Approximations for the Isoperimetric and Spectral Profile of Graphs and Related Parameters

Prasad Raghavendra
MSR New England
Cambridge, MA

David Steurer*
Princeton University
Princeton, NJ

Prasad Tetali
Georgia Tech
Atlanta, GA

ABSTRACT

The spectral profile of a graph is a natural generalization of the classical notion of its Rayleigh quotient. Roughly speaking, given a graph G, for each $0 < \delta < 1$, the spectral profile $\Lambda_G(\delta)$ minimizes the Rayleigh quotient (from the variational characterization) of the spectral gap of the Laplacian matrix of G over vectors with support at most δ over a suitable probability measure. Formally, the spectral profile Λ_G of a graph G is a function $\Lambda_G : [0, 1/2] \to \mathbb{R}$ defined as:

$$\Lambda_G(\delta) \stackrel{\text{def}}{=} \min_{\substack{x \in \mathbb{R}^V \\ d(\operatorname{supp}(x)) \leqslant \delta}} \frac{\sum g_{ij}(x_i - x_j)^2}{\sum_i d_i x_i^2} .$$

where g_{ij} is the weight of the edge (i, j) in the graph, d_i is the degree of vertex i, and $d(\operatorname{supp}(x))$ is the fraction of edges incident on vertices within the support of vector x.

While the notion of the spectral profile has numerous applications in Markov chain, it is also is closely tied to its isoperimetric profile of a graph. Specifically, the spectral profile is a relaxation for the problem of approximating edge expansion of small sets in graphs.

In this work, we obtain an efficient algorithm that yields a $\log(1/\delta)$-factor approximation for the value of $\Lambda_G(\delta)$. By virtue of its connection to edge-expansion, we also obtain an algorithm for the problem of approximating edge expansion of small linear sized sets in a graph. This problem was recently shown to be intimately connected to the Unique Games Conjecture in [18].

Finally, we extend the techniques to obtain approximation algorithms with similar guarantees for restricted eigenvalue problems on diagonally dominant matrices.

Categories and Subject Descriptors

F.2.2 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity—*Nonnumerical Algorithms and Problems*

*Supported by NSF grants 0830673, 0832797, 528414. Part of this work done while at Microsoft Research New England.

General Terms

Algorithms, Theory

Keywords

Spectral Profile, Graph Expansion, Semidefinite Programming, Approximation Algorithm, Small-Set Expansion, Sparse Principal Component Analysis

1. INTRODUCTION

Motivated by the need for improvements over the elementary bounds on total variation mixing times of a Markov chain using spectral gap and conductance, Lovász and Kannan [15] introduced the notion of an average conductance, which takes into account the gain from large expansion small sets (of states of a Markov chain). By deriving bounds on the stronger L^2-mixing time, Morris and Peres [17] further strengthened the above result, using the notion of *evolving sets*. A corresponding functional analog of the average conductance was then derived in the form of the spectral profile by Goel-Montegro-Tetali [14], building on and further simplifying earlier works in the continuous setting of heat kernels on manifolds [5, 10]. As explained in detail in [16], besides providing tighter estimates on mixing times for various walks (by avoiding the penalty of a slow start), the approach to mixing time in L^2 using the spectral profile facilitated easier derivation of estimates on mixing using other functional approaches – notably the logarithmic Sobolev and Nash inequalities.

Generalizing the standard Cheeger-type inequality relating the spectral gap and conductance (or sparsest cut in the context of a graph), [14] also showed a Cheeger-type inequality relating the spectral profile to the average conductance (or isoperimetric profile). Since one is often interested in knowing when *small sets* expand (as this is sufficient for many applications of expanders), the question of computing or approximating the spectral profile becomes relevant and interesting – with the hope of obtaining a certificate for such small-set expansion. In this work we consider a very natural semidefinite relaxation of the profile and obtain a logarithmic factor approximation. A precise statement of our results appears below.

Inspired by recent applications in optimization over sparse subspaces, we consider a natural extension of the notion of spectral profile to the class of diagonally dominant matrices. Recall that a matrix (over the reals) is diagonally dominant if in every row, the diagonal entry is larger or equal (in absolute value) to the sum of the absolute values of the off-diagonal entries. The Laplacian matrix of a graph (being the diagonal matrix of degrees minus the adjacency matrix), with a row and a corresponding column removed, is easily seen to be a special case of such matrices. Our second main contribution is in formulating the spectral profile problem for the diagonally dominant matrices and providing a similar logarithmic factor approximation. This turns out to be a rather nontrivial extension of our first result for graphs. Once again, the precise statement of our second main theorem is provided below.

In the following we briefly mention some recent related work. In the currently active topic of sparse principal component analysis [11, 12, 13], a general problem is to decompose a covariance matrix as a combination of projections on orthogonal "sparse subspaces". A related problem to this turns out to be to maximize a quadratic form over sparse vectors. For this problem, d'Aspremont et al. [13] consider a semidefinite relaxation very similar to the one considered in our work. We show the first approximation algorithm for this problem. If the covariance matrix is diagonally dominant, then we can find sparse vectors that are almost maximally correlated with this covariance matrix (see Theorem 1.6 for a more precise statement).

1.1 Results

Let $G = (g_{ij})$ be a weighted graph with vertex set $V = \{1, \ldots, n\}$. Let $d = (d_1, \ldots, d_n)$ be the *degree vector* of G, that is, $d_i := \sum_j g_{ij}$. We assume that the weights of G are scaled such that $\sum_i d_i = 1$. Hence, we can think of d as a probability measure on V and we denote $d(S) := \sum_{i \in S} d_i$ for a subset $S \subseteq V$. For a vector $x \in \mathbb{R}^V$, let $\operatorname{supp}(x) := \{i \in V \mid x_i \neq 0\}$ denote its *support*. We define the *spectral profile* of G as

$$\Lambda_G(\delta) \stackrel{\text{def}}{=} \min_{\substack{x \in \mathbb{R}^V \\ d(\operatorname{supp}(x)) \leqslant \delta}} \frac{\sum_{ij} g_{ij}(x_i - x_j)^2}{\sum_i d_i x_i^2}. \tag{1.1}$$

We consider a natural semidefinite programming relaxation $\tilde{\Lambda}_G$ for the spectral profile of G (see §2.1, (2.3)). Our main result is the following relation of the spectral profile and its SDP relaxation. (The theorem follows by combining Theorem 3.1 and Theorem 3.2 in Section 3.)

THEOREM 1.1. *For every graph G and all $\varepsilon, \delta > 0$,*

$$\tilde{\Lambda}_G(\delta) \leqslant \Lambda_G(\delta) \leqslant \tilde{\Lambda}_G\left(\delta/1+\varepsilon\right) \cdot O\left(\log\left(1/\delta\right)/\varepsilon^3\right).$$

In the theorem above, we can think of ε as an absolute constant, say $\varepsilon = 0.01$, and δ as tending to zero.

Standard arguments also show that the rounding can be implemented efficiently, yielding approximation algorithms for the spectral profile of graphs. In particular, given a graph G and a parameter $\delta > 0$, we can compute a vertex set of volume $O(\delta)$ such that the corresponding principal submatrix of $\mathcal{L}_G$ has smallest eigenvalue $\lambda \leqslant \Lambda_G(\delta) \cdot O(\log(1/\delta))$. Given such a submatrix, we can efficiently find a vertex set S of volume $O(\delta)$ with conductance $\Phi(S) \leqslant O(\sqrt{\lambda})$ [14, 8]. Recall that the conductance $\Phi(S)$ of a vertex set S is the weight of the edge leaving the set divided by the weight of all edges incident to the set.

The combination of these results leads to following approximation algorithm for SMALL-SET EXPANSION (as considered in [18]).

THEOREM 1.2. *There exists a polynomial time algorithm that given a graph G and a parameter $\delta > 0$, finds a set $S \subseteq [n]$ with volume $d(S) = O(\delta)$ and conductance*

$$\Phi(S) \leqslant O\big(\tilde{\Lambda}(\delta)\log(1/\delta)\big)^{1/2}.$$

To the best of our knowledge, there are no previous approximation results for SMALL-SET EXPANSION mentioned in the literature (at least for the regime where the optimal conductance is small but constant).

We observe that the approximation algorithm for SPARSEST CUT based on Cheeger's inequality [1, 2] implies quite easily a somewhat worse approximation for SMALL-SET EXPANSION. The rough idea is to apply this algorithm SPARSEST CUT recursively until the graph is partitioned into sets of volume not much more than δ. We can then argue that one of these sets has conductance $O(\log(1/\delta)\sqrt{\varepsilon})$ if the optimal conductance of a set of volume δ was ε. Theorem 1.2 improves this approximation guarantee by a factor $\sqrt{\log(1/\delta)}$.

We note that one can prove Theorem 1.2 also directly, without going through the approximation for the spectral profile. The idea is to combine the techniques of [6, 7] with the techniques in this paper.

We also note that the approximations of Theorem 1.1 and Theorem 1.2 are optimal (up to constant factors) with respect to the relaxation we are using. The Gaussian noise graph in $\mathbb{R}^n$ (Ornstein–Uhlenbeck operator) is an integrality gap instance with matching parameters.

Finally, we remark that the approximation algorithms for SMALL-SET EXPANSION and the spectral profile can be implemented with quasi-linear running time if $\tilde{\Lambda}(\delta)$ is lower-bounded by a constant. (We measure the instance size by the number of edges in the graph.) In this case, we can compute near-optimal solutions for the SDP relaxation $\tilde{\Lambda}(\delta)$ in quasi-linear time using the framework of Arora and Kale [3] (also [19]).

Diagonally Dominant Matrices.

Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ be a symmetric diagonally dominant[1] matrix. Let μ be some probability measure on the coordinate set $[n]$. Let $R_A(x) = \sum_{ij} a_{ij}x_i x_j / \sum_i \mu_i x_i^2$ be the Rayleigh quotient of A at x. We define the spectral profile of A as $\Lambda_A(\delta) = \min_{x \in \mathbb{R}^n, \mu(\operatorname{supp}(x)) \leqslant \delta} R_A(x)$. We consider a natural semidefinite relaxation $\tilde{\Lambda}_A(\delta)$ of the spectral profile. We can generalize Theorem 1.1 in the following way. (The theorem follows by combining Theorem 4.1 and Theorem 4.3 in Section 4.)

THEOREM 1.3. *There exists an absolute constant $c > 0$ such that for every matrix A as above and all $\delta > 0$,*

$$\tilde{\Lambda}_A(\delta) \leqslant \Lambda_A(\delta) \leqslant \tilde{\Lambda}_A(c\delta) \cdot O\big(\log(1/\delta)\big).$$

The rounding that achieves the above bounds is quite different from our construction for Theorem 1.1.

A relatively straight-forward modification of the construction for Theorem 1.1 achieves almost the above bound, but with an additional additive error of $\tilde{\Lambda}_A(c\delta)\log(1/\tilde{\Lambda}_A(c\delta))$. We present the significantly simpler analysis of this slightly suboptimal construction in Section 4.2.

Again the rounding can be implemented efficiently.

[1] diagonally dominant means $a_{ii} \geqslant \sum_{j \neq i} |a_{ij}|$ for all $i \in [n]$

THEOREM 1.4. *There exists a polynomial time algorithm that given a matrix A as above and a parameter $\delta > 0$, finds an $O(\delta)$-sparse vector x such that*

$$R_A(x) \leqslant \tilde{\Lambda}_A\,(c\delta) \cdot O\big(\log(1/\delta)\big)\,.$$

Trevisan [20] showed that Cheeger's inequality also applies in this setting.

THEOREM 1.5. *There exists a polynomial time algorithm that given a matrix A as above and a parameter $\delta > 0$, finds an $O(\delta)$-sparse vector $x \in \{-1, 0, 1\}^n$ such that*

$$R_A(x) \leqslant O\Big(\tilde{\Lambda}_A\,(c\delta) \cdot \log(1/\delta)\Big)^{1/2}\,.$$

As before, the framework of Arora and Kale [3] allows us to implement these approximation algorithms in quasi-linear if $\tilde{\Lambda}_A(c\delta)$ is bounded from below by a constant.

D'Aspremont et al. [13] consider the problem of maximizing $R_A(x)$ over sparse vectors x. If the measure μ is proportional to the diagonal of A (which is a natural assumption), this problem reduces to the problem of computing the spectral profile of the matrix $A' = 2\operatorname{diag}(A) - A$. (In words, A' has the same diagonal as A, but all off-diagonal entries are negated.) Notice that A' is diagonally dominant if and only if A is so. Hence, we can apply Theorem 1.4 to A'. In this way, we get the following approximation algorithm for the problem of maximizing $R_A(x)$ over sparse vectors x.

THEOREM 1.6. *Let A be a matrix as above and let $\delta > 0$. Suppose the measure μ is proportional to the diagonal of A. If there exists a δ-sparse vector x such that $R_A(x) \geqslant (2-\eta)\operatorname{Tr} A$, then we can find in polynomial time a $O(\delta)$-sparse vector y such that*

$$R_A(y) \geqslant \big(2 - O\big(\eta \cdot \log(1/\delta)\big)\big)\operatorname{Tr} A\,.$$

(In our normalization, 2 is the maximum possible value that $R_A(x)$ can achieve.)

1.2 Proof Overview and Techniques

SDP Relaxation of sparsity constraints.

Let G be a graph on vertex set $\{1, \ldots, n\}$. For the sake of simplicity, let us assume that G is unweighted and regular. Our goal is to find a vector x with about δn non-zero coordinates that minimizes the following Rayleigh quotient (from the variational characterization of the spectral gap)

$$R_G(x) \stackrel{\text{def}}{=} \frac{\sum_{i\sim j}(x_i - x_j)^2}{\sum_i x_i^2}\,.$$

We note that without loss of generality, it is enough to consider non-negative vectors (the Rayleigh quotient can only decrease if we replace every coordinate by its absolute value). Ignoring the sparsity constraint, the minimization of $R_G(x)$ is equivalent to a semidefinite program, where we minimize the function $\sum_{i\sim j}\|v_i - v_j\|^2$ over all collections of vectors $\{v_i\}$ such that $\sum_i \|v_i\|^2 = 1$. The challenge is to deal with the sparsity constraint. Since every δ-sparse vector x satisfies $\|x\|_1^2 \leqslant \delta n\|x\|_2^2$, we can relax the sparsity constraint to $\|\sum_i v_i\|^2 \leqslant \delta n \sum_i \|v_i\|^2$ in the SDP. In addition to this constraint, we also include non-negativity constraints for the inner products, $\langle v_i, v_j\rangle \geqslant 0$ for all $i, j \in [n]$. We note that unlike for other SDP relaxations (e.g. for UNIQUE GAMES), these non-negativity constraints are crucial. (Without these constraints the SDP value for two disjoint cliques would be 0 for all $\delta > 0$, even though the spectral profile is close to 1 for small enough $\delta > 0$.)

Before describing the rounding, we mention a natural generalization: For a symmetric positive semidefinite matrix $A = (a_{ij})$, let $R_A(x) = \langle x, Ax\rangle / \langle x, x\rangle$ be its Rayleigh quotient. Our goal is to minimize this quotient over all δ-sparse vectors x. The objective of our SDP relaxation for this problem is $\sum_{ij} a_{ij}\langle v_i, v_j\rangle / \sum_i \|v_i\|^2$. In contrast to the previous case, we can no longer assume that the coordinates of the minimizer x are non-negative. Hence, we cannot include non-negativity constraints for the inner products. Instead, we add the constraint $\sum_{ij}|\langle v_i, v_j\rangle| \leqslant \delta n \sum_i \|v_i\|^2$. The resulting SDP relaxation has been suggested in [13] (for the related problem of maximizing $R_A(x)$ over sparse vectors x).

Two-phase rounding.

Our rounding proceeds in two phases: In the first phase, we start with an optimal solution to the SDP relaxation and produce a new SDP solution that satisfies a stronger relaxation of the sparsity constraint. In this phase, our objective value increases by at most a factor of order $\log(1/\delta)$. In the second phase, we extract a sparse vector from the SDP solution that satisfies the stronger sparsity constraint. In this phase, we only increase the objective value by a constant factor.

The advantage of this two-phase rounding is that it allows us to separate local and non-local arguments. In the first phase, the transformation of the SDP solution is quite drastic (we lose a $\log(1/\delta)$ factor in the objective). However, the transformation in this phase is completely local. In fact, every vector gets mapped to a new vector in a way that is oblivious to all other vectors in the SDP solution. On the other hand, the rounding in the second phase is inherently non-local, because it has to ensure that the resulting vector x is sparse. In particular, the rounding for the ith coordinate of x cannot be oblivious to all other coordinates. What allows us to analyze the non-local rounding in this phase, is the stronger sparsity constraint that we established in the first phase.

The stronger sparsity constraint that is established in the first phase has the form

$$\big\|\textstyle\sum_i |v_i|\big\|^2 \leqslant \delta n \displaystyle\sum_i \|v_i\|^2\,, \tag{1.2}$$

where $|v_i|$ is the vector v_i with each coordinate replaced by its absolute value. In Section 3.1, we show that vectors satisfying this constraints can be rounded to a sparse vector x such that $R_G(x)$ is not much larger than the SDP value of the vectors $v_1, \ldots, v_n$. The main ingredient is the following lemma (see Lemma 2.5 and its consequences Lemma 3.6 and Lemma 4.7): Suppose x satisfies $\|x\|_1^2 \leqslant \delta n\|x\|_2^2$. Then, we can choose a parameter $t > 0$ such that the vector y defined as $y_i = \mathsf{soft\text{-}chop}_t(x_i)$ is 3δ-sparse (say) and at the same time $\|y\|_2^2 \geqslant \frac{1}{3}\|x\|_2^2$ (a good fraction of the mass is preserved). Here, $\mathsf{soft\text{-}chop}_t$ is the soft-thresholding function, defined as $\mathsf{soft\text{-}chop}_t(x) = x - t$ if $x > t$, $\mathsf{soft\text{-}chop}_t(x) = 0$ if $|x| \leqslant t$, and $\mathsf{soft\text{-}chop}_t(x) = x + t$ if $x < -t$.

This lemma allows us to show that in order to approximate $\Lambda_G(\delta)$, it is enough to find a vector x such that $R_G(x)$ is small and $\|x\|_1^2 \leqslant \delta n\|x\|_2^2$. More concretely, it is easy to show that $R_G(y) \leqslant 3R_G(x)$ where y is the 3δ-sparse vector obtained from x by soft-thresholding as above. (This argument works not only for Laplacian matrices but also for symmetric diagonally dominant matrices, see Section 4.1.)

In the first phase of the rounding, our task is to transform an arbitrary SDP solution to a solution that satisfies the stronger sparsity constraint (1.2). For Laplacian matrices, we use the same transformation as Barak et al. [4]. This transformation has the property that it maps arbitrary vectors to vectors with only non-negative coordinates and it preserves distances in an approximate sense (see Section 3.2). For diagonally dominant matrices, we require that the transformation is symmetric (an antipodal pair of vectors gets mapped to an antipodal pair again). A straight forward adaptation of the transformation for Laplacian matrices already gives a relatively good approximation for diagonally dominant matrices (see Section 4.2). Unfortunately, this construction incurs an additional additive error term and therefore the final approximation guarantee is not multiplicative. In Section 4.3, we describe a transformation that is quite different from the construction of Barak et al. [4] and avoids the additional additive error even for diagonally dominant matrices. The disadvantage of this construction is that the analysis is technically more involved (it relies on Gaussian noise sensitivity bounds for soft threshold functions, see Section B).

2. PRELIMINARIES

2.1 Semidefinite Relaxation for the Spectral Profile of Graphs

Let $G = (g_{ij})$ be a weighted graph with vertex set $V = \{1, \dots, n\}$. Let $d = (d_1, \dots, d_n)$ be the *degree vector* of G, that is, $d_i := \sum_j g_{ij}$. We assume that the weights of G are scaled such that $\sum_i d_i = 1$. We denote $d(S) := \sum_{i \in S} d_i$ for a subset $S \subseteq V$. For a vector $x \in \mathbb{R}^V$, let $\operatorname{supp}(x) := \{i \in V \mid x_i \neq 0\}$ denote its *support*. The *spectral profile* of G is defined as

$$\Lambda_G(\delta) \stackrel{\text{def}}{=} \min_{\substack{x \in \mathbb{R}^V \\ d(\operatorname{supp}(x)) \leqslant \delta}} \frac{\sum_{ij} g_{ij}(x_i - x_j)^2}{\sum_i d_i x_i^2} . \tag{2.1}$$

Note that $\Lambda_G(\delta)$ is monotonically decreasing in δ. It is easy to show that $\Lambda_G(1/2)$ is up to constant factors equal to the usual spectral gap of G (e.g., see [14]). The spectral profile is characterized by the smallest eigenvalues of certain restrictions of the Laplacian of G (Dirichlet eigenvalues). Let $\mathcal{L}$ denote the normalized Laplacian of G (as defined e.g. in [9]). For a set $S \subseteq V$, let $\mathcal{L}_S$ be the $S \times S$ submatrix of $\mathcal{L}$. Then,

$$\Lambda_G(\delta) = \min_{\substack{S \subseteq V \\ d(S) \leqslant \delta}} \lambda_{\min}(\mathcal{L}_S) . \tag{2.2}$$

We consider the following semidefinite programming relaxation for the spectral profile,

$$\tilde{\Lambda}_G(\delta) \stackrel{\text{def}}{=} \min_{v_1, \dots, v_n \in \mathbb{R}^n} \frac{\sum_{ij} g_{ij} \|v_i - v_j\|^2}{\sum_i d_i \|v_i\|^2} , \tag{2.3}$$

where the minimum is over all vector configurations $v_1, \dots, v_n \in \mathbb{R}^n$ that satisfy

$$\left\|\textstyle\sum_i d_i v_i\right\|^2 \leqslant \delta \cdot \textstyle\sum_i d_i \|v_i\|^2 , \tag{2.4}$$

$$\langle v_i, v_j \rangle \geqslant 0 . \qquad (i, j \in V) \tag{2.5}$$

Let us show that $\tilde{\Lambda}_G(\delta)$ is indeed a relaxation of the spectral profile. Let v_0 be any unit vector in $\mathbb{R}^n$. For a vector $x \in \mathbb{R}^V$ with $S = \operatorname{supp}(x)$ and $d(S) \leqslant \delta$, define vectors $v_1, \dots, v_n \in \mathbb{R}^n$ as $v_i = |x_i| \cdot v_0$. Let $\mathbb{1}_S \in \{0,1\}^V$ denote the indicator function of the set S. By Cauchy–Schwarz,

$$\begin{aligned} \|\textstyle\sum d_i v_i\|^2 &= \Big(\sum_i d_i |x_i|\Big)^2 \\ &\leqslant \sum_i d_i \mathbb{1}_S(i) \cdot \sum_i d_i x_i^2 \leqslant \delta \cdot \sum_i d_i \|v_i\|^2 . \end{aligned}$$

Furthermore, it is easy to verify that $\langle v_i, v_j \rangle \geqslant 0$ for all $i, j \in V$ and that $\sum_{ij} g_{ij} \|v_i - v_j\|^2 \leqslant \sum_{ij} g_{ij}(x_i - x_j)^2$. It follows that $\tilde{\Lambda}_G(\delta)$ is indeed a relaxation for the spectral profile.

LEMMA 2.1. *For every graph G and all $\delta > 0$,*

$$\tilde{\Lambda}_G(\delta) \leqslant \Lambda_G(\delta) .$$

2.2 Semidefinite Relaxation for the Spectral Profile of Matrices

Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ be a symmetric positive semidefinite matrix and let μ be a probability measure on $\{1, \dots, n\}$.

We define the *spectral profile* $\Lambda_A \colon [0,1] \to \mathbb{R}$ as

$$\Lambda_A(\delta) \stackrel{\text{def}}{=} \min_{\substack{x \in \mathbb{R}^n \\ \mu(\operatorname{supp}(x)) \leqslant \delta}} \frac{\sum_{ij} a_{ij} x_i x_j}{\sum_i \mu_i x_i^2} .$$

Let $D \in \mathbb{R}^{n \times n}$ be the diagonal matrix with entries $\mu_1, \dots, \mu_n$. Note that

$$\Lambda_A(\delta) = \min_{\substack{S \subseteq [n] \\ \mu(S) \leqslant \delta}} \lambda_{\min}(D_S^{-1/2} A_S D_S^{-1/2}) ,$$

where D_S and A_S denote the principal submatrices of D and A corresponding to the coordinate set S.

We consider the following semidefinite programming relaxation for the spectral profile,

$$\tilde{\Lambda}_A(\delta) \stackrel{\text{def}}{=} \min_{v_1, \dots, v_n \in \mathbb{R}^n} \frac{\sum_{ij} a_{ij} \langle v_i, v_j \rangle}{\sum_i \mu_i \|v_i\|^2} , \tag{2.6}$$

where the minimum is over all vector configurations $v_1, \dots, v_n \in \mathbb{R}^n$ that satisfy

$$\sum_{ij} \mu_i \mu_j \left|\langle v_i, v_j \rangle\right| \leqslant \delta \cdot \sum_i \mu_i \|v_i\|^2 . \tag{2.7}$$

This semidefinite relaxation has been introduced in [13] (for the corresponding maximization problem).

Note that the constraint (2.7) is weaker than the previous constraints (2.4) and (2.5). We have to use the weaker constraint (2.7) for $\tilde{\Lambda}_A$ because we can no longer assume that the optimal vector $x \in \mathbb{R}^n$ has only nonnegative coordinates. As before, it is easy to verify that $\tilde{\Lambda}_A$ is a relaxation of Λ_A.

LEMMA 2.2. *For every symmetric matrix A and all $\delta > 0$,*

$$\tilde{\Lambda}_A(\delta) \leqslant \Lambda_A(\delta) .$$

2.3 Gaussian Distributions

Let λ^n denote the usual Lebesgue measure on $\mathbb{R}^n$. Let $N(0, \sigma^2)^n$ be the Gaussian measure on $\mathbb{R}^n$ with mean 0 and covariance $\sigma^2 I$ (each coordinate is independent Gaussian with mean 0 and standard deviation σ). Let $\phi_\sigma \colon \mathbb{R}^n \to \mathbb{R}_+$

be the density of $N(0,\sigma^2)^n$ with respect to the Lebesgue measure,

$$\phi_\sigma(x) \stackrel{\text{def}}{=} \tfrac{1}{(\sigma\sqrt{2\pi})^n} e^{-\|x\|^2/2\sigma^2} .$$

Let T_u be the translation operator, i.e.,

$$T_u f(x) \stackrel{\text{def}}{=} f(x-u) .$$

We will use the following bound on the (Hellinger) affinity of translated Gaussians (e.g. in Barak et al. [4]).

Lemma 2.3 ([4]). *Let u and v be two unit vectors in $\mathbb{R}^n$. Then*

$$\int \sqrt{T_u\phi_\sigma \cdot T_v\phi_\sigma} \;\; \mathrm{d}\lambda^n = e^{-\|u-v\|^2/8\sigma^2} .$$

Proof. Immediate from the identity

$$\sqrt{T_u\phi_\sigma \cdot T_v\phi_\sigma} = e^{-\|u-v\|^2/8\sigma^2} T_{\frac{1}{2}(u+v)}\phi_\sigma . \quad \square$$

We equip the space $\{f\colon \mathbb{R}^n \to \mathbb{R}\}$ with the inner product $\langle f,g\rangle := \int fg \, \mathrm{d}\lambda^n$ and define the corresponding norm $\|f\| := \langle f,f\rangle^{1/2}$. Consider the mapping

$$u \mapsto f_u := \|u\| \sqrt{T_{\bar u}\phi_\sigma} .$$

Here, $\bar u$ denotes the unit vector in the direction of u. In §3.2, we will use this mapping to transform arbitrary SDP solutions to non-negative SDP solutions. Notice that $\|f_u\| = \|u\|$ for every vector $u \in \mathbb{R}^n$. Furthermore, for unit vectors $u,v\in\mathbb{R}^n$ that are close, say $\|u-v\|^2 \leqslant \varepsilon$, we have $\|f_u - f_v\|^2 \leqslant O(\varepsilon/\sigma^2)$. On the other hand, for unit vectors $u,v\in\mathbb{R}^n$ that are far apart, say $\langle u,v\rangle \leqslant 1/2$, we have $\langle f_u, f_v\rangle \leqslant 2^{-\Omega(1/\sigma^2)}$. In this sense, $u\mapsto f_u$ is a distance preserving mapping from the sphere in $\mathbb{R}^n$ to the non-negative orthant of the space $\mathbb{R}^n \to \mathbb{R}$.

The following technical fact shows that in order for a mapping to preserve distances it is enough to preserve lengths and distances of unit vectors (angles).

Fact 2.4. *For any two vectors $u,v\in\mathbb{R}^n$, we have*

$$\|u-v\|^2 = (\|u\| - \|v\|)^2 + \|u\|\,\|v\| \cdot \|\bar u - \bar v\|^2 .$$

2.4 Nonnegative Random Variables

The following technical lemma shows the following intuitive fact. If a non-negative random variable X satisfies $\mathbb{E}\,X^2 > (1+\varepsilon)k(\mathbb{E}\,X)^2$, then a significant fraction of the 2-norm of X has to be on values larger than $\tau = k\,\mathbb{E}\,X$, in the sense that the 2-norm of the variable $Y = \max\{X-\tau, 0\}$ is comparable to the 2-norm of X.

Lemma 2.5. *Let X be a non-negative random variable with mean μ. For $k\geqslant 1$, let $Y = \max\{X-k\mu, 0\}$. Then, for every $\varepsilon > 0$,*

$$\mathbb{E}\,Y^2 \geqslant \varepsilon^2 \cdot \left(\mathbb{E}\,X^2 - k\mu^2/(1-\varepsilon)\right) .$$

Proof. For $u=\varepsilon^2$, let A be the event $Y^2\geqslant uX^2$ and let B denote the complementary event. In other words, B is the event $(1-\varepsilon)X \leqslant k\mu$. Let $\mathbb{1}_A$ and $\mathbb{1}_B$ be 0/1-indicator variables for these events. We can lower bound $\mathbb{E}\,Y^2$ as follows

$$\begin{aligned}\mathbb{E}\,Y^2 &\geqslant u\,\mathbb{E}\,\mathbb{1}_A X^2 \\ &= u\left(\mathbb{E}\,X^2 - \mathbb{E}\,\mathbb{1}_B X^2\right) \\ &\geqslant u\left(\mathbb{E}\,X^2 - \mathbb{E}\,X\cdot k\mu/(1-\varepsilon)\right) . \quad \square\end{aligned}$$

3. SPECTRAL PROFILE OF GRAPHS

We prove the following two theorems in this section (one for each phase of our rounding). Putting them together yields Theorem 1.1.

Let $G=(g_{ij})$ be a graph with vertex set $V=\{1,\ldots,n\}$ and degrees $d_1,\ldots,d_n$ such that $\sum_i d_i = 1$.

Theorem 3.1. *For all $\delta > 0$ and $t\geqslant 1$, there exist non-negative functions $f_1,\ldots,f_n\colon \mathbb{R}^n\to\mathbb{R}_+$ such that*

$$\frac{\sum_{ij} g_{ij}\|f_i - f_j\|^2}{\sum_i d_i\|f_i\|^2} \leqslant t\cdot \tilde\Lambda_G(\delta)$$

and

$$\|\textstyle\sum_i d_i f_i\|^2 \leqslant (\delta + e^{-t}) \sum_i d_i \|f_i\|^2 .$$

Theorem 3.2. *Let $f_1,\ldots,f_n\colon\mathbb{R}^n\to\mathbb{R}_+$ be non-negative functions that satisfy*

$$\left\|\textstyle\sum_i d_i f_i\right\|^2 \leqslant \delta \sum_i d_i \left\|f_i\right\|^2 . \tag{3.1}$$

Then for every $\varepsilon > 0$,

$$\Lambda_G(\delta/1-2\varepsilon) \leqslant \frac{\sum_{ij} g_{ij}\|f_i - f_j\|^2}{\varepsilon^3\sum_i d_i\|f_i\|^2} .$$

3.1 From Non-negative SDP Solutions to Sparse Vectors (Theorem 3.2)

Let $f_1,\ldots,f_n\colon\mathbb{R}^n\to\mathbb{R}_+$ be non-negative functions that satisfy the constraint (3.1). For $\delta' := \delta/(1-2\varepsilon)$, let $t\colon \mathbb{R}^n\to\mathbb{R}_+$ be the function

$$t \stackrel{\text{def}}{=} \tfrac{1}{\delta'} \sum_i d_i f_i .$$

Define non-negative functions $f'_1,\ldots,f'_n\colon\mathbb{R}^n\to\mathbb{R}_+$ as

$$f'_i \stackrel{\text{def}}{=} \max\{f_i - t, 0\} .$$

Let $F'\colon\mathbb{R}^n\to\mathbb{R}^n_+$ denote $F' := (f'_1,\ldots,f'_n)$.

For the next lemma we crucially rely on the fact that the functions $f_1,\ldots,f_n$ are non-negative (in fact it is the only part of the proof where the non-negativity is used).

Lemma 3.3. *For every $x\in\mathbb{R}^n$,*

$$d(\operatorname{supp}(F'(x))) \leqslant \delta' .$$

Proof. The lemma is a consequence of Markov's inequality. Imagine that we choose i at random with probability d_i. Then $d(\operatorname{supp}(F'(x)))$ is the probability that $f_i(x)$ is larger than $t(x)$. On the other hand, $\sum_i d_i f_i(x)$ is the expected value of $f_i(x)$ if i is chosen from this distribution. (Here it is important that $f_i(x)$ is non-negative.) Thus,

$$t(x)\cdot d(\operatorname{supp}(F'(x))) \leqslant \sum_i d_i f_i(x) .$$

By the choice of t, it follows that $d(\operatorname{supp}(F'(x)))\leqslant\delta'$. $\square$

Lemma 3.4.

$$\Lambda_G(\delta') \leqslant \frac{\sum_{ij} g_{ij}\|f'_i - f'_j\|^2}{\sum_i d_i\|f'_i\|^2} .$$

Proof. For every $x\in\mathbb{R}^n$, the support of the vector $F'(x)$ has weight at most δ' (Lemma 3.3) and therefore,

$$\Lambda_G(\delta') \leqslant \frac{\sum_{ij} g_{ij}(f'_i(x) - f'_j(x))^2}{\sum_i d_i f'_i(x)^2} .$$

Furthermore, by an averaging argument, there exists an $x \in \mathbb{R}^n$ such that

$$\frac{\sum_{ij} g_{ij}(f'_i(x) - f'_j(x))^2}{\sum_i d_i f'_i(x)^2} \leqslant \frac{\sum_{ij} g_{ij}\|f'_i - f'_j\|^2}{\sum_i d_i \|f'_i\|^2}\,.$$

The combination of the previous two equations implies the lemma. $\square$

LEMMA 3.5.

$$\sum_{ij} g_{ij}\|f'_i - f'_j\|^2 \leqslant \sum_{ij} g_{ij}\|f_i - f_j\|^2\,.$$

PROOF. The truncation operation is contractive, i.e., for any two functions f and g, we have $\|\max\{f,0\} - \max\{g,0\}\| \leqslant \|f - g\|$. This fact implies the lemma. $\square$

LEMMA 3.6.

$$\sum_i d_i\|f'_i\|^2 \geqslant \varepsilon^3 \sum_i d_i\|f_i\|^2\,.$$

PROOF. By Lemma 2.5, for every $x \in \mathbb{R}^n$,

$$\sum\nolimits_i d_i f'_i(x)^2 \geqslant \varepsilon^2\Big(\sum\nolimits_i d_i f_i(x)^2 - \Big(\sum\nolimits_i d_i f_i(x)\Big)^2 \tfrac{1}{\delta'(1-\varepsilon)}\Big)\,.$$

Hence,

$$\begin{aligned}\sum\nolimits_i d_i\|f'_i\|^2 &\geqslant \varepsilon^2 \cdot \Big(\sum\nolimits_i d_i\|f_i\|^2 - \Big\|\sum\nolimits_i d_i f_i\Big\|^2 \tfrac{1}{\delta'(1-\varepsilon)}\Big)\\ &\geqslant \varepsilon^2 \cdot \Big(1 - \tfrac{\delta}{\delta'(1-\varepsilon)}\Big)\sum\nolimits_i d_i\|f_i\|^2 \qquad \text{(by (3.1))}\\ &= \varepsilon^2 \cdot \tfrac{\varepsilon}{1-\varepsilon}\sum\nolimits_i d_i\|f_i\|^2 \qquad \text{(choice of } \delta')\,. \quad \square\end{aligned}$$

Putting together the previous three lemmas (Lemma 3.4, Lemma 3.5 and Lemma 3.6) yields Theorem 3.2.

3.2 From Arbitrary SDP Solutions to Non-negative SDP Solutions (Theorem 3.1)

Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be an optimal solution for the SDP relaxation $\widetilde{\Lambda}_G(\delta)$ (see (2.3)–(2.5)). The vectors satisfy

$$\Big\|\sum\nolimits_i d_i v_i\Big\|^2 \leqslant \delta \sum\nolimits_i d_i\|v_i\|^2\,, \tag{3.2}$$

$$\langle v_i, v_j\rangle \geqslant 0\,. \qquad (i,j \in V) \tag{3.3}$$

We define non-negative functions $f_1, \ldots, f_n\colon \mathbb{R}^n \to \mathbb{R}_+$ as

$$f_i := \|v_i\| \cdot \sqrt{T_{\bar{v}_i}\phi_\sigma}\,.$$

Here, $T_{\bar{v}_i}\phi_\sigma$ is the density function of the Gaussian distribution with mean $\bar{v}_i$ and standard deviation $\sigma \leqslant 1/2$ in each coordinate (see §2.3 for the formal definitions).

The following lemma shows that up to factor $O(1/\sigma^2)$ the functions $f_1, \ldots, f_n$ have the same objective value as the SDP solution $v_1, \ldots, v_n$ (which we assumed to be optimal).

LEMMA 3.7.

$$\frac{\sum_{ij} g_{ij}\|f_i - f_j\|^2}{\sum_i d_i\|f_i\|^2} \leqslant \frac{\sum_{ij} g_{ij}\|v_i - v_j\|^2}{4\sigma^2\sum_i d_i\|v_i\|^2}\,.$$

PROOF. Since $\|v_i\| = \|f_i\|$, it is enough to show that we have $\|f_i - f_j\|^2 \leqslant \|v_i - v_j\|^2/4\sigma^2$ for all $i,j \in V$. From Lemma 2.3 and the fact $1 - x \leqslant e^{-x}$, it follows that

$$\|\bar{f}_i - \bar{f}_j\|^2 = 2 - 2e^{-\|\bar{v}_i - \bar{v}_j\|^2/8\sigma^2} \leqslant \|\bar{v}_i - \bar{v}_j\|^2/4\sigma^2$$

Together with Fact 2.4, we get as desired

$$\begin{aligned}\|f_i - f_j\|^2 &= (\|v_i\| - \|v_j\|)^2 + \|v_i\|\,\|v_j\| \cdot \|\bar{f}_i - \bar{f}_j\|^2\\ &\leqslant (\|v_i\| - \|v_j\|)^2 + \|v_i\|\,\|v_j\| \cdot \|\bar{v}_i - \bar{v}_j\|^2/4\sigma^2\\ &\leqslant \|v_i - v_j\|^2/4\sigma^2.\end{aligned}$$

The last inequality uses the assumption $\sigma \leqslant 1/2$ and again Fact 2.4. $\square$

In the next lemma, we crucially use that the vectors $v_1, \ldots, v_n$ satisfy the constraints (3.2) and (3.3). In fact, it is the only part of the proof where these constraints are important.

LEMMA 3.8.

$$\Big\|\sum\nolimits_i d_i f_i\Big\|^2 \leqslant (e^{-1/4\sigma^2} + \delta)\sum_i d_i\,\|v_i\|^2\,.$$

PROOF. Using Lemma 2.3 and the fact $e^{cx} \leqslant 1 - (1 - e^c)x$ for $x \in [0,1]$, we can bound the inner products of the normalized functions by

$$\begin{aligned}\langle \bar{f}_i, \bar{f}_j\rangle &= e^{-(1-\langle \bar{v}_i, \bar{v}_j\rangle)/4\sigma^2}\\ &\leqslant 1 - (1 - e^{-1/4\sigma^2})(1 - \langle \bar{v}_i, \bar{v}_j\rangle) \leqslant e^{-1/4\sigma^2} + \langle \bar{v}_i, \bar{v}_j\rangle.\end{aligned}$$

(Both inequalities above use the non-negativity of $\langle v_i, v_j\rangle$.) This bound allows us to estimate the length of $\sum_i d_i f_i$,

$$\begin{aligned}\Big\|\sum\nolimits_i d_i f_i\Big\|^2 &\leqslant \sum_{ij} d_i d_j \|v_i\|\,\|v_j\|\big(e^{-1/4\sigma^2} + \langle \bar{v}_i, \bar{v}_j\rangle\big)\\ &= e^{-1/4\sigma^2}\big(\sum\nolimits_i d_i\|v_i\|\big)^2 + \big\|\sum\nolimits_i d_i v_i\big\|^2\,.\end{aligned}$$

By Jensen's inequality, the first term contributes not more than $e^{-1/4\sigma^2}\sum_i d_i\|v_i\|^2$. From (3.2) it follows that the second term is at most $\delta\sum_i d_i\|v_i\|^2$. The lemma follows. $\square$

Putting together the previous two lemmas (Lemma 3.7 and Lemma 3.8) yields Theorem 3.1 (for $t = 1/4\sigma^2$).

4. SPECTRAL PROFILE OF DIAGONALLY DOMINANT MATRICES

We will prove Theorem 1.3 in this section. The proof has the same outline as the proof of Theorem 1.1 in the previous section (§3). The main difference is in the transformation that is used to map an optimal SDP solution to a collection of functions that satisfies a stronger relaxation of the sparsity constraint.

Let $A \in \mathbb{R}^{n\times n}$ be a symmetric diagonally dominant matrix and let $\mu = (\mu_1, \ldots, \mu_n)$ be a probability measure on $[n]$ (natural choices for μ are uniform, $\mu_i = 1/n$, or proportional to the diagonal of the matrix, $\mu_i = a_{ii}/\operatorname{Tr} A$).

In §4.1, we prove the following analog of Theorem 3.2.

THEOREM 4.1. *Let $f_1, \ldots, f_n\colon \mathbb{R}^n \to \mathbb{R}$ be functions that satisfy*

$$\sum_{ij} \mu_i\mu_j\langle |f_i|, |f_j|\rangle \leqslant \delta\sum_i \mu_i\|f_i\|^2\,. \tag{4.1}$$

Here, $|f|$ is the function $x \mapsto |f(x)|$. Then for every $\varepsilon > 0$,

$$\Lambda_A(\delta/1-2\varepsilon) \leqslant \frac{\sum_{ij} a_{ij}\langle f_i, f_j\rangle}{\varepsilon^3\sum_i \mu_i\|f_i\|^2}\,.$$

In §4.2, we prove an analog of Theorem 3.1. However, in addition to a multiplicative error there is also an additive error. (Furthermore, we can control this additive error only when μ is proportional to the diagonal of A.) The construction is a relatively straight-forward modification of the construction we used for Laplacian matrices.

THEOREM 4.2. *Suppose that μ is proportional to the diagonal of A. For all $\delta > 0$ and $t \geqslant 1$, there exist functions $f_1, \ldots, f_n \colon \mathbb{R}^n \to \mathbb{R}$ such that*

$$\sum_{ij} \mu_i \mu_j \langle |f_i|, |f_j| \rangle \leqslant (\delta + O(2^{-t})) \sum_i \mu_i \|f_i\|^2$$

and

$$\frac{\sum_{ij} a_{ij} \langle f_i, f_j \rangle}{\sum_i \mu_i \|f_i\|^2} \leqslant O(t) \cdot \tilde{\Lambda}_A(\delta) + 2^{-t} \sum_i a_{ii} \,.$$

In §4.3, we prove an analog of Theorem 3.1 that avoids the additive error of the previous theorem. The construction here is quite different from the previous constructions. Our analysis of the construction is unfortunately somewhat more involved.

THEOREM 4.3. *For all $\delta > 0$ and $t \geqslant 1$, there exist functions $f_1, \ldots, f_n \colon \mathbb{R}^n \to \mathbb{R}$ such that*

$$\sum_{ij} \mu_i \mu_j \langle |f_i|, |f_j| \rangle \leqslant O(\delta + 2^{-t}) \sum_i \mu_i \|f_i\|^2$$

and

$$\frac{\sum_{ij} a_{ij} \langle f_i, f_j \rangle}{\sum_i \mu_i \|f_i\|^2} \leqslant O(t) \cdot \tilde{\Lambda}_A(\delta) \,.$$

Putting together Theorem 4.1 and Theorem 4.3 yields an approximation for the spectral profile of diagonally dominant matrices (Theorem 1.3).

4.1 From SDP Solutions with Small 1-Norms to Sparse Vectors (Theorem 4.1)

Let $f_1, \ldots, f_n \colon \mathbb{R}^n \to \mathbb{R}$ be functions that satisfy (4.1). For $\delta' := \delta/(1-2\varepsilon)$, let $t \colon \mathbb{R}^n \to \mathbb{R}_+$ be the function

$$t \stackrel{\text{def}}{=} \tfrac{1}{\delta'} \sum_i \mu_i |f_i| \,.$$

Define non-negative functions $f'_1, \ldots, f'_n \colon \mathbb{R}^n \to \mathbb{R}$ as

$$f'_i(x) \stackrel{\text{def}}{=} \begin{cases} f_i(x) - t(x) & \text{if } f_i(x) > t(x)\,, \\ 0 & \text{if } -t(x) \leqslant f_i(x) \leqslant t(x)\,, \\ f_i(x) + t(x) & \text{if } f_i(x) < -t(x)\,. \end{cases}$$

This method of "sparsifying" the functions is known as *soft thresholding* [21]. Let $\mathsf{soft\text{-}chop}_\tau \colon \mathbb{R} \to \mathbb{R}$ denote the soft-thresholding function $\mathsf{soft\text{-}chop}_\tau(a) = \operatorname{sign}(a) \cdot \max\{|a| - \tau, 0\}$. Note that $f'_i(x) = \mathsf{soft\text{-}chop}_{t(x)} f_i(x)$. Let $F' \colon \mathbb{R}^n \to \mathbb{R}^n_+$ denote $F' := (f'_1, \ldots, f'_n)$.

The proof of the next lemma is the same as the proof of Lemma 3.3 (essentially Markov's inequality).

LEMMA 4.4. *For every $x \in \mathbb{R}^n$,*

$$\mu(\operatorname{supp}(F'(x))) \leqslant \delta' \,.$$

The proof of the next lemma is the same as the proof of Lemma 3.4. (An averaging argument shows that we can find $x \in \mathbb{R}^n$ such that the Rayleigh quotient of A at the vector $F'(x)$ is at most the right-hand side of the equation below. By the previous lemma, this vector $F'(x)$ is δ'-sparse.)

LEMMA 4.5.

$$\Lambda_A(\delta') \leqslant \frac{\sum_{ij} a_{ij} \langle f'_i, f'_j \rangle}{\sum_i \mu_i \|f'_i\|^2} \,.$$

The next lemma corresponds to Lemma 3.5. In the proof of Lemma 3.5 we used properties specific to Laplacians. In the next lemma, we show that the same bound holds also for diagonally dominant matrices.

LEMMA 4.6.

$$\sum_{ij} a_{ij} \langle f'_i, f'_j \rangle \leqslant \sum_{ij} a_{ij} \langle f_i, f_j \rangle \,.$$

PROOF. Since $\mathsf{soft\text{-}chop}_\tau$ is contractive (has Lipschitz constant at most 1) and symmetric (an odd function), i.e., $\sigma \cdot \mathsf{soft\text{-}chop}_\tau(b) = \mathsf{soft\text{-}chop}_\tau(\sigma \cdot b)$, we have for all $a, b \in \mathbb{R}$, $\sigma \in \{\pm 1\}$, and $\tau \geqslant 0$,

$$|\mathsf{soft\text{-}chop}_\tau(a) - \sigma \cdot \mathsf{soft\text{-}chop}_\tau(b)| \leqslant |a - \sigma \cdot b| \,.$$

In particular, $\|f'_i - \sigma f'_j\| \leqslant \|f_i - \sigma f_j\|$ for all $i, j \in [n]$ and $\sigma \in \{\pm 1\}$. By Lemma A.1, there exist matrices $B \in \mathbb{R}^{n \times n}_+$ and $C \in \{1, -1\}^{n \times n}$ such that $\sum_{ij} a_{ij} x_i x_j = \sum_{ij} b_{ij} (x_i - c_{ij} x_j)^2$ for all $x \in \mathbb{R}^n$. We conclude

$$\begin{aligned} \sum_{ij} a_{ij} \langle f_i, f_j \rangle &= \sum_{ij} b_{ij} \|f_i - c_{ij} f_j\|^2 \\ &\leqslant \sum_{ij} b_{ij} \|f'_i - c_{ij} f'_j\|^2 = \sum_{ij} a_{ij} \langle f'_i, f'_j \rangle \,. \quad \square \end{aligned}$$

The next lemma has the same proof as Lemma 3.6.

LEMMA 4.7.

$$\sum_i \mu_i \|f'_i\|^2 \geqslant \varepsilon^3 \sum_i \mu_i \|f_i\|^2 \,.$$

Putting together the previous three lemmas (Lemma 4.5, Lemma 4.6, and Lemma 4.7) yields Theorem 4.1.

4.2 From Arbitrary SDP Solutions to SDP Solutions with Small 1-Norms (Theorem 4.2)

Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be an optimal solution to the semidefinite relaxation $\tilde{\Lambda}_A(\delta)$. The vectors satisfy

$$\sum_{ij} \mu_i \mu_j |\langle v_i, v_j \rangle| \leqslant \delta \sum_i \mu_i \|v_i\|^2 \,. \tag{4.2}$$

For a parameter $\sigma \geqslant 0$ (which will be determined later), we define functions $f_1, \ldots, f_n \colon \mathbb{R}^n \to \mathbb{R}$ as

$$f_i := \tfrac{1}{\sqrt{2}} \|v_i\| \cdot \sqrt{T_{\bar{v}_i} \phi_\sigma} - \tfrac{1}{\sqrt{2}} \|v_i\| \cdot \sqrt{T_{-\bar{v}_i} \phi_\sigma} \,. \tag{4.3}$$

(Recall the notation from §2.3.) Using Lemma 2.3, we compute the inner products of the functions $f_1, \ldots, f_n$ as

$$\langle f_i, f_j \rangle = \|v_i\| \, \|v_j\| \Big(e^{-\|\bar{v}_i - \bar{v}_j\|^2/8\sigma^2} - e^{-\|\bar{v}_i + \bar{v}_j\|^2/8\sigma^2} \Big). \tag{4.4}$$

For the rest of the proof, we will use the following bounds on these inner products.

LEMMA 4.8. *For all $i, j \in \{1, \ldots, n\}$,*

1. $\|f_i\|^2 = \big(1 - e^{-1/2\sigma^2}\big)\|v_i\|^2$,
2. $\langle|\bar{f}_i|, |\bar{f}_i|\rangle \leqslant |\langle\bar{v}_i, \bar{v}_j\rangle| + 2e^{-1/4\sigma^2}$,
3. $\|\bar{f}_i - \bar{f}_j\|^2 \leqslant {}^1\!/_{4\sigma^2}\|\bar{v}_i - \bar{v}_j\|^2 + 2e^{-1/4\sigma^2}$,
4. $\|\bar{f}_i + \bar{f}_j\|^2 \leqslant {}^1\!/_{4\sigma^2}\|\bar{v}_i + \bar{v}_j\|^2 + 2e^{-1/4\sigma^2}$,

Proof. Item 1 is immediate from (4.4). To demonstrate Item 2, we bound the inner product of $|\bar{f}_i|$ and $|\bar{f}_j|$ analog to (4.4) by

$$\langle|\bar{f}_i|, |\bar{f}_j|\rangle \leqslant e^{-(1-\langle\bar{v}_i,\bar{v}_j\rangle)/4\sigma^2} + e^{-(1+\langle\bar{v}_i,\bar{v}_j\rangle)/4\sigma^2}.$$

By symmetry, we may assume $\langle\bar{v}_i, \bar{v}_j\rangle \geqslant 0$. In this case, the second term on the right hand side is at most $e^{-1/4\sigma^2}$, and the first term on the right hand side is at most

$$1 - (1 - e^{-1/4\sigma^2})(1 - \langle\bar{v}_i, \bar{v}_j\rangle/4\sigma^2) \leqslant e^{-1/4\sigma^2} + \langle\bar{v}_i, \bar{v}_j\rangle/4\sigma^2.$$

Here, we used the bound $e^{cx} \leqslant 1 - (1 - e^c)x$ for $x \in [0, 1]$ as in the proof of Lemma 3.8. Item 2 follows by combining the bounds on the two terms. Towards proving Item 3, we bound the inner product of $\bar{f}_i$ and $\bar{f}_j$ from below. If $\langle\bar{v}_i, \bar{v}_j\rangle \geqslant 0$, then $\|\bar{v}_i + \bar{v}_j\|^2 \geqslant 2$ and thus we get from (4.4),

$$\langle\bar{f}_i, \bar{f}_j\rangle \geqslant 1 - \|\bar{v}_i - \bar{v}_j\|^2/8\sigma^2 - e^{-1/4\sigma^2},$$

using the approximation $e^{-x} \geqslant 1 - x$ for $x \in \mathbb{R}$. Otherwise, if $\langle\bar{v}_i, \bar{v}_j\rangle < 0$, we get from (4.4),

$$\langle\bar{f}_i, \bar{f}_j\rangle \geqslant (1 - e^{-1/4\sigma^2})(1 + \langle\bar{v}_i, \bar{v}_j\rangle) - 1 \geqslant \langle\bar{v}_i, \bar{v}_j\rangle - e^{-1/4\sigma^2},$$

using $-e^{cx} \geqslant (1 - e^c)x - 1$ for $x \in [0, 1]$. In both cases, our bound on the inner product $\langle\bar{f}_i, \bar{f}_j\rangle$ implies Item 3. To prove Item 4, we note that the transformation (4.3) is symmetric, i.e., if v is mapped to f, then $-v$ is mapped to $-f$. Hence, Item 3 implies Item 4. □

The next lemma corresponds to Lemma 3.7. However, in our current construction, an additive error term appears (in contrast to the purely multiplicative error in Lemma 3.7). We can control the error by decreasing the standard deviation σ of the Gaussians in our construction.

Lemma 4.9.

$$\frac{\sum_{ij} a_{ij}\langle f_i, f_j\rangle}{\sum_i \mu_i\|f_i\|^2} \leqslant O(\tfrac{1}{\sigma^2})\frac{\sum_{ij} a_{ij}\langle v_i, v_j\rangle}{\sum_i \mu_i\|v_i\|^2} + 2^{-\Omega(\sigma^{-2})}\textstyle\sum_j a_{jj}.$$

Proof. Using Fact 2.4 (similar to the proof of Lemma 3.7), Item 3 and Item 4 of Lemma 4.8 imply for all $i, j \in [n]$ and $\sigma \in \{\pm 1\}$,

$$\|f_i - \sigma f_j\|^2 \leqslant O\big(\tfrac{1}{\sigma^2}\big)\|v_i - v_j\|^2 + 2^{-\Omega(\sigma^{-2})}\|v_i\|\,\|v_j\|.$$

Let $B \in \mathbb{R}_+^{n\times n}$ and $C \in \{1, -1\}^{n\times n}$ as in Lemma A.1. Then,

$$\begin{aligned}\sum_{ij} a_{ij}\langle f_i, f_j\rangle &= \sum_{ij} b_{ij}\|f_i - c_{ij}f_j\|^2\\ &\leqslant O\big(\tfrac{1}{\sigma^2}\big)\sum_{ij} b_{ij}\|v_i - c_{ij}v_j\|^2 + 2^{-\Omega(\sigma^{-2})}\sum_{ij} b_{ij}\|v_i\|\,\|v_j\|\\ &\leqslant O\big(\tfrac{1}{\sigma^2}\big)\sum_{ij} a_{ij}\langle v_i, v_j\rangle + 2^{-\Omega(\sigma^{-2})}\sum_i a_{ii}\|v_i\|^2.\end{aligned}$$

In the last inequality, we used $2\|v_i\|\,\|v_j\| \leqslant \|v_i\|^2 + \|v_j\|^2$ and the observation that $\sum_j b_{ij} \leqslant a_{ii}$. The lemma follows because $\|f_i\|^2 = \Theta(\|v_i\|^2)$ and $\mu_i = a_{ii}/\sum_j a_{jj}$ (one of the assumptions of Theorem 4.2 was that μ is proportional to the diagonal of A). □

The next lemma corresponds to Lemma 3.8. In this section, it is a straight-forward consequence of Item 2 of Lemma 4.8 and the fact that the vectors satisfy (4.2).

Lemma 4.10.

$$\sum_{ij} \mu_i\mu_j\langle|f_i|, |f_j|\rangle \leqslant (\delta + 2^{-\Omega(1/\sigma^2)})\sum_i \mu_i\|f_i\|^2.$$

Proof. From Item 2 of Lemma 4.8, it follows that $\langle|f_i|, |f_j|\rangle \leqslant |\langle v_i, v_j\rangle| + \|v_i\|\,\|v_j\|e^{-1/4\sigma^2}$. Therefore,

$$\sum_{ij} \mu_i\mu_j\langle|f_i|, |f_j|\rangle \leqslant \sum_{ij} \mu_i\mu_j|\langle v_i, v_j\rangle| + e^{-1/4\sigma^2}\big(\textstyle\sum_i \mu_i\|v_i\|\big)^2.$$

Since the vectors satisfy (4.2), the first term of the right-hand side is at most $\delta\sum_i \mu_i\|v_i\|^2$. By convexity, the second term contributes at most $e^{-1/4\sigma^2}\sum_i \mu_i\|v_i\|^2$, as desired. □

Theorem 4.2 follows from Lemma 4.9 and Lemma 4.10.

4.3 Better Construction for Diagonally Dominant Matrices (Theorem 4.3)

Suppose $v_1, \ldots, v_n$ is an optimal solution for the SDP relaxation $\tilde{\Lambda}_A(\delta)$. Recall the soft-thresholding function $\mathsf{soft\text{-}chop}_t(x) = \operatorname{sign}(x)\max\{|x| - t, 0\}$. For a parameter $t > 0$, we define functions $f_1, \ldots, f_n$ as

$$f_i(x) = \|v_i\| \cdot \mathsf{soft\text{-}chop}_t(\langle\bar{v}_i, x\rangle)/\|\mathsf{soft\text{-}chop}_t\|_\gamma.$$

Here, $\|\cdot\|_\gamma$ denotes the 2-norm with respect to the Gaussian measure γ. We equip the space $\{f\colon \mathbb{R}^n \to \mathbb{R}\}$ with the corresponding inner product $\langle f, g\rangle_\gamma = \int f(x)g(x)\,\mathrm{d}\gamma^n(x)$, where γ^n denotes the n-dimensional Gaussian measure. With these definitions, we can check that $\|f_i\|_\gamma = \|v_i\|$ for all $i \in [n]$.

In the next section (§B), we show that the functions satisfy the following properties.

Lemma 4.11. *For all $i, j \in \{1, \ldots, n\}$,*

1. $\langle|\bar{f}_i|, |\bar{f}_j|\rangle \leqslant O(1)|\langle\bar{v}_i, \bar{v}_j\rangle| + 2^{-\Omega(t^2)}$,
2. $\|\bar{f}_i - \bar{f}_j\|^2 \leqslant O(t^2)\|\bar{v}_i - \bar{v}_j\|^2$
3. $\|\bar{f}_i + \bar{f}_j\|^2 \leqslant O(t^2)\|\bar{v}_i + \bar{v}_j\|^2$

Proof of Theorem 4.3. In the previous section (§4.2) we showed that the properties asserted by Lemma 4.11 are enough to conclude as desired

$$\frac{\sum_{ij} a_{ij}\langle f_i, f_j\rangle}{\sum_i \mu_i\|f_i\|^2} \leqslant O(t^2)\frac{\sum_{ij} a_{ij}\langle v_i, v_j\rangle}{\sum_i \mu_i\|v_i\|^2},$$

and

$$\sum_{ij} \mu_i\mu_j\langle|f_i|, |f_j|\rangle \leqslant O(\delta + 2^{-\Omega(1/\sigma^2)})\sum_i \mu_i\|f_i\|^2. \quad \square$$

5. REFERENCES

[1] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. Theory of computing (Singer Island, Fla., 1984).

[2] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[3] S. Arora and S. Kale. A combinatorial, primal-dual approach to semidefinite programs. In *STOC*, pages 227–236, 2007.

[4] B. Barak, M. Hardt, I. Haviv, A. Rao, O. Regev, and D. Steurer. Rounding parallel repetitions of unique games. In *FOCS*, pages 374–383, 2008.

[5] M. Barlow, T. Coulhon, and A. Grigor'yan. Manifolds and graphs with slow heat kernel decay. *Invent. Math.*, 144:609–649, 2001.

[6] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for unique games. In *STOC*, pages 205–214, 2006.

[7] E. Chlamtac, K. Makarychev, and Y. Makarychev. How to play unique games using embeddings. In *FOCS*, pages 687–696, 2006.

[8] F. Chung. Random walks and local cuts in graphs. *Linear Algebra Appl.*, 423(1):22–32, 2007.

[9] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

[10] T. Coulhon, Grigor'yan, and C. Pittet. A geometric approach to on-diagonal heat kernel lower bound on groups. *Ann. Inst. Fourier*, 51:1763–1827, 2001.

[11] A. d'Aspremont, F. R. Bach, and L. E. Ghaoui. Full regularization path for sparse principal component analysis. In *ICML*, pages 177–184, 2007.

[12] A. d'Aspremont, F. R. Bach, and L. E. Ghaoui. Optimal solutions for sparse principal component analysis. *CoRR*, abs/0707.0705, 2007.

[13] A. d'Aspremont, L. E. Ghaoui, M. I. Jordan, and G. R. G. Lanckriet. A direct formulation for sparse pca using semidefinite programming. In *NIPS*, 2004.

[14] S. Goel, R. Montenegro, and P. Tetali. Mixing time bounds via the spectral profile. *Electron. J. Probab.*, 11:no. 1, 1–26 (electronic), 2006.

[15] L. Lovász and R. Kannan. Faster mixing via average conductance. In *Proc. of the Symp. on Theory of Computing*, volume 31 of *STOC*, pages 282–287. ACM, ACM, 1999.

[16] R. Montenegro and P. Tetali. *Mathematical Aspects of Mixing Times in Markov Chains*, volume 1 of *Foundations and Trends in Theoretical Computer Science*. NOW, Boston-Delft, 2006.

[17] B. Morris and Y. Peres. Evolving sets, mixing and heat kernel bounds. *Probab. Th. and Related Fields*, 133(2):245–266, 2005.

[18] P. Raghavendra and D. Steurer. Graph expansion and the unique games conjecture. In *STOC*, 2010.

[19] D. Steurer. Fast SDP algorithms for constraint satisfaction problems. In *SODA*, 2010. To appear.

[20] L. Trevisan. Max cut and the smallest eigenvalue. In *STOC*, pages 263–272, 2009.

[21] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *J. Comput. Graph. Statist.*, 15(2):265–286, 2006.

APPENDIX

A. CHARACTERIZATION OF DIAGONALLY DOMINANT MATRICES

The following lemma is easy to show, e.g., by induction on the number of non-zero entries of the matrix A.

LEMMA A.1. *Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ be a symmetric diagonally dominant matrix. Then, there exists a non-negative matrix $B = (b_{ij}) \in \mathbb{R}_+^{n\times n}$ and a sign matrix $C = (c_{ij}) \in \{1,-1\}^{n\times n}$ such that for every $x \in \mathbb{R}^n$,*

$$\langle x, Ax\rangle = \sum_{ij} b_{ij}(x_i - c_{ij}x_j)^2 .$$

B. PROPERTIES OF TRUNCATED GAUSSIANS

In this section, we will prove Lemma 4.11. For brevity, we write $\theta_t = \mathsf{soft\text{-}chop}_t$ for the soft-thresholding function. We think of θ_t as a member of $L_2(\mathbb{R},\gamma)$, where γ is the standard Gaussian measure on $\mathbb{R}$. In §4.3, we implicitly considered the following mapping T from $\mathbb{R}^n$ to $L_2(\mathbb{R}^n,\gamma^n)$,

$$T(u)\colon g \mapsto \|u\|\cdot\theta_t(\langle g,\bar u\rangle)/\|\theta_t\| .$$

In the following, we will show that for any two unit vectors $u, v \in \mathbb{R}^n$,

$$\langle |T(u)|, |T(v)|\rangle \leqslant |\langle u,v\rangle| + 2^{-\Omega(t^2)} , \tag{B.1}$$

$$\|T(u) - T(v)\|^2 \leqslant O(t^2)\|u-v\|^2 . \tag{B.2}$$

Since $T(-u) = -T(u)$, the bounds (B.1) and (B.2) directly imply Lemma 4.11.

For $\rho \in [-1,1]$, let U_ρ denote the operator on $L_2(\mathbb{R})$

$$\langle f, U_\rho f\rangle = \int f(x) f(\rho x + \sqrt{1-\rho^2}y)\,\mathrm{d}\gamma(x)\,\mathrm{d}\gamma(y) .$$

For any two unit vectors $u, v \in \mathbb{R}^n$ with inner product $\langle u,v\rangle = \rho$, the distribution of $(\langle g,\bar u\rangle, \langle g,\bar v\rangle)$ for a random Gaussian vector $g \in \mathbb{R}^n$ is the same as the distribution of $(x, \rho x + \sqrt{1-\rho^2}y)$ for two independent Gaussians x and y. Hence, $\langle T(u), T(v)\rangle = \langle\theta_t, U_\rho\theta_t\rangle$ and $\langle |T(u)|, |T(v)|\rangle = \langle|\theta_t|, U_\rho|\theta_t|\rangle$. The quantity $\langle\theta_t, U_\rho\theta_t\rangle$ is the Gaussian noise sensitivity of the function θ_t. In the rest of this section, we derive bounds on this sensitivity that imply the bounds (B.1) and (B.2) and thereby Lemma 4.11.

B.1 Gaussian noise sensitivity bounds for soft threshold functions

For $x \in \mathbb{R}$, let $(x)_+ = \max\{x, 0\}$. Note that $\theta_t = \operatorname{sign}(x)(|x|-t)_+$. Recall that $\mathrm{d}\gamma(x) = (\sqrt{2\pi})^{-1}e^{-x^2/2}\,\mathrm{d}x$.

We need the following fact about moments of the exponential distribution.

FACT B.1. *For $k \in \mathbb{N}$ and $z \geqslant 1/t$,*

$$\int_{u=0}^{z} u^k t e^{-tu}\mathrm{d}u = \tfrac{k!}{t^{k+1}} \quad \pm e^{-tz}\cdot O_k(tz)^k .$$

LEMMA B.2. $\|\theta_t\|^2 = \Theta(e^{-t^2/2}/t^3)$.

PROOF. We have

$$\|\theta_t\|^2 = 2\int (x-t)_+^2\,\mathrm{d}\gamma(x) = 2e^{-t^2/2}\int_{u\geqslant 0} u^2 e^{-tu}\,\mathrm{d}\gamma(u) .$$

The lemma follows from Fact B.1 and the fact that $\mathrm{d}\gamma(u) \approx \mathrm{d}u/\sqrt{2\pi}$ for small enough u. □

LEMMA B.3. *If $|\rho| \leqslant 1/2$, then for $\alpha = \sqrt{4/3}$*

$$\langle|\theta_t|, U_\rho|\theta_t|\rangle \leqslant 6\|\theta_{\alpha t}\|^2 .$$

PROOF. Define $\vartheta_t(x) = (x-t)_+$. Let $\tau = \sqrt{1-\rho^2}$. First, we estimate $\langle \vartheta_t, U_\rho \vartheta_t \rangle$,

$$\begin{aligned}
&\int (x-t)_+(\rho x + \tau y - t)_+ \, \mathrm{d}\gamma(x,y) \\
&\leqslant \int (x + \rho x + \tau y - 2t)_+^2 \, \mathrm{d}\gamma(x,y) \qquad \text{(pointwise)} \\
&= \int (\sqrt{2(1+\rho)}\, z - 2t)_+^2 \, \mathrm{d}\gamma(z) \quad \text{(Gaussians are 2-stable)} \\
&= 2(1+\rho) \int \left(z - \sqrt{2/1+\rho}\; t\right)_+^2 \mathrm{d}\gamma(z) \\
&= (1+\rho)\|\theta_{t'}\|^2 \qquad (\text{for } t' = \sqrt{2/1+\rho}\; t) \\
&\leqslant 3/2 \cdot \|\theta_{\alpha t}\|^2 \qquad (\text{for } \alpha = \sqrt{4/3} \leqslant t'/t)
\end{aligned}$$

Let $z = \rho x + \sqrt{1-\rho^2} y$. Notice that

$$|\theta_t(x)\theta_t(z)| = \vartheta_t(|x|)\vartheta_t(|z|) = \sum_{\sigma,\sigma' \in \{\pm 1\}} \vartheta_t(\sigma x)\vartheta_t(\sigma' z)\,.$$

It follows that $\langle |\theta_t|, U_\rho |\theta_t| \rangle = 2\langle \vartheta_t, U_\rho \vartheta_t\rangle + 2\langle \vartheta_t, U_{-\rho}\vartheta_t\rangle$, which implies the lemma using the previous bound $\langle \vartheta_t, U_{\pm\rho}\vartheta_t \rangle \leqslant 3/2 \cdot \|\theta_{\alpha t}\|^2$. ☐

As a direct consequence of previous two lemmas, we get the following bound on $\langle |\theta_t|, U_\rho |\theta_t|\rangle$.

LEMMA B.4.

$$\langle |\theta_t|, U_\rho |\theta_t| \rangle \leqslant 2|\rho| \|\theta_t\|^2 + O(e^{-t^2/6})\|\theta_t\|^2\,.$$

PROOF. Since U_ρ is a Markov operator, $\langle |\theta_t|, U_\rho|\theta_t|\rangle \leqslant \|\theta_t\|^2$. Hence the lemma is trivially true if $|\rho| \geqslant 1/2$. Otherwise, if $|\rho| \leqslant 1/2$, Lemma B.3 asserts $\langle |\theta_t|, U_\rho |\theta_t|\rangle \leqslant \|\theta_{\alpha t}\|^2$ for $\alpha = \sqrt{4/3}$. By Lemma B.2, $\|\theta_{\alpha t}\|^2 = \|\theta_t\|^2 \cdot \Theta(e^{-t^2/6})$. ☐

LEMMA B.5. *If $\rho = 1 - \varepsilon$, then*

$$\langle \theta_t, (I - U_\rho)\theta_t \rangle \leqslant O(t^2 \varepsilon)\|\theta_t\|^2$$

PROOF. Let $\tau = \sqrt{1-\rho^2} \approx \sqrt{2\varepsilon}$ and let $f(x,y) = \theta_t(x) - \theta_t(\rho x + \tau y)$. The inner product $\langle \theta_t, (I-U_\rho)\theta_t\rangle$ is equal to

$$\tfrac{1}{2} \int \big(\theta_t(x) - \theta_t(\rho x + \tau y)\big)^2 \mathrm{d}\gamma^2(x,y) = \tfrac{1}{2}\int f^2 \, \mathrm{d}\gamma^2\,.$$

The function f is invariant under the following isometries,

$$(x,y) \mapsto (-x,-y) \text{ and } (x,y) \mapsto (\rho x + \tau y, \tau x - \rho y).$$

Since the Gaussian measure is also invariant under these isometries, it follows that the integrals of f^2 over the sets $\{(x,y) \mid x \geqslant t\}$, $\{(x,y) \mid x \leqslant -t\}$, $\{(x,y) \mid \rho x + \tau y \geqslant t\}$, and $\{(x,y) \mid \rho x + \tau y \leqslant -t\}$ have the same value. Since these four sets cover the support of f, it follows that

$$\int f^2 \, \mathrm{d}\gamma^2 \leqslant 4 \int_{x \geqslant t} f^2 \, \mathrm{d}\gamma^2(x,y)\,.$$

Since θ_t is 1-Lipschitz, we can bound $|f(x,y)|$ by

$$\begin{aligned}
|\theta_t(x) - \theta_t(\rho x + \tau y)| &\leqslant |\theta_t(x) - \theta_t(\rho x)| + |\theta_t(\rho x) - \theta_t(\rho x + \tau y)| \\
&\leqslant \varepsilon |x| + \tau |y|\,.
\end{aligned}$$

If $x \geqslant t$, then $f(x,y) = x - t - \theta_t(\rho x + \tau y)$. Therefore,

$$\int_{x \geqslant t} f^2 \, \mathrm{d}\gamma^2(x,y) \leqslant 2\varepsilon^2 \int_{x\geqslant t} x^2 \, \mathrm{d}\gamma(x) + 2\tau^2 \int_{x \geqslant t} 1 \, \mathrm{d}\gamma(x)\,.$$

The standard Gaussian tail bound asserts

$$\int_{x \geqslant t} 1 \, \mathrm{d}\gamma(x) \leqslant \tfrac{1}{\sqrt{2\pi}} e^{-t^2/2}/t = t^2 \|\theta_t\|^2\,.$$

Similarly, it is also straight forward to show

$$\int_{x\geqslant t} x^2 \, \mathrm{d}\gamma(x) \leqslant O(te^{-t^2/2}) = t^4 \|\theta_t\|^2\,.$$

Putting the two bounds together, we can conclude

$$\langle \theta_t, (I - U_\rho)\theta_t \rangle \leqslant O(\varepsilon^2 t^4)\|\theta_t\|^2 + O(\varepsilon t^2)\|\theta_t\|^2.$$

This bound implies the lemma, since we can assume $\varepsilon t^2 \leqslant 1$ (otherwise, the lemma is trivially true). ☐