity, where $B_{m\times n}$ is the signed edge-vertex incidence matrix of G with rows $\{(\chi_u - \chi_v)^T\}_{(u,v)\in E}$. The projection Π has a decomposition $\Pi = \sum_{e\in G} \pi_e \pi_e^T$ for certain vectors π_e which correspond to edges of G; they show, by a concentration lemma of Rudelson [31], that randomly sampling $O(n \log n)$ of the π_e with probabilities proportional to $\|\pi_e\|^2$ gives a matrix $\tilde{\Pi}$ that approximates Π in the spectral norm and corresponds to a graph sparsifier, with high probability. We note that the vectors π_e are the same as the vectors $(L_G)^{\frac{1}{2}}(\chi_u - \chi_v)$ considered above after an appropriate change of basis, and their squared norms have a physical meaning: they are the effective resistances across the edges in the resistive circuit corresponding to G. This paper is different from [34] in two ways: we eliminate Π in order to simplify notation, since we are no longer following the intuition of sampling by effective resistances, and, instead of Rudelson's sampling lemma, we use Theorem 3.1 to *deterministically* select $O(n)$ vectors.

The rest of this section is devoted to proving Theorem 3.1. The proof is constructive and yields a deterministic polynomial time algorithm for finding the scalars s_i, which can then be used to sparsify graphs, as advertised.

Given vectors $\{\mathbf{v}_i\}_{i\leq m}$ with $\sum_i \mathbf{v}_i \mathbf{v}_i^T = I$, our goal is to choose a small set of coefficients s_i so that $A = \sum_i s_i \mathbf{v}_i \mathbf{v}_i^T$ is well-conditioned, i.e., the ratio of its largest and smallest eigenvalues is small. We will build the matrix A in steps, starting with $A = 0$ and adding one vector $s_i \mathbf{v}_i \mathbf{v}_i^T$ at a time. We will show that at every step there is a choice of $\mathbf{v}_i$ and scaling s_i which causes the eigenvalues of A to increase at a steady and predictable rate, in a way which we will make precise soon. This will cause the condition number to converge to 1 as we continue the process, and sparsity of the chosen weights s_i will be guaranteed simply by terminating the process after $O(n)$ steps.

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 into the structure of the proof and explain the origin of the "twice-Ramanujan" bound $\left(\frac{\sqrt{d}+1}{\sqrt{d}-1}\right)^2$, which appears in our final result.

3.1. Intuition for the Proof. It is well known that the eigenvalues of $A + \mathbf{v}\mathbf{v}^T$ interlace those of A. In fact, the new eigenvalues can be determined exactly by looking

at the characteristic polynomial of $A + \mathbf{v}\mathbf{v}^T$, which is computed using Lemma 2.2 as follows:

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

where λ_i are the eigenvalues of A and u_j are the corresponding eigenvectors. The polynomial $p_{A+\mathbf{v}\mathbf{v}^T}(x)$ 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 therefore do not "move" upon adding $\mathbf{v}\mathbf{v}^T$.
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 (see Figure 1), which gives intuition into where these new eigenvalues are located.

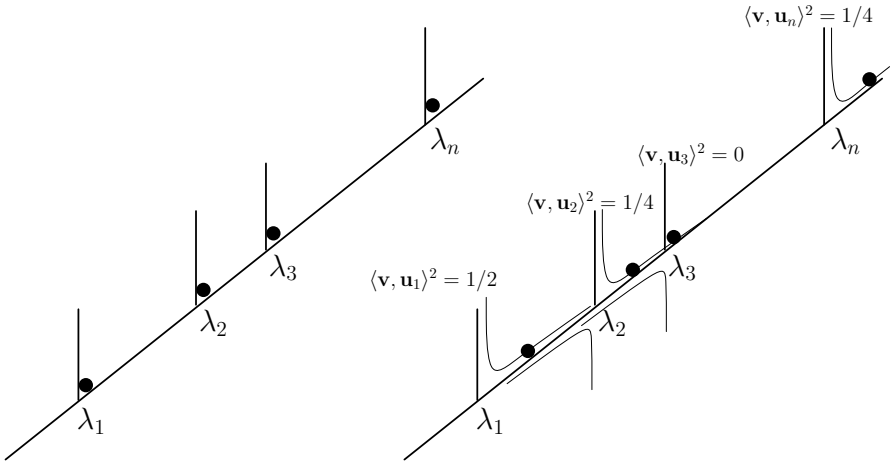

Fig. 1 *Physical model of interlacing eigenvalues.*

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 push 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 the total "downward pull" $f(\lambda)$ being equal to zero.

With this physical model in mind, let us 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 $\mathbf{v} \in \{\mathbf{v}_i\}_{i \le m}$ 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[1] 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:

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

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 a standard family of orthogonal polynomials—the *associated Laguerre polynomials* [11]. 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 [11] to be

$$\frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}},$$

which is exactly what we want.

It is worth mentioning that the associated Laguerre polynomials and the bound above also arise naturally in random matrix theory. In particular, let $g_1, \ldots, g_m \sim \mathcal{N}(0, I/n)$ be independent Gaussian random vectors in $\mathbb{R}^n$ (i.e., each with independent normal entries with variance $1/n$), and suppose $m/n = d > 1$. Then, it is well known

[1] For concreteness, we remark that this "average" vector would be precisely $\mathbf{v}_{\mathrm{avg}} = \frac{1}{\sqrt{m}} \sum_j u_j$.

[10] that the eigenvalues of the Wishart ensemble

$$\frac{1}{m}\sum_{i\le m} g_i g_i^T$$

are contained with high probability in the interval

$$[(1-1/\sqrt{d})^2 - o(1), (1+1/\sqrt{d})^2 + o(1)],$$

where the $o(1)$ terms tend to zero as $m, n \to \infty$ while keeping $m/n = d$ fixed. Intuitively, this happens because each g_i is rotationally invariant, so each random rank-one update $g_i g_i^T$ has about the same projection on each of the eigenvectors of the sum of the previous terms $\sum_{j\le i-1} g_j g_j^T$, which is similar to the "expected" case described above.

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 $\mathbf{v}_{\mathsf{avg}}$) as long as we are allowed to add arbitrary fractional amounts of the $\mathbf{v}_i\mathbf{v}_i^T$ via the weights $s_i \ge 0$. 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.

3.2. 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 previous section.

DEFINITION 3.2. *For $u, l \in \mathbb{R}$ and A a symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, define*

$$\Phi^u(A) \stackrel{\text{def}}{=} \operatorname{Tr}(uI - A)^{-1} = \sum_i \frac{1}{u - \lambda_i} \quad \textit{(upper potential)},$$

$$\Phi_l(A) \stackrel{\text{def}}{=} \operatorname{Tr}(A - lI)^{-1} = \sum_i \frac{1}{\lambda_i - l} \quad \textit{(lower potential)}.$$

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 the fact that they reflect the locations of all the eigenvalues simultaneously: for instance, $\Phi^u(A) \le 1$ implies that no λ_i is within distance 1 of u, no two 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[2] u_0, l_0, δ_U, δ_L, ϵ_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, δ_U, δ_L, ϵ_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.

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 toward 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.

[2]We suggest that, on first reading the paper, 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 3.6.

LEMMA 3.3 (upper barrier shift). *Suppose that* $\lambda_{\max}(A) < u$, *and* $\mathbf{v}$ *is any vector. If*

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

then

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

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 $U_A(\mathbf{v})$ 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

$$\begin{aligned}
&\Phi^{u+\delta_U}(A + t\mathbf{v}\mathbf{v}^T) = \operatorname{Tr}(u'I - A - t\mathbf{v}\mathbf{v}^T)^{-1} \\
&= \operatorname{Tr}\left((u'I - A)^{-1} + \frac{t(u'I - A)^{-1}\mathbf{v}\mathbf{v}^T(u'I - A)^{-1}}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}\right) \\
&= \operatorname{Tr}(u'I - A)^{-1} + \frac{t\operatorname{Tr}(\mathbf{v}^T(u'I - A)^{-1}(u'I - A)^{-1}\mathbf{v})}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}} \\
&\qquad \text{since Tr is linear and } \operatorname{Tr}(XY) = \operatorname{Tr}(YX) \\
&= \Phi^{u+\delta_U}(A) + \frac{t\mathbf{v}^T(u'I - A)^{-2}\mathbf{v}}{1 - t\mathbf{v}^T(u'I - A)^{-1}\mathbf{v}} \\
&= \Phi^u(A) - (\Phi^u(A) - \Phi^{u+\delta_U}(A)) + \frac{\mathbf{v}^T(u'I - A)^{-2}\mathbf{v}}{1/t - \mathbf{v}^T(u'I - A)^{-1}\mathbf{v}}.
\end{aligned}$$

As $U_A(\mathbf{v}) > \mathbf{v}^T(u'I - A)^{-1}\mathbf{v}$, the last term is finite for $1/t \geq U_A(\mathbf{v})$. By now, substituting any $1/t \geq U_A(\mathbf{v})$, we find that $\Phi^{u+\delta_U}(A + t\mathbf{v}\mathbf{v}^T) \leq \Phi^u(A)$. This also tells us that $\lambda_{\max}(A + t\mathbf{v}\mathbf{v}^T) < u + \delta_U$, because 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. ∎

The second lemma concerns 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 toward 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 3.4 (lower barrier shift). *Suppose that* $\lambda_{\min}(A) > l$, $\Phi_l(A) \leq 1/\delta_L$, *and* $\mathbf{v}$ *is any vector. If*

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

then

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

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

$$\begin{aligned}
&\Phi_{l+\delta_L}(A + t\mathbf{v}\mathbf{v}^T) = \operatorname{Tr}(A + t\mathbf{v}\mathbf{v}^T - l'I)^{-1} \\
&= \operatorname{Tr}\left((A - l'I)^{-1} - \frac{t(A - l'I)^{-1}\mathbf{v}\mathbf{v}^T(A - l'I)^{-1}}{1 + t\mathbf{v}^T(A - l')^{-1}\mathbf{v}}\right) \\
&= \operatorname{Tr}(A - l'I)^{-1} - \frac{t\operatorname{Tr}(\mathbf{v}^T(A - l'I)^{-1}(A - l'I)^{-1}\mathbf{v})}{1 + t\mathbf{v}^T(A - l'I)^{-1}\mathbf{v}} \\
&= \Phi_{l+\delta_L}(A) - \frac{t\mathbf{v}^T(A - l'I)^{-2}\mathbf{v}}{1 + t\mathbf{v}^T(A - l'I)^{-1}\mathbf{v}} \\
&= \Phi_l(A) + (\Phi_{l+\delta_L}(A) - \Phi_l(A)) - \frac{\mathbf{v}^T(A - l'I)^{-2}\mathbf{v}}{1/t + \mathbf{v}^T(A - l'I)^{-1}\mathbf{v}}.
\end{aligned}$$

Rearranging shows that $\Phi_{l+\delta_L}(A + t\mathbf{v}\mathbf{v}^T) \leq \Phi_l(A)$ when $1/t \leq L_A(\mathbf{v})$. ∎

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 $\mathbf{v}_{\mathsf{avg}}$. The success of the averaging argument crucially relies on the fact that the trace of a matrix can be written as the sum of its quadratic form over any decomposition of the identity. Notice that the use of variable weights t, from which the eventual s_i arise, is essential to this part of the proof.

LEMMA 3.5 (both barriers). *If $\lambda_{\max}(A) < u$, $\lambda_{\min}(A) > l$, $\Phi^u(A) \leq \epsilon_U$, $\Phi_l(A) \leq \epsilon_L$, and ϵ_U, ϵ_L, δ_U, and δ_L satisfy*

$$0 \leq \frac{1}{\delta_U} + \epsilon_U \leq \frac{1}{\delta_L} - \epsilon_L, \tag{4}$$

then there exist an i and positive t for which

$$\begin{aligned}
L_A(\mathbf{v}_i) \geq 1/t \geq U_A(\mathbf{v}_i), \\
\lambda_{\max}(A + t\mathbf{v}_i\mathbf{v}_i^T) < u + \delta_U, \quad \textit{and} \\
\lambda_{\min}(A + t\mathbf{v}_i\mathbf{v}_i^T) > l + \delta_L.
\end{aligned}$$

Proof. We will show that

$$\sum_i L_A(\mathbf{v}_i) \geq \sum_i U_A(\mathbf{v}_i),$$

from which the claim will follow by Lemmas 3.3 and 3.4. We begin by bounding

$$\begin{aligned}
\sum_i U_A(\mathbf{v}_i) &= \frac{\sum_i \mathbf{v}_i^T((u+\delta_U)I - A)^{-2}\mathbf{v}_i}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \sum_i \mathbf{v}_i^T((u+\delta_U)I - A)^{-1}\mathbf{v}_i \\
&= \frac{((u+\delta_U)I - A)^{-2} \bullet (\sum_i \mathbf{v}_i\mathbf{v}_i^T)}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + ((u+\delta_U)I - A)^{-1} \bullet \left(\sum_i \mathbf{v}_i\mathbf{v}_i^T\right) \\
&= \frac{\mathrm{Tr}((u+\delta_U)I - A)^{-2}}{\Phi^u(A) - \Phi^{u+\delta_U}(A)} + \mathrm{Tr}((u+\delta_U)I - A)^{-1} \\
&\quad \text{since } \sum_i \mathbf{v}_i\mathbf{v}_i^T = I \text{ and } X \bullet I = \mathrm{Tr}(X) \\
&= \frac{\sum_i (u+\delta_U-\lambda_i)^{-2}}{\sum_i (u-\lambda_i)^{-1} - \sum_i (u+\delta_U-\lambda_i)^{-1}} + \Phi^{u+\delta_U}(A) \\
&= \frac{\sum_i (u+\delta_U-\lambda_i)^{-2}}{\delta_U \sum_i (u-\lambda_i)^{-1}(u+\delta_U-\lambda_i)^{-1}} + \Phi^{u+\delta_U}(A) \\
&\le 1/\delta_U + \Phi^{u+\delta_U}(A) \\
&\quad \text{as } \sum_i (u-\lambda_i)^{-1}(u+\delta_U-\lambda_i)^{-1} \ge \sum_i (u+\delta_U-\lambda_i)^{-2} \\
&\le 1/\delta_U + \Phi^u(A) \le 1/\delta_U + \epsilon_U.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\sum_i L_A(\mathbf{v}_i) &= \frac{\sum_i \mathbf{v}_i^T(A - (l+\delta_L))^{-2}\mathbf{v}_i}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - \sum_i \mathbf{v}_i^T(A - (l+\delta_L)I)^{-1}\mathbf{v}_i \\
&= \frac{(A - (l+\delta_L)I)^{-2} \bullet (\sum_i \mathbf{v}_i\mathbf{v}_i^T)}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - (A - (l+\delta_L)I)^{-1} \bullet \left(\sum_i \mathbf{v}_i\mathbf{v}_i^T\right) \\
&= \frac{\mathrm{Tr}(A - (l+\delta_L)I)^{-2}}{\Phi_{l+\delta_L}(A) - \Phi_l(A)} - \mathrm{Tr}(A - (l+\delta_L)I)^{-1} \\
&\quad \text{since } \sum_i \mathbf{v}_i\mathbf{v}_i^T = I \text{ and } X \bullet I = \mathrm{Tr}(X) \\
&= \frac{\sum_i (\lambda_i - l - \delta_L)^{-2}}{\sum_i (\lambda_i - l - \delta_L)^{-1} - \sum_i (\lambda_i - l)^{-1}} - \sum_i (\lambda_i - l - \delta_L)^{-1} \\
&\ge 1/\delta_L - \sum_i (\lambda_i - l)^{-1} = 1/\delta_L - \epsilon_L
\end{aligned}$$

by Claim 3.6.

Putting these together, we find that

$$\sum_i U_A(\mathbf{v}_i) \le \frac{1}{\delta_U} + \epsilon_U \le \frac{1}{\delta_L} - \epsilon_L \le \sum_i L_A(\mathbf{v}_i),$$

as desired. □

CLAIM 3.6. *If $\lambda_i > l$ for all i, $0 \le \sum_i (\lambda_i - l)^{-1} \le \epsilon_L$, and $1/\delta_L - \epsilon_L \ge 0$, then*

$$\frac{\sum_i (\lambda_i - l - \delta_L)^{-2}}{\sum_i (\lambda_i - l - \delta_L)^{-1} - \sum_i (\lambda_i - l)^{-1}} - \sum_i \frac{1}{\lambda_i - l - \delta_L} \ge \frac{1}{\delta_L} - \sum_i \frac{1}{\lambda_i - l}. \tag{5}$$

Proof. We have

$$\delta_L \le 1/\epsilon_L \le \lambda_i - l$$

for every i. So, the denominator of the leftmost term on the left-hand side is positive, and the claimed inequality is equivalent to

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

which, by moving the first term on the right-hand side to the left-hand side, is just

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

By Cauchy–Schwarz,

$$\begin{aligned}
\left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)(\lambda_i - l)} \right)^2 &\le \left(\delta_L \sum_i \frac{1}{\lambda_i - l} \right) \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right) \\
&\le (\delta_L \epsilon_L) \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right) \\
&\qquad \text{since } \sum (\lambda_i - l)^{-1} \le \epsilon_L \\
&\le 1 \left(\delta_L \sum_i \frac{1}{(\lambda_i - l - \delta_L)^2 (\lambda_i - l)} \right) \\
&\qquad \text{since } \frac{1}{\delta_L} - \epsilon_L \ge 0,
\end{aligned}$$

and so (5) is established. ∎

Proof of Theorem 3.1. All we need to do now is set $\epsilon_U, \epsilon_L, \delta_U$, and δ_L in a manner that satisfies Lemma 3.5 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

$$L_{A^{(q)}}(\mathbf{v}_i) \ge U_{A^{(q)}}(\mathbf{v}_i)$$

(such a vector is guaranteed to exist by Lemma 3.5) and setting $A^{(q+1)} = A^{(q)} + t\mathbf{v}_i\mathbf{v}_i^T$ for any $t \geq 0$ satisfying

$$L_{A^{(q)}}(\mathbf{v}_i) \geq \frac{1}{t} \geq U_{A^{(q)}}(\mathbf{v}_i).$$

It is sufficient to take

$$\begin{aligned} \delta_L &= 1, & \epsilon_L &= \frac{1}{\sqrt{d}}, & l_0 &= -n/\epsilon_L, \\ \delta_U &= \frac{\sqrt{d}+1}{\sqrt{d}-1}, & \epsilon_U &= \frac{\sqrt{d}-1}{d+\sqrt{d}}, & u_0 &= n/\epsilon_U. \end{aligned}$$

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 (4) 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

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

as desired. □

To turn this proof into an algorithm, one must first compute the vectors $\mathbf{v}_i$, which can be done in time $O\left(n^2m\right)$. 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 edge to add in each iteration by computing $U_A(\mathbf{v}_i)$ and $L_A(\mathbf{v}_i)$ for each edge, which can be done in time $O\left(n^2m\right)$. As we run for dn iterations, the total time of the algorithm is $O\left(dn^3m\right)$.

4. Sparsifiers of the Complete Graph. Let $G = (V, E)$ be the complete graph on n vertices, and let $H = (V, F, w)$ be a weighted graph of average degree d that $(1+\epsilon)$-approximates G. As $x^T L_G x = n\|x\|^2$ for every x orthogonal to $\mathbf{1}$, it is immediate that every vertex of H has weighted degree between n and $(1+\epsilon)n$. Thus, one should think of H as being an expander graph in which each edge weight has been multiplied by n/d.

As H is weighted and can be irregular, it may at first seem strange to view it as an expander. However, it may easily be shown to have the properties that define expanders: it has high edge conductance, random walks mix rapidly on H and converge to an almost uniform distribution, and it satisfies the expander mixing property (see [2] or [16, 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 nonuniform distribution, and the star does not satisfy the expander mixing property. For the convenience of the reader, we include below a proof demonstrating that H has the expander mixing property.

LEMMA 4.1. *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 that

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

Thus,

$$\begin{aligned} x^T L_H y &= \left(1 + \frac{\epsilon}{2}\right) x^T L_G y + x^T M y \\ &= -\left(1 + \frac{\epsilon}{2}\right) |S|\,|T| + x^T M y. \end{aligned}$$

The lemma now follows by observing that

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

Using the proof of the lower bound on the spectral gap of Alon and Boppana (see [29]), one can show that a d-regular unweighted graph cannot κ-approximate a complete graph for κ asymptotically better than

$$\frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}},$$

which is the bound achieved by Ramanujan graphs [24, 26]. We conjecture that this bound also holds for weighted graphs of average degree d. Presently, we prove the following weaker result for such graphs.

PROPOSITION 4.2. *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 that 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 that 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

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

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

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:

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

and

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

The ratio in question is thus

$$\begin{aligned} \frac{y^T L_H y}{x^T L_H x} &= \frac{\sum_i w_i + \sum_i (\delta_i + w_i)/d + 2 \sum_i w_i/\sqrt{d}}{\sum_i w_i + \sum_i (\delta_i + w_i)/d - 2 \sum_i w_i/\sqrt{d}} \\ &= \frac{1 + \frac{1}{\sqrt{d}} \frac{2 \sum_i w_i}{\sum_i w_i + \sum_i (\delta_i + w_i)/d}}{1 - \frac{1}{\sqrt{d}} \frac{2 \sum_i w_i}{\sum_i w_i + \sum_i (\delta_i + w_i)/d}}. \end{aligned}$$

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,

$$\frac{y^T L_H y}{x^T L_H x} \geq \frac{1+\frac{1}{\sqrt{d}}\frac{2}{1+\kappa}}{1-\frac{1}{\sqrt{d}}\frac{2}{1+\kappa}}. \tag{6}$$

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 $n \to \infty$,

$$\frac{\|x^*\|^2}{\|y^*\|^2} = 1 - O\left(\frac{\sqrt{d}}{n}\right). \tag{7}$$

Combining (6) and (7), 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 left-hand side 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. □

5. Conclusion. The most pressing question raised by our work is whether sparsifiers of a quality similar to those we construct can be computed in nearly linear time. While we have presented a polynomial time algorithm, it is too slow to use on large graphs. It is well understood how to quickly approximate complete graphs: random d-regular graphs are almost as good as Ramanujan graphs [13]. However, we are unaware of an analogous way of approximating arbitrary graphs.

In the original version of this paper, we observed that Theorem 3.1 appeared to be very similar to a conjecture of Weaver [39] that was known to imply a positive solution to the Kadison–Singer problem. The differences are that Weaver's conjecture requires all of the nonzero scalars s_i to be identical, but it adds as a condition an upper bound on the norms of the vectors $\mathbf{v}_i$. The barrier function technique introduced in section 3.2 of this paper played an important role in the recent proof of Weaver's conjecture and the solution of the Kadison–Singer problem [25].

REFERENCES

[1] C. A. Akemann and J. Anderson, *Lyapunov theorems for operator algebras*, Mem. Amer. Math. Soc., 94 (458) (1991).

[2] N. Alon and F. Chung, *Explicit construction of linear sized tolerant networks*, Discrete Math., 72 (1988), pp. 15–19.

[3] A. Barvinok, *Thrifty approximations of convex bodies by polytopes*, Int. Math. Res. Not. (2013); DOI: 10.1093/imrn/rnt078.

[4] A. A. Benczúr and D. R. Karger, *Approximating s-t minimum cuts in $O(n^2)$ time*, in Proceedings of the 28th Annual Symposium on the Theory of Computing (STOC '96), 1996, pp. 47–55.

[5] J. Bourgain and L. Tzafriri, *On a problem of Kadison and Singer*, J. Reine Angew. Math., 420 (1991), pp. 1–43.

[6] C. Boutsidis, P. Drineas, and M. Magdon-Ismail, *Near optimal column-based matrix reconstruction*, in Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE, 2011, pp. 305–314.

[7] M. K. de Carli Silva, N. J. A. Harvey, and C. M. Sato, *Sparse Sums of Positive Semidefinite Matrices*, preprint, CoRR, abs/1107.0088, 2011; available online from http://arxiv.org/abs/1107.0088.

[8] P. G. Casazza and J. C. Tremain, *The Kadison-Singer problem in mathematics and engineering*, Proc. Natl. Acad. Sci. USA, 103 (2006), pp. 2032–2039.

[9] P. Chew, *There is a planar graph almost as good as the complete graph*, in Proceedings of the 2nd Annual Symposium on Computational Geometry (SoCG '86), 1986, pp. 169–177.

[10] K. R. Davidson and S. J. Szarek, *Local operator theory, random matrices and Banach spaces*, in Handbook of the Geometry of Banach Spaces, Vol. I, North-Holland, Amsterdam, 2001, pp. 317–366.

[11] H. Dette and W. J. Studden, *Some new asymptotic properties for the zeros of Jacobi, Laguerre, and Hermite polynomials*, Constr. Approx., 11 (1995), pp. 227–238.

[12] C. Fefferman, A. Israel, and G. Luli, *Sobolev extension by linear operators*, J. Amer. Math. Soc., 27 (2014), pp. 69–145.

[13] J. Friedman, *Relative expanders or weakly relatively Ramanujan graphs.* Duke Math. J., 118 (2003), pp. 19–35.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.

[15] N. Goyal, L. Rademacher, and S. Vempala, *Expanders via random spanning trees*, in Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '09), 2009, pp. 576–585.

[16] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc., 43 (2006), pp. 439–561.

[17] M. Kapralov and R. Panigrahy, *Spectral sparsification via random spanners*, in Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 2012, pp. 393–398.

[18] J. Kelner and A. Levin, *Spectral Sparsification in the semi-streaming setting*, Theory Comput. Syst., 53 (2013), pp. 243–262.

[19] R. Khandekar, S. Rao, and U. Vazirani, *Graph partitioning using single commodity flows*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC '06), 2006, pp. 385–390.

[20] A. Kolla, Y. Makarychev, A. Saberi, and S.-H. Teng, *Subgraph sparsification and nearly optimal ultrasparsifiers*, in Proceedings of the 42nd Annual Symposium on the Theory of Computing (STOC '10), 2010, pp. 57–66.

[21] I. Koutis, A. Levin, and R. Peng, *Improved spectral sparsification and numerical algorithms for SDD Matrices*, in Proceedings of 29th Symposium on Theoretical Aspects of Computer Science, 2012, pp. 266–277.

[22] I. Koutis, G. L. Miller, and R. Peng, *Approaching optimality for solving SDD linear systems*, in Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science, 2010, pp. 235–244.

[23] I. Koutis, G. L. Miller, and R. Peng, *A nearly-m log n time solver for SDD linear systems*, in Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science, 2011, pp. 590-598.

[24] A. Lubotzky, R. Phillips, and P. Sarnak, *Ramanujan graphs*, Combinatorica, 8 (1988), pp. 261–277.

[25] A. Marcus, D. A. Spielman, and N. Srivastava, *Interlacing Families* II: *Mixed Characteristic Polynomials and the Kadison-Singer Problem*, preprint, arXiv:1306.3969, 2013.

[26] G. A. Margulis, *Explicit group theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators*, Probl. Inf. Transm., 24 (1988), pp. 39–46.
[27] A. Naor, *Sparse quadratic forms and their geometric applications (after Batson, Spielman and Srivastava)*, Ast'erisque, 348 (2012), pp. 189–217.
[28] I. Newman and Y. Rabinovich, *Finite Volume Spaces and Sparsification*, preprint, arXiv: 1002.3541, 2010.
[29] A. Nilli, *On the second eigenvalue of a graph*, Discrete Math., 91 (1991), pp. 207–210.
[30] D. Peleg and J. D. Ullman, *An optimal synchronizer for the hypercube*, SIAM J. Comput., 18 (1989), pp. 740–747.
[31] M. Rudelson, *Random vectors in the isotropic position*, J. Funct. Anal., 163 (1999), pp. 60–72.
[32] G. Schechtman, *Tight embedding of subspaces of l_p in ℓ_p^n for even p*, Proc. Amer. Math. Soc., 139 (2011), pp. 4419–4421.
[33] D. A. Spielman and N. Srivastava, *An elementary proof of the restricted invertibility theorem*, Israel J. Math., 190 (2012), pp. 83–91.
[34] D. A. Spielman and N. Srivastava, *Graph sparsification by effective resistances*, in Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC '08), 2008, pp. 563–568.
[35] D. A. Spielman and S.-H. Teng, *Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC '04), 2004, pp. 81–90.
[36] D. A. Spielman and S.-H. Teng, *Nearly-Linear Time Algorithms for Preconditioning and Solving Symmetric, Diagonally Dominant Linear Systems*, preprint, CoRR, abs/cs/0607105, 2008; available online from http://www.arxiv.org/abs/cs.NA/0607105.
[37] D. A. Spielman and S.-H. Teng, *Spectral sparsification of graphs*, SIAM J. Comput., 40 (2011), pp. 981–1025.
[38] N. Srivastava and R. Vershynin, *Covariance estimation for distributions with $2+\epsilon$ moments*, Ann. Probab., 41 (2013), pp. 3081–3111.
[39] N. Weaver, *The Kadison-Singer problem in discrepancy theory*, Discrete Math., 278 (2004), pp. 227–239.