CS 860: Eigenvalues and Polynomials

University of Waterloo

April 5, 2022

List of Lectures

Chapter 1

Overview

The course has three parts. The first part is about eigenvalues, from classical to recent results in spectral graph theory. The second part is about polynomials, mostly on the method of interlacing polynomials and the theory of real-stable polynomials. The third part is about eigenvalues and polynomials, on high-dimensional expanders and log-concave polynomials.

1.1 First Part

The classical spectral graph theory relates (i) the second eigenvalue of the adjacency/Laplacian matrix and (ii) the graph expansion and (iii) the mixing time of random walks. We will start with the fundamental Cheeger's inequality, and then see its applications in analyzing mixing time and in constructing expander graphs.

Around 2010, there are a few interesting extensions/generalizations of Cheeger's inequality using other eigenvalues of the matrix. In previous offerings, we studied these generalizations in details. In this offering, we will just have an overview of these results. Instead, we will study a new Cheeger's inequality for vertex expansion from 2021.

Also around 2010, there are a few interesting results on a linear algebraic formulation of the graph sparsification problem. We will study a random sampling algorithm, and a deterministic algorithm using barrier functions. Then we will also study a related concept called spectral rounding, and see its applications in experimental design and network design.

To provide a more concrete idea, the graph sparsification problem can be formulated as the following pure linear algebraic problem. Given $v_1, \ldots, v_n \in \mathbb{R}^d$ such that $\sum_{i=1}^n v_i v_i^T = I_d$, find scalars $s_1, \ldots, s_n$ with few nonzeros such that $\sum_{i=1}^n s_i v_i v_i^T \approx I_d$. The deterministic result says that it is possible to have only $O(d)$ non-zeros scalars to achieve a constant factor approximation, implying that any undirected graph has a linear-sized sparsifier. It is striking that this linear algebraic formulation provides the best-known way to look at this combinatorial graph problem.

1.2 Second Part

The ideas and techniques developed in spectral sparsification turned out to be surprisingly powerful. It was observed that the deterministic spectral sparsification result is reminiscent to a major open problem in mathematics called the Kadison-Singer problem. This major problem is very remarkably solved in 2013 by a novel probabilistic method and a multivariate extension of the barrier method.

Interestingly, the new probabilistic method is based on viewing eigenvalues as roots of polynomials, and exploiting interlacing properties of these roots. Besides this method of interlacing family and the multivariate barrier method, the solution is also built on a beautiful theory for real-stable polynomials.

In the second part of the course, we will take this polynomial perspective and study the theory of real-stable polynomials to some extent. Then we will see how this is used in establishing interlacing properties for the new probabilistic method. And then we will see the multivariate barrier method and the solution to the Kadison-Singer problem.

Besides the Kadison-Singer problem, this method of interlacing family has several other interesting applications, including the construction of Ramanujan graphs and even the traveling salesman problem. To give one example, consider the following special case of the restricted invertibility problem. Given $v_1, \ldots, v_n \in \mathbb{R}^d$ such that $\sum_{i=1}^n v_i v_i^T = I_d$ and an integer k, the goal is to prove the existence of a subset S of k vectors with large minimum non-zero eigenvalue $\lambda_{\min}\big(\sum_{i\in S} v_i v_i^T\big)$. It turns out that the method of interlacing family allows us to reduce the problem to bounding the maximum root of the polynomial $\partial^n x^n (x-1)^n$!

1.3 Third Part

In the third part, we will study an active research topic about high-dimensional expanders. This new concept provides a local-to-global way to bound the second eigenvalue of the random walk matrix, and it leads to an elegant solution to a long standing open problem called the matriod expansion conjecture in 2019. Since then, lots of progress have been made in using this new approach to analyze mixing time of random walks.

Interestingly, this approach of bounding eigenvalues for random walks is also closely related to analytical properties of some associated polynomials. Consider the following natural algorithm for sampling a random spanning tree of an undirected graph $G = (V, E)$. Start with an arbitrary spanning tree T_0. In the i-th iteration, we add a random edge e to the tree and remove a random edge f on the cycle created and set $T_i := T_{i-1} + e - f$. And we simply repeat many iterations and hope that the tree will look random very soon. Amazingly, the analysis of this algorithm depends on the analytical properties of the following generating polynomials. Given an undirected graph $G = (V, E)$, we associate a variable x_e for each edge $e \in E$, and consider the generating polynomial of spanning trees $p(x) = \sum_{T\in\mathcal{T}} \prod_{e\in T} x_e$ where $\mathcal{T}$ denotes the set of all spanning trees of G. For spanning trees, we will see that this polynomial is completely log-concave and this implies that the above random sampling algorithm is fast.

This polynomial approach has been extended very nicely to prove optimal bounds on mixing time for several other problems, through the so-called log-Sobolev inequality. For these problems, a more general property called fractionally log-concavity is used to prove strong bounds on log-Sobolev inequality. This connection between polynomials and mixing time is very elegant.

High-dimensional expanders are also used in a recent breakthrough in constructing locally testable codes. It is an exciting and very active research area that has shown great promise.

Chapter 2

Linear Algebra

2.1 Eigenvalues and Eigenvectors

Definition 2.1 (Eigenvalues and Eigenvectors)**.** *Let A be an $n \times n$ matrix. A nonzero vector v is called an eigenvector of A if $Av = \lambda v$ for some scalar λ. A scalar λ is called an eigenvalue of A if there exists an eigenvector v with $Av = \lambda v$.*

The multi-set of eigenvalues of A is given by the roots of the characteristic polynomial. This viewpoint will not be used often in the first part of the course, but it will be of central importance in the second part of the course.

Definition 2.2 (Characteristic Polynomial)**.** *Let A be an $n \times n$ matrix. The characteristic polynomial of A is $p_A(x) := \det(xI - A)$.*

Two matrices are said to be similar if one is obtained from another by a change of basis.

Definition 2.3 (Similar Matrices)**.** *A matrix X is similar to another matrix Y if there exists a non-singular matrix B so that $X = BYB^{-1}$.*

It is well known that similar matrices have the same spectrum.

Fact 2.4 (Spectrum of Similar Matrices)**.** *If X is similar to Y, then the multi-set of eigenvalues of X and that of Y are the same.*

Proof. One way to see it is that they have the same characteristic polynomial, as

$$p_X(x) = \det(xI - X) = \det(xI - BYB^{-1}) = \det(B(xI - Y)B^{-1}) = \det(xI - Y) = p_Y(x),$$

where the second last equality is by Fact 2.27. □

Real Symmetric Matrices

In this course, we mostly work with real symmetric matrices, which have all eigenvalues being real numbers.

Theorem 2.5 (Spectral Theorem). *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then all eigenvalues of A are real numbers. Furthermore, there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.*

See L01 of 2019 for a proof, which is based on the proof given in the book by Godsil and Royle [GR].

Remark 2.6 (Undirected and Directed Graphs). *Theorem 2.5 applies to the adjacency/Laplacian matrices of undirected graphs, but not for directed graphs. This is the main reason that spectral graph theory is much more devleoped in undirected graphs. It has been an open direction to develop spectral graph theory for directed graphs.*

Diagonalization: Using the spectral theorem, real symmetric matrices can be written in the following form. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be an orthonormal basis of eigenvectors guaranteed by Theorem 2.5 with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Let V be the $n \times n$ matrix with the i-th column being v_i. Let D be the $n \times n$ diagonal matrix with the (i,i)-th entry being λ_i. Then the conditions $Av_i = \lambda_i v_i$ for $1 \le i \le n$ can be compactly written as $AV = VD$. Since the columns in V form an orthonormal basis, it follows that $V^TV = I$ and thus $V^{-1} = V^T$. So, we can rewrite $AV = VD$ as

$$A = VDV^{-1} = VDV^T.$$

Power of Matrices: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The diagonalization form $A = VDV^T$ is very convenient in computations. To compute A^k, observe that it is simply $A^k = (VDV^T)^k = VD^kV^T$ where D^k is readily computed as D is a diagonal matrix.

This is very useful in analyzing random walks, as P^t is the transition matrix of the random walks after t steps where P is the transition matrix in one step. We will use the eigenvalues of the transition matrix to bound the mixing time of random walks.

Eigen-Decomposition: If $v_1, \ldots, v_n$ form an orthonormal basis, then any $x \in \mathbb{R}^n$ can be written as a linear combination $c_1v_1 + \ldots + c_nv_n$. By orthonormality, for any $1 \le i \le n$,

$$\langle x, v_i\rangle = \langle c_1v_1 + \ldots + c_nv_n, v_i\rangle = \langle c_iv_i, v_i\rangle = c_i.$$

Therefore, for any $x \in \mathbb{R}^n$,

$$x = \langle x, v_1\rangle v_1 + \ldots + \langle x, v_n\rangle v_n = v_1v_1^Tx + \ldots v_nv_n^Tx = \big(v_1v_1^T + \ldots + v_nv_n^T\big)x.$$

Since this is true for all $x \in \mathbb{R}^n$, it follows that

$$v_1v_1^T + \ldots + v_nv_n^T = I_n.$$

Now, if $v_1, \ldots, v_n$ are also eigenvectors of a matrix $A \in \mathbb{R}^{n\times n}$, then for any $x \in \mathbb{R}^n$,

$$Ax = A(v_1v_1^T + \ldots + v_nv_n^T)x = \big(\lambda_1v_1v_1^T + \ldots + \lambda_nv_nv_n^T\big)x.$$

This implies that

$$A = \lambda_1v_1v_1^T + \ldots + \lambda_nv_nv_n^T.$$

Verify that we can also write the inverse using the eigen-decomposition as

$$A^{-1} = \frac{1}{\lambda_1}v_1v_1^T + \ldots + \frac{1}{\lambda_n}v_nv_n^T.$$

Later, this form will also be used to define the "psuedo-inverse" of a matrix A when A is not of full rank.

Positive Semidefinite Matrices

An important class of real symmetric matrices is the class of positive semidefinite matrices. A real symmetric matrix is called positive semidefinite if all of its eigenvalues are nonnegative. This can be seen as a matrix analog of a non-negative number. The following are some equivalent characterizations of a positive semidefinite matrix.

Fact 2.7 (Positive Semidefinite Matrix)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. The following statements are equivalent.*

1. *A is positive semidefinite, i.e. all eigenvalues of A are non-negative.*
2. *For any $x \in \mathbb{R}^n$, it holds that $x^T Ax \geq 0$, i.e. all quadratic forms are non-negative.*
3. *$A = U^T U$ for some matrix $U \in \mathbb{R}^{n\times n}$.*

The notation $A \succcurlyeq 0$ will be used to denote that A is a positive semidefinite matrix.

It is a good exercise to prove this fact; see L01 from 2019 for a proof. A matrix is called positive definite if all eigenvalues of A are positive. It is left as an exercise to find the equivalent characterizations for positive definite matrices as in Fact 2.7.

Check that the set of positive semidefinite matrices forms a convex set. Optimizing a linear function over the set of positive semidefinite matrices with linear constraints is called semidefinite programming. This is a very important class of convex optimization problems that can be solved in polynomial time. We will see it once in this course and we will explain more when we use it.

It is also a good exercise to prove the following useful fact.

Fact 2.8. *For any two positive semidefinite matrices $A, B \in \mathbb{R}^{n\times n}$,*

$$\langle A, B\rangle := \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ij} \geq 0.$$

Optimization Formulation for Eigenvalues

The main reason why eigenvalues are useful for optimization is through the following formulation, which is basically the quadratic form but normalized by the vector length.

Definition 2.9 (Rayleigh Quotient)**.** *The Rayleigh quotient of a vector $x \in \mathbb{R}^n$ with respect to a matrix $A \in \mathbb{R}^{n\times n}$ is defined to be*

$$R_A(x) := \frac{x^T Ax}{x^T x} = \frac{\sum_{i=1}^n \sum_{j=1}^n A_{ij}x_i x_j}{\sum_{i=1}^n x_i^2}.$$

The largest eigenvalue is the maximum value of the Rayleigh quotient.

Lemma 2.10 (Optimization Formulation for α_1)**.** *Suppose $A \in \mathbb{R}^{n\times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$. Then*

$$\alpha_1 = \max_{x\in\mathbb{R}^n} \frac{x^T Ax}{x^T x}.$$

Proof. Let $v_1, v_2, \ldots, v_m$ be the corresponding orthonormal basis of eigenvectors guaranteed by Theorem 2.5. As $v_1, \ldots, v_n$ forms a basis of $\mathbb{R}^n$, any vector $x \in \mathbb{R}^n$ can be written as a linear combination $x = c_1 v_1 + \cdots + c_n v_n$. Then, the numerator can be written as

$$x^T A x = (c_1 v_1 + \cdots + c_n v_n)^T A (c_1 v_1 + \cdots + c_n v_n) = (c_1 v_1 + \cdots + c_n v_n)^T (c_1 \alpha_1 v_1 + \cdots + c_n \alpha_n v_n) = \sum_{i=1}^{n} c_i^2 \alpha_i,$$

where the second equality is because $v_1, \cdots, v_n$ are eigenvectors and the last equality is because $v_1, \cdots, v_n$ are orthonormal. Similarly, the denominator can be written as

$$x^T x = (c_1 v_1 + \cdots + c_n v_n)^T (c_1 v_1 + \cdots + c_n v_n) = \sum_{i=1}^{n} c_i^2.$$

So, the Rayleigh quotient of x is

$$\frac{x^T A x}{x^T x} = \frac{\sum_{i=1}^n c_i^2 \alpha_i}{\sum_{i=1}^n c_i^2} \leq \frac{\alpha_1 \sum_{i=1}^n c_i^2}{\sum_{i=1}^n c_i^2} = \alpha_1.$$

On the other hand, note that v_1 attains the maximum, and the lemma follows. □

This can be extended to characterize other eigenvalues. In particular, we will use the following lemma for the second largest eigenvalue later.

Lemma 2.11 (Optimization Formulation for α_k)**.** *Suppose $A \in \mathbb{R}^{n \times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ and corresponding orthonormal eigenvectors $v_1, \ldots, v_n$. Let T_k be the set of vectors that are orthogonal to $v_1, v_2, \ldots, v_{k-1}$. Then*

$$\alpha_k = \max_{x \in T_k} \frac{x^T A x}{x^T x}.$$

Proof. Let $x \in T_k$. Write $x = c_1 v_1 + \cdots + c_n v_n$. Recall that $c_i = \langle x, v_i \rangle$ from the eigen-decomposition subsubsection. Since $x \in T_k$, it follows that $c_1 = c_2 = \cdots = c_{k-1} = 0$. Using the same calculation as in Lemma 2.10,

$$\frac{x^T A x}{x^T x} = \frac{\sum_{i=k}^n c_i^2 \alpha_i}{\sum_{i=k}^n c_i^2} \leq \frac{\alpha_k \sum_{i=k}^n c_i^2}{\sum_{i=k}^n c_i^2} = \alpha_k.$$

On the other hand, $v_k \in T_k$ and $v_k^T A v_k / v_k^T v_k = \alpha_k$, and the lemma follows. □

The above result gives a characterization of α_k, but it requires the knowledge of the previous eigenvectors. The Courant-Fischer theorem gives a characterization without knowing the eigenvectors, and is more useful in giving bounds on eigenvalues. In words, the Courant-Fischer theorem says that to prove a lower bound on α_k, one needs to show a k-dimensional subspace in which every vector has large Rayleigh quotient, and the best k-dimensional subspace gives the tight lower bound. And to prove an upper bound on α_k, one needs to show a $(n-k+1)$-dimensional subspace in which every vector has small Rayleigh quotient, and the best $(n-k+1)$-dimensional subspace gives the tight upper bound.

Theorem 2.12 (Courant-Fischer Theorem)**.** *Suppose $A \in \mathbb{R}^{n \times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$. Then*

$$\alpha_k = \max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} = \min_{S \subseteq \mathbb{R}^n : \dim(S) = n-k+1} \max_{x \in S} \frac{x^T A x}{x^T x}.$$

Proof. We prove the max-min equality. The min-max equality is similar and is left as an exercise.

Let S_k be the k-dimensional subspace spanned by the first k orthonormal eigenvectors $v_1, \ldots, v_k$, i.e. $S_k = \{x \mid x = c_1 v_1 + \cdots + c_k v_k \text{ for some } c_1, \ldots, c_k \in \mathbb{R}\}$. Then, for any $x \in S_k$,

$$\frac{x^T A x}{x^T x} = \frac{(c_1 v_1 + \cdots + c_k v_k)^T A (c_1 v_1 + \cdots + c_k v_k)}{(c_1 v_1 + \cdots + c_k v_k)^T (c_1 v_1 + \cdots + c_k v_k)} = \frac{\sum_{i=1}^k c_i^2 \alpha_i}{\sum_{i=1}^k c_i^2} \geq \frac{\alpha_k \sum_{i=1}^k c_i^2}{\sum_{i=1}^k c_i^2} = \alpha_k.$$

Therefore,

$$\max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} \geq \min_{x \in S_k} \frac{x^T A x}{x^T x} \geq \alpha_k.$$

To prove that the maximum cannot be greater than α_k, observe that any k-dimensional subspace must intersect the $(n-k+1)$-dimensional subspace T_k spanned by $\{v_k, v_{k+1}, \ldots, v_n\}$. For any $x \in T_k$,

$$\frac{x^T A x}{x^T x} = \frac{\sum_{i=k}^n c_i^2 \alpha_i}{\sum_{i=k}^n c_i^2} \leq \alpha_k.$$

Therefore,

$$\max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} \leq \min_{x \in S \cap T_k} \frac{x^T A x}{x^T x} \leq \alpha_k.$$

□

One consequence of the Courant-Fischer theorem is the eigenvalue interlacing theorem, which will be useful in the second part of the course.

Theorem 2.13 (Cauchy's Interlacing Theorem)**.** *Let $A \in \mathbb{R}^{n \times n}$ be a real symmetric matrix and B be a $(n-1) \times (n-1)$ principle submatrix of A. Then*

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \ldots \geq \alpha_{n-1} \geq \beta_{n-1} \geq \alpha_n,$$

where $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ are the eigenvalues of A and $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_{n-1}$ are the eigenvalues of B.

Proof. Assume without loss of generality that B is in the top left corner of A, that is, the first $n-1$ coordinates.

It should be clear that $\alpha_k \geq \beta_k$ because the search space for α_k is larger than than for β_k. More precisely,

$$\alpha_k = \max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} \geq \max_{S \subseteq \mathbb{R}^{n-1} : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} = \max_{S \subseteq \mathbb{R}^{n-1} : \dim(S) = k} \min_{x \in S} \frac{x^T B x}{x^T x} = \beta_k.$$

Next, we show $\beta_k \geq \alpha_{k+1}$. For any $S \subseteq \mathbb{R}^n$ with $\dim(S) = k+1$, its restriction to the first $n-1$ coordinates (i.e. $S \cap \mathbb{R}^{n-1}$) is of dimension at least k. So, if there is a good $(k+1)$-dimensional subspace for A, then there is a good k-dimensional subspace for B, and so β_k can do as well as α_{k+1}. More formally, let S_{k+1} be the $(k+1)$-dimensional subspace that attains maximum for α_{k+1},

$$\alpha_{k+1} = \min_{x \in S_{k+1}} \frac{x^T A x}{x^T x} \leq \min_{x \in S_{k+1} \cap \mathbb{R}^{n-1}} \frac{x^T A x}{x^T x} \leq \max_{S \subseteq \mathbb{R}^{n-1} : \dim(S) = k} \min_{x \in S} \frac{x^T A x}{x^T x} = \max_{S \subseteq \mathbb{R}^{n-1} : \dim(S) = k} \min_{x \in S} \frac{x^T B x}{x^T x} = \beta_k.$$

□

Perron-Frobenius Theorem

The Perron-Frobenius theorem is the most important result on the eigenvalues and eigenvectors on non-negative matrices. To state it, we need the definitions of an irreducible matrix and the spectral radius of a matrix.

Definition 2.14 (Irreducible Matrix)**.** *A matrix $A \in \mathbb{R}^{n\times n}$ is irreducible if its underlying directed graph $G = (V, E)$ is strongly connected, where the vertex set of G is $V = [n]$ and the edge set of G is $E = \{ij \mid A_{ij} \neq 0\}$.*

The spectral radius of a real symmetric matrix is simply the eigenvalue with largest absolute value. The following is the more general definition for matrices with complex eigenvalues.

Definition 2.15 (Spectral Radius)**.** *The spectral radius $\rho(A)$ of a matrix A is the maximum of the moduli of its eigenvalues.*

The Perron-Frobenius theorem is about the largest eigenvalue and its corresponding eigenvectors, which will be useful in the study of random walks. See chapter 8.8 in [GR] and chapter 8.4 in [HJ13] for more details and proofs.

Theorem 2.16 (Perron-Frobenius Theorem)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a non-negative irreducible matrix.*

1. *The spectral radius $\rho(A)$ is an eigenvalue of A with multiplicity one. In particular, for a real symmetric matrix, the largest eigenvalue is of multiplicity one and its absolute value is the largest.*
2. *If v is an eigenvector with eigenvalue $\rho(A)$, then all the entries of v are nonzero and they have the same sign.*

Matrix Norms

Definition 2.17 (Operator Norm)**.** *Let A be an $m \times n$ matrix. The operator norm $\|A\|_{\text{op}}$ of A is defined as*

$$\|A\|_{\text{op}} := \sup_{x\in\mathbb{R}^n, x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

This is also denoted by $\|A\|_2$ to denote that it is the ratio of the 2-norm of the vectors after and before the linear transformation. But sometimes it is confused with the Schatten 2-norm of A which is also denoted by $\|A\|_2$, so we use the notation $\|A\|_{\text{op}}$ that is also a common notation.

Exercise 2.18 (Operator Norm of a Real Symmetric Matrix)**.** *Show that $\|A\|_{\text{op}}$ is equal to the largest eigenvalue of A when A is a real symmetric matrix.*

The following are some simple properties that will be useful.

Fact 2.19 (Properties of Operator Norm)**.** *Let $A \in \mathbb{R}^{m\times n}$.*

1. *$\|A\|_{\text{op}} \geq 0$ and $\|A\|_{\text{op}} = 0$ if and only if $A = 0$.*
2. *$\|cA\|_{\text{op}} = |c|\|A\|_{\text{op}}$ for every scalar c.*

3. $\|A+B\|_{\text{op}} \le \|A\|_{\text{op}} + \|B\|_{\text{op}}$.
4. $\|Ax\|_2 \le \|A\|_{\text{op}}\|x\|_2$ *for every* $x \in \mathbb{R}^n$.
5. $\|BA\|_{\text{op}} \le \|B\|_{\text{op}}\|A\|_{\text{op}}$

2.2 Formulas and Inequalities

We record some useful formulas and inequalities in this section. A general reference is the book by Horn and Johnson [HJ13].

Inverse

The following formulas are for updating the inverse of a matrix. See wiki for proofs.

Fact 2.20 (Sherman-Morrison Formula)**.** *Suppose* $A \in \mathbb{R}^{n\times n}$ *is an invertible square matrix and* $u, v \in \mathbb{R}^n$ *are column vectors. Then* $A + uv^T$ *is invertible if and only if* $1 + v^T A^{-1} u \neq 0$*. In this case,*

$$(A+uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}$$

Fact 2.21 (Woodbury Formula)**.** *Given a square invertible* $n \times n$ *matrix* A*, a* $n \times k$ *matrix* U*, and a* $k \times n$ *matrix* V*, assuming* $(I_k + VA^{-1}U)$ *is invertible, then*

$$(A+UV)^{-1} = A^{-1} - A^{-1}U\big(I_k + VA^{-1}U\big)^{-1}VA^{-1}.$$

The following formula is for inverting a block matrix, and the Schur complement is a useful definition.

Fact 2.22 (Block Matrix Inversion)**.** *Let* A *and* D *be square matrices and*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

If A *and the Schur complement* $D - CA^{-1}B$ *are invertible, then*

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B\big(D-CA^{-1}B\big)^{-1}CA^{-1} & -A^{-1}B\big(D-CA^{-1}B\big)^{-1} \\ -\big(D-CA^{-1}B\big)^{-1}CA^{-1} & \big(D-CA^{-1}B\big)^{-1} \end{pmatrix}.$$

When a matrix A is not invertible, we can work with the Moore-Penrose pseudoinverse of A.

Definition 2.23 (Pseudoinverse)**.** *Let* $A \in \mathbb{R}^{n\times n}$ *be a real symmetric matrix with eigen-decomposition* $A = \sum_{i=1}^n \lambda_i v_i v_i^T$*. The pseudoinverse of* A*, denoted by* $A^\dagger$*, is defined as*

$$A^\dagger := \sum_{i:\lambda_i \neq 0} \frac{1}{\lambda_i} v_i v_i^T.$$

Check the following properties of pseudoinverse.

Fact 2.24 (Properties)**.** *Let* A *be a real symmetric matrix and* $A^\dagger$ *be its pseudoinverse. Then*

$$AA^\dagger A = A \quad \text{and} \quad A^\dagger A A^\dagger = A^\dagger \quad \text{and} \quad \big(A^\dagger\big)^\dagger = A.$$

Determinant

These formulas about determinants will be used in the second part of the course.

Fact 2.25 (Laplace Co-Factor Expansion). *Let A be a $n \times n$ matrix. For every $1 \le i \le n$,*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} \det\left(A_{[n]\setminus i,[n]\setminus j}\right),$$

where $A_{S,T}$ is the submatrix with rows in $S \subseteq [n]$ and columns in $T \subseteq [n]$.

Applying Laplace expansion recursively gives the Leibniz formula.

Fact 2.26 (Leibniz Formula). *Let A be a $n \times n$ matrix. Then*

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{\sigma(i),i},$$

where S_n is the set of permutations of the set $[n] = \{1, \ldots, n\}$ and $\operatorname{sgn}(\sigma)$ is the sign function of permutation σ which returns $+1$ and -1 for even and odd permutations.

The following is a simple fact.

Fact 2.27 (Product).

$$\det(AB) = \det(A)\det(B).$$

The following result, sometimes known as Sylvester's determinant identity, can be used to deduce that the nonzero eigenvalues of AB and BA are the same (with multiplicity).

Fact 2.28 (Weinstein-Aronszajn Identity).

$$\det(I + AB) = \det(I + BA).$$

The matrix determinant formula keeps track of how the determinant changes after a rank-one update.

Fact 2.29 (Matrix Determinant Formula).

$$\det\left(M - uv^T\right) = \det(M)\left(1 - v^T M^{-1} u\right).$$

The Cauchy-Binet formula will be very useful. One application is to compute the number of spanning trees of a graph.

Fact 2.30 (Cauchy-Binet Formula). *Let A be an $m \times n$ matrix and B be an $n \times m$ matrix. Then*

$$\det(AB) = \sum_{S \in \binom{[n]}{m}} \det\left(A_{[m],S}\right) \det\left(B_{S,[m]}\right)$$

The following formula gives the coefficients of the characteristic polynomials.

Fact 2.31 (Characteristic Polynomial). *Let A be an $n \times n$ matrix.*

$$\det(\lambda I_n - A) = \sum_{k=0}^{n} \lambda^{n-k}(-1)^k \sum_{S \in \binom{[n]}{k}} \det(A_{S,S}).$$

Using Fact 2.31 and Cauchy-Binet formula in Fact 2.30, we have the following identity for the characteristic polynomial of a sum of outer products.

Fact 2.32 (Characteristic Polynomial of Sum of Outer Products). *Let $u_1, \ldots, u_m \in \mathbb{R}^n$.*

$$\det\left(xI - \sum_{i=1}^{m} u_i u_i^T\right) = \sum_{k=0}^{n} \lambda^{n-k}(-1)^k \sum_{S \in \binom{[m]}{k}} \det_k\left(\sum_{i \in S} u_i u_i^T\right),$$

where $\det_k(A) = \sum_{S \in \binom{[n]}{k}} \det(A_{S,S})$.

Trace

Definition 2.33 (Trace). *The trace of a matrix $A \in \mathbb{R}^{n \times n}$, denoted by $\mathrm{Tr}(A)$, is defined as the sum of the diagonal entries of A.*

Fact 2.34 (Cyclic Property of Trace). *For two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$,*

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

By looking at the coefficient of x^{n-1} of the characteristic polynomial $\det(xI - A)$ of $A \in \mathbb{R}^{n \times n}$ in two ways, one can obtain the following useful fact (see L01 from 2019 for a proof).

Fact 2.35 (Trace is Sum of Eigenvalues). *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A \in \mathbb{R}^{n \times n}$. Then*

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

The following are two advanced results about traces; see [Bha97]. We may not need them explicitly in this course.

Fact 2.36 (Golden-Thompson Inequality).

$$\mathrm{Tr}(e^{A+B}) \leq \mathrm{Tr}(e^A) \cdot \mathrm{Tr}(e^B)$$

Fact 2.37 (Lieb-Thirring Inequality). *Let A and B be positive definite matrices and $q \geq 1$. Then*

$$\mathrm{Tr}\left((BAB)^q\right) \leq \mathrm{Tr}(B^q A^q B^q).$$

Matrix Calculus

The formula for differenting the inverse is obtained by differentiating the identity $AA^{-1} = I$.

Fact 2.38 (Inverse).

$$d(A^{-1}) = -A^{-1}(dA)A^{-1}.$$

Jacobi's formula is obtained by differentiating the cofactor expansion in Fact 2.25.

Fact 2.39 (Jacobi's Formula).

$$\frac{d}{dt}\det A(t) = \mathrm{Tr}\left(\mathrm{adj}(A(t)) \cdot \frac{dA(t)}{dt}\right) = \left(\det A(t)\right) \cdot \mathrm{Tr}\left(A(t)^{-1} \cdot \frac{dA(t)}{dt}\right).$$

2.3 References

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169. Springer, 1997. 15

[GR] Christopher D. Godsil and Gordon F. Royle. *Algebraic Graph Theory*. 8, 12

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge; New York, 2nd edition, 2013. 12, 13, 47, 53

Chapter 3

Graph Spectrum

The linear algebraic approach to algorithmic graph theory is to view graphs as matrices, and use concepts and tools in linear algebra to design and analyze algorithms for graph problems. Spectral graph theory uses eigenvalues and eigenvectors of matrices associated with the graph to study its combinatorial properties. In this chapter, we consider the adjacency matrix and the Laplacian matrix of a graph, and see some basic results in spectral graph theory. A general reference for this chapter is the upcoming book by Spielman [Spi19].

3.1 Adjacency Matrix

Definition 3.1 (Adjacency Matrix)**.** *Given an undirected graph $G = (V, E)$ with $V(G) = [n]$, the adjacency matrix $A(G)$ is an $n \times n$ matrix with $A_{ij} = A_{ji} = 1$ if $ij \in E(G)$ and $A_{ij} = A_{ji} = 0$ otherwise.*

The adjacency matrix of an undirected graph is symmetric. So, by the spectral theorem for real symmetric matrices in Theorem 2.5, the adjacency matrix has an orthonormal basis of eigenvectors with real eigenvalues. We denote the eigenvalues of the adjacency matrix by

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n.$$

It is not clear that these eigenvalues should provide any useful information about the combinatorial properties of the graph, but they do, and surprisingly much information can be obtained from them. Let's start with some examples and compute their spectrum.

Example 3.2 (Complete Graphs)**.** *If G is a complete graph, then $A(G) = J - I$ where J denotes the all-one matrix. Any vector is an eigenvector of I with eigenvalue 1. Hence the eigenvalues of A are one less than that of J. Since J is of rank 1, there are $n - 1$ eigenvalues of 0. The all-one vector is an eigenvector of J with eigenvalue n. So, $n - 1$ is an eigenvalue of A with multiplicity one, and -1 is an eigenvalue of A with multiplicity $n - 1$. The is the example with the largest gap between the largest eigenvalue and the second largest eigenvalue.*

Example 3.3 (Complete Bipartite Graphs)**.** *Let $K_{p,q}$ be the complete bipartite graph with p vertices on one side and q vertices on the other side. Its adjacency matrix $A(K_{p,q})$ is of rank 2, so 0 is an eigenvalue with multiplicity $p + q - 2$, and there are two non-zero eigenvalues α and β. By Fact 2.35, the sum of the eigenvalues is equal to the trace of A, which is equal to zero as there*

are no self-loops, and so $\alpha = -\beta$. To determine α, we consider the characteristic polynomial $\det(xI - A) = (x-\alpha)(x+\alpha)x^{p+q-2} = x^{p+q} - \alpha^2 x^{p+q-2}$. Using the Leibniz formula for determinant in Fact 2.26, any term that contributes to x^{p+q-2} must have $p+q-2$ diagonal entries, and the remaining two entries must be $-A_{ij}$ and $-A_{ji}$ for some i, j. There are totally pq such terms, one for each edge, where the sign of the corresponding permutation is -1 as it only has one inversion pair. So, $\alpha^2 = pq$, and thus $|\alpha| = \sqrt{pq}$. To conclude, the spectrum is $(\sqrt{pq}, 0, \ldots, 0, -\sqrt{pq})$, where 0 is an eigenvalue with multiplicity $p+q-2$.

Bipartiteness

It turns out that bipartite graphs can be characterized by the spectrum of their adjacency matrix. The following lemma says that the spectrum of a bipartite graph must be symmetric around the origin on the real line.

Lemma 3.4 (Spectrum of Bipartite Graph is Symmetric)**.** *If G is a bipartite graph and α is an eigenvalue of $A(G)$ with multiplicity k, then $-\alpha$ is an eigenvalue of $A(G)$ with multiplicity k.*

Proof. If G is a bipartite graph, then we can permute the rows and columns of G to obtain the form

$$A(G) = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}.$$

Suppose $u = \begin{pmatrix} x \\ y \end{pmatrix}$ is an eigenvector of $A(G)$ with eigenvalue α. Then

$$\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \alpha \begin{pmatrix} x \\ y \end{pmatrix} \qquad \Longrightarrow \qquad B^T x = \alpha y \ \text{ and } \ By = \alpha x.$$

It follows that

$$\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -By \\ B^T x \end{pmatrix} = \begin{pmatrix} -\alpha x \\ \alpha y \end{pmatrix} = -\alpha \begin{pmatrix} x \\ -y \end{pmatrix},$$

and thus $\begin{pmatrix} x \\ -y \end{pmatrix}$ is an eigenvector of $A(G)$ with eigenvalue $-\alpha$. By this construction, note that k linearly independent eigenvectors with eigenvalue α would give k linearly independent eigenvectors with eigenvalue $-\alpha$, and so their multiplicity is the same. □

The next lemma shows that the converse is also true.

Lemma 3.5 (Symmetric Spectrum Implies Bipartiteness)**.** *Let G be an undirected graph and let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. If $\alpha_i = -\alpha_{n-i+1}$ for each $1 \leq i \leq n$, then G is a bipartite graph.*

Proof. Let k be any positive odd number. Then $\sum_{i=1}^n \alpha_i^k = 0$, by the symmetry of the spectrum. Note that $\alpha_1^k \geq \alpha_2^k \geq \ldots \geq \alpha_n^k$ are the eigenvalues of A^k, because if $Av = \alpha v$ then $A^k v = \alpha^k v$. By Fact 2.35, it follows that $\mathrm{Tr}(A^k) = \sum_{i=1}^n \alpha_i^k = 0$. Observe that $A^k_{i,j}$ is the number of length k walks from i to j in G, which can be proved by a simple induction. So, if G has an odd cycle of length k, then $A^k_{i,i} > 0$ for each vertex i in the odd cycle, and this would imply that $\mathrm{Tr}(A^k) > 0$ as each diagonal entry $A^k_{i,i}$ is non-negative. Therefore, since $\mathrm{Tr}(A^k) = 0$, G must have no odd cycles and is thus a bipartite graph. □

Combining Lemma 3.4 and Lemma 3.5, a graph is bipartite if and only if the spectrum of its adjacency matrix is symmetric around the origin.

Proposition 3.6 (Spectral Characterization of Bipartite Graphs)**.** *Let G be an undirected graph and let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. Then G is a bipartite graph if and only if $\alpha_i = -\alpha_{n-i+1}$ for each $1 \leq i \leq n$.*

When the graph is connected, the characterization is even simpler.

Problem 3.7 (Spectral Characterization of Connected Bipartite Graphs)**.** *Let G be a connected undirected graph and let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. Prove that G is bipartite if and only if $\alpha_1 = -\alpha_n$.*

You may need to use the result of Perron-Frobenius in Theorem 2.16 and also the optimization formulation of eigenvalues in Definition 2.9 to solve this problem.

Largest Eigenvalue

We see some upper and lower bounds on the largest eigenvalue of the adjacnecy matrix in this subsection.

Lemma 3.8 (Max Degree Upper Bound)**.** *Let $G = (V, E)$ be an undirected graph with maximum degree d and let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. Then $\alpha_1 \leq d$.*

Proof. Let v be an eigenvector with eigenvalue α_1. Let j be a vertex with $v(j) \geq v(i)$ for all $i \in V(G)$. Then

$$\alpha_1 \cdot v(j) = (Av)(j) = \sum_{i:ij \in E(G)} v(i) \leq \sum_{i:ij \in E(G)} v(j) = \deg(j) \cdot v(j) \leq d \cdot v(j),$$

which implies that $\alpha_1 \leq d$. □

Look at the proof more closely, we can characterize the connected graphs with α_1 equal to the maximum degree.

Exercise 3.9 (Tight Max Degree Upper Bound)**.** *Let G be a connected undirected graph with maximum degree d and the largest eigenvalue $\alpha_1 = d$. Then G is d-regular and the eigenvalue α_1 is of multiplicity one.*

The maximum degree upper bound can be far from tight. The following problem provides such an example, whose bound is also important in the study of Ramanujan graphs in the second part of the course.

Problem 3.10 (Largest Eigenvalue of a Tree)**.** *Prove that the maximum eigenvalue of the adjacency matrix of a tree of maximum degree d is at most $2\sqrt{d-1}$.*

On the other hand, the average degree is a lower bound on the largest eigenvalue. More generally, the largest eigenvalue is at least the average degree of the densest induced subgraph. One corollary of the following exercise is that the largest eigenvalue is at least the size of a maximum clique minus one.

Exercise 3.11 (Average Degree Lower Bound). *Let $G = (V, E)$ be an undirected graph with largest eigenvalue α_1. For a subset $S \subseteq V$ and a vertex $v \in S$, let $\deg_S(v) := \big|\{u \mid uv \in E \text{ and } u \in S\}\big|$ be the degree of v induced in S. Then*

$$\alpha_1 \geq \max_{S:S\subseteq V} \frac{1}{|S|} \sum_{v\in S} \deg_S(v).$$

We remark that the largest eigenvalue of the adjacency matrix of a connected graph is always of multiplicity one by the Perron-Frobenius Theorem 2.16, and the spectrum of the adjacency matrix satisfies

$$d \geq \alpha_1 > \alpha_2 \geq \ldots \geq \alpha_n \geq -d.$$

In general, I do not know of a nice combinatorial characterization of the largest eigenvalue of the adjacency matrix.

Question 3.12 (Largest Eigenvalue of the Adjacency Matrix). *Is there a better "combinatorial" characterization of the largest eigenvalue of the adjacency matrix of an undirected graph?*

3.2 Laplacian Matrix

The Laplacian matrix plays a more important role in spectral graph theory than the adjacency matrix, as we will see some reasons soon.

Definition 3.13 (Diagonal Degree Matrix). *Let $G = (V, E)$ be an undirected graph with $V(G) = [n]$. The diagonal degree matrix $D(G)$ of G is the $n \times n$ diagonal matrix with $D_{i,i} = \deg(i)$ for each $1 \leq i \leq n$.*

Definition 3.14 (Laplacian Matrix). *Let G be an undirected graph. The Laplacian matrix $L(G)$ of G is defined as $L(G) := D(G) - A(G)$, where $D(G)$ is the diagonal degree matrix in Definition 3.13 and $A(G)$ is the adjacency matrix in Definition 3.1.*

When G is a d-regular graph, the diagonal degree matrix $D(G)$ is simply $d \cdot I_n$, and so the spectrums of the adjacency matrix and the Laplacian matrix are basically the same. That is, let $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be the eigenvalues of the adjacency matrix, and $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be the eigenvalues of the Laplacian matrix. When G is d-regular, it holds that $\lambda_i = d - \alpha_i$ for $1 \leq i \leq n$, and so the i-th largest eigenvalue of A corresponds to the i-th smallest eigenvalue of L. We will use this convention throughout that the eigenvalues of A are denoted by α_i and the eigenvalues of L are denoted by λ_i, and also the eigenvalues of A are ordered in non-increasing order while the eigenvalues of L are ordered in non-decreasing order. So, later on, when we say the k-th eigenvalue of a graph, we either mean the k-th largest eigenvalue of the adjacency matrix or the k-th smallest eigenvalue of the Laplacian matrix.

When the graph is not a regular graph, it may not be easy to relate the eigenvalues of the adjacency matrix and the Laplacian matrix. On one hand, as we mentioned in the previous subsection, it is not so easy to give a characterization of α_1 when the graph is not regular. On the other hand, λ_1 is equal to zero for every graph as we will soon see. We define a matrix for the proof which will also be useful later.

Definition 3.15 (Edge Incidence Matrix). *Let $G = (V, E)$ be an undirected graph and with $V(G) = [n]$ and $m = |E|$. For each edge $e = ij \in E$, let b_e be an n-dimensional vector with the i-th position equal to $+1$ and the j-th position equal to -1 and all other positions equal to 0. Let $B(G)$ be the $n \times m$ edge incidence matrix whose columns are $\{b_e \mid e \in E\}$.*

For an edge $e \in E$, let L_e be its Laplacian matrix with $(L_e)_{i,i} = (L_e)_{j,j} = 1$ and $(L_e)_{i,j} = (L_e)_{j,i} = -1$. Note that the Laplacian L_e of an edge e can be written as $b_e b_e^T$, and the Laplacian of the graph G can be written as

$$L(G) = \sum_{e \in E} L_e = \sum_{e \in E} b_e b_e^T = B(G) \cdot B(G)^T.$$

With this definition in place, the proof that zero is the smallest eigenvalue of Laplacian matrix is straightforward.

Lemma 3.16 (Smallest Eigenvalue of Laplacian Matrix). *The Laplacian matrix $L(G)$ of an undirected graph G is positive semidefinite, and its smallest eigenvalue is zero with the all-one vector being a corresponding eigenvector.*

Proof. As L can be written as BB^T as shown in Definition 3.15, it follows that L is a positive semidefinite matrix by Fact 2.7, and so all eigenvalues of L are non-negative. It is easy to check that $L\vec{1} = 0$, and so 0 is the smallest eigenvalue and $\vec{1}$ is a corresponding eigenvector. □

Having a trivial smallest eigenvalue and a simple corresponding eigenvector is one reason that Laplacian matrix is easier to work with. Another reason is that the Laplacian matrix has a nice quadratic form.

Lemma 3.17 (Quadratic Form for Laplacian Matrix). *Let L be the Laplacian matrix of an undirected graph $G = (V, E)$ with $V(G) = [n]$. For any vector $x \in \mathbb{R}^n$,*

$$x^T L x = \sum_{ij \in E} \big(x(i) - x(j)\big)^2.$$

Proof. Using the decomposition of L in Definition 3.15,

$$x^T L x = x^T \Big(\sum_{ij \in E} L_{ij} \Big) x = x^T \Big(\sum_{ij \in E} b_{ij} b_{ij}^T \Big) x = \sum_{ij \in E} x^T b_{ij} b_{ij}^T x = \sum_{ij \in E} \big(x(i) - x(j)\big)^2.$$

□

We will use Lemma 3.16 and Lemma 3.17 to write a nice formulation for the second smallest eigenvalue when we study Cheeger's inequality.

Connectedness

The second smallest eigenvalue of the Laplacian matrix can be used to determine whether the graph is connected.

Proposition 3.18 (Spectral Characterization of Connected Graphs). *Let G be an undirected graph and let $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues of its Laplacian matrix L. Then G is a connected graph if and only if $\lambda_2 > 0$.*

Proof. Suppose G is disconnected. Then the vertex set can be partitioned into two sets S_1 and S_2 such that there are no edges between them. For a subset $S \subseteq V$, let $\chi_S \in \mathbb{R}^n$ be the characteristic vector of S. Check that both χ_{S_1} and χ_{S_2} are eigenvectors of L with eigenvalue 0. Since χ_{S_1} and χ_{S_2} are linearly independent, it follows that 0 is an eigenvalue with multiplicity at least 2, and thus $\lambda_2 = 0$.

Suppose G is connected. Let x be an eigenvector with eigenvalue 0. Then its quadratic form $x^T Lx = 0$, and so $\sum_{ij \in E}(x(i) - x(j))^2 = 0$ by Lemma 3.17, which implies that $x(i) = x(j)$ for every edge $ij \in E$. Since G is connected, it follows that $x = c \cdot \vec{1}$ for some c, and thus the eigenspace of eigenvalue 0 is of one dimension. Therefore, the eigenvalue 0 has multiplicity one and thus $\lambda_2 > 0$. □

The proof of Proposition 3.18 can be extended to prove the following generalization.

Exercise 3.19 (Spectral Characterization of Number of Components)**.** *Prove that the Laplacian matrix $L(G)$ of an undirected graph G has 0 as its eigenvalue with multiplicity k if and only if the graph G has k connected components.*

3.3 Normalized Adjacency and Laplacian Matrix

Recall that the spectrum of the adjacency matrix satisfies

$$d \geq \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq -d,$$

where the upper bound and the lower bound depend on the maximum degree d of the graph. So, when we relate the eigenvalues of the adjacency matrix to some combinatorial parameters, there is usually a dependency on the maximum degree of the graph.

To remove this dependency and state the Cheeger's inequality nicely, we will use the following normalized version of the adjacency matrix and the Laplacian matrix.

Definition 3.20 (Normalized Adjacency and Laplacian Matrix)**.** *Let G be an undirected graph with no isolated vertices. The normalized adjacency matrix $\mathcal{A}(G)$ of G is defined as $\mathcal{A}(G) := D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, where D is the diagonal degree matrix in Definition 3.13 and A is the adjacency matrix in Definition 3.1.*

The normalized Laplacian matrix $\mathcal{L}(G)$ of G is defined as $\mathcal{L}(G) := D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, where L is the Laplacian matrix in Definition 3.14. Note that $\mathcal{L}(G) = I - \mathcal{A}(G)$.

We will overload notations and still use $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ to denote the eigenvalues of $\mathcal{A}(G)$ and $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ to denote the eigenvalues of $\mathcal{L}(G)$. Since $\mathcal{L}(G) = I - \mathcal{A}(G)$ as stated in Definition 3.20, the spectrums of $\mathcal{L}(G)$ and $\mathcal{A}$ are basically equivalent such that $\lambda_i = 1 - \alpha_i$ for $1 \leq i \leq n$. After the normalization, the eigenvalues are bounded as follows.

Lemma 3.21 (Normalized Eigenvalues)**.** *Let G be an undirected graph with no isolated vertices. Let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its normalized adjacency matrix and $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues of its normalized Laplacian matrix. Then $1 = \alpha_1 \geq \alpha_n \geq -1$ and $0 = \lambda_1 \leq \lambda_n \leq 2$.*

Proof. First we prove that $\lambda_1 = 0$. Note that 0 is an eigenvalue of $\mathcal{L}$, as

$$\mathcal{L}\big(D^{\frac{1}{2}}\vec{1}\big) = \big(D^{-\frac{1}{2}} L D^{-\frac{1}{2}}\big)\big(D^{\frac{1}{2}}\vec{1}\big) = \big(D^{-\frac{1}{2}} L \vec{1}\big) = 0.$$

Note also that

$$\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = D^{-\frac{1}{2}} B B^T D^{-\frac{1}{2}} = \left(D^{-\frac{1}{2}} B\right)\left(D^{-\frac{1}{2}} B\right)^T$$

where B is the edge incidence matrix in Definition 3.15. It follows that $\mathcal{L} = I - \mathcal{A}$ is a positive semidefinite matrix by Fact 2.7, and thus 0 is the smallest eigenvalue of $\mathcal{L}$ and hence $\lambda_1 = 0$. This implies that $\alpha_1 = 1$ as $\lambda_1 = 1 - \alpha_1$.

Next we prove that $\alpha_n \geq -1$. We will show that $D + A$ is also a positive semidefinite matrix. Then the same argument as in the above paragraph can be used to show that $I + \mathcal{A} = D^{-\frac{1}{2}}(D + A)D^{-\frac{1}{2}}$ is also a positive semidefinite matrix, and this would imply that $1 + \alpha_n \geq 0$ and thus $\alpha_n \geq -1$. There are at least two ways to see that $D + A$ is positive semidefinite. One way is to define $\bar{B}$ to be the "unsigned" matrix of B where $\bar{B}_{ij} = |B_{ij}|$ for all $i, j \in V$, and go through the same argument in Definition 3.15 and check that $D + A = \bar{B}\bar{B}^T$. Another way is to use a similar decomposition as in Definition 3.15 and see that the quadratic form of $D + A$ can be written as

$$x^T(D + A)x = \sum_{ij \in E} \left(x_i + x_j\right)^2,$$

which is a sum of squares and thus non-negative. This implies that $\lambda_n \leq 2$ as $\lambda_n = 1 - \alpha_n$. □

3.4 Robust Generalizations

So far we have used the graph spectrum to deduce some simple combinatorial properties of the graph, such as bipartiteness and connectedness, which are easy to deduce directly by simple combinatorial methods such as breadth first search and depth first search. So one may wonder why these spectral characterizations are useful. The key feature of the spectral characterizations is that they can be generalized quantitatively to prove the following robust generalizations of the basic results:

- λ_2 is close to zero if and only if the graph is close to being disconnected. This is the content of Cheeger's inequality.
- λ_n is close to 2 if and only if the graph has a structure close to a bipartite component. This is an analog of Cheeger's inequality for λ_n.
- λ_k is close to zero if and only if the graph is close to having k connected components. This is a generalization called the higher-order Cheeger's inequality.

We will make these statements precise in the next two chapters.

3.5 Problems

The following are some additional problems that are relevant and interesting.

Problem 3.22 (Cycles). *Compute the Laplacian spectrum of C_n, the cycle with n vertices.*

Problem 3.23 (Hypercubes). *A hypercube of n-dimension is an undirected graph with 2^n vertices. Each vertex corresponds to a string of n bits. Two vertices have an edge if and only if their corresponding strings differ by exactly one bit.*

1. *Given two undirected graphs $G = (V, E)$ and $H = (U, F)$, we define $G \times H$ as the undirected graph with vertex set $V \times U$ and two vertices (v_1, u_1), (v_2, u_2) have an edge if and only if either (1) $v_1 = v_2$ and $u_1 u_2 \in F$ or (2) $u_1 = u_2$ and $v_1 v_2 \in E$. Let x be an eigenvector of the Laplacian of G with eigenvalue α, and let y be an eigenvector of the Laplacian of H with eigenvalue β. Show that we can use x and y to construct an eigenvector of the Laplacian of $G \times H$ with eigenvalue $\alpha + \beta$.*

2. *Use (1), or otherwise, to compute the Laplacian spectrum of the hypercube of n dimension.*

Problem 3.24 (Number of Spanning Trees)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$.*

1. *Let B be the edge incidence matrix of G in Definition 3.15. Prove that the determinant of any $(n-1) \times (n-1)$ submatrix of B is ± 1 if and only if the $n-1$ edges corresponding to the columns form a spanning tree of G.*

2. *Let L be the Laplacian matrix of G and let L' be the matrix obtained from L by deleting the last row and last column. Use (1), or otherwise, to prove that $\det(L')$ is equal to the number of spanning trees in G. You can use the Cauchy-Binet formula in Fact 2.30 to solve this problem.*

Problem 3.25 (Wilf's Theorem)**.** *Let G be an undirected graph and α_1 be the largest eigenvalue of its adjacency matrix. Prove that $\chi(G) \leq \lfloor \alpha_1 \rfloor + 1$, where $\chi(G)$ is the chromatic number of G. You may find the Cauchy interlacing Theorem 2.13 useful.*

3.6 References

[Spi19] Daniel A. Spielman. *Spectral and Algebraic Graph Theory*. 2019. 17, 45

Chapter 4

Cheeger's Inequality

Cheeger's inequality is the fundamental result in spectral graph theory, which connects a combinatorial property of a graph and an algebraic quantity of its associated matrix. This connection is important in the theory of expander graphs and the theory of random walks, which we will see in later chapters. The proof is also very useful in graph partitioning which we will see soon.

Cheeger [Che70] proved this inequality in the manifold setting, and the inequality in the graph setting was proved in several works in the 80's [AM85, Alo86, SJ89] with motivations from expander graphs and random walks.

4.1 Graph Conductance

Recall from Proposition 3.18 that a graph G is connected if and only if $\lambda_2 > 0$ where λ_2 is the second smallest eigenvalue of the normalized Laplacian matrix. Cheeger's inequality is the robust generalization that λ_2 is large if and only if the graph is well-connected.

To state this formally, we need to define a measure of how well a graph is connected. There are different natural definitions and we state two of them here.

Definition 4.1 (Edge Expansion)**.** *Let $G = (V, E)$ be an undirected graph. The expansion of a subset $S \subseteq V$ and the expansion of the graph G are defined as*

$$\Phi(S) := \frac{|\delta(S)|}{|S|} \qquad \text{and} \qquad \Phi(G) := \min_{S:|S|\leq|V|/2} \Phi(S),$$

where $\delta(S)$ denotes the set of edges with one endpoint in S and one endpoint in $V - S$.

Definition 4.2 (Edge Conductance)**.** *Let $G = (V, E)$ be an undirected graph. The conductance of a subset $S \subseteq V$ and the conductance of the graph G are defined as*

$$\phi(S) := \frac{|\delta(S)|}{\mathrm{vol}(S)} \qquad \text{and} \qquad \phi(G) := \min_{S:\mathrm{vol}(S)\leq|E|} \phi(S),$$

where $\mathrm{vol}(S) := \sum_{v\in S} \deg(v)$ is called the volume of the subset S. Note that for all $S \subseteq V$, $0 \leq \phi(S) \leq 1$, as it is the ratio of the number of edges cut by S to the total degree in S.

For graph partitioning, a subset $S \subseteq V$ corresponds to the partition of the vertex set V into two parts $(S, V - S)$. In the above definitions for $\Phi(G)$ and $\phi(G)$, note that we only consider the part

with smaller denominator. Both of these definitions try to identify the "bottleneck" in the graph. When the graph G is d-regular, the two definitions are basically equivalent, with $\Phi(G) = d \cdot \phi(G)$. When the graph G is non-regular, the relation between the edge conductance and the second smallest eigenvalue is more elegant.

Theorem 4.3 (Cheeger's Inequality)**.** *Let $G = (V, E)$ be an undirected graph and let λ_2 be the second smallest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(G)$ in Definition 3.20. Then*

$$\frac{1}{2}\lambda_2 \leq \phi(G) \leq \sqrt{2\lambda_2}.$$

The first inequality is called the easy direction, and the second inequality is called the hard direction. We will see that the easy direction corresponds to using the second eigenvalue as a "relaxation" for graph conductance, and the hard direction corresponds to "rounding" a fractional solution to graph conductance to an integral solution.

We say that a graph is an *expander graph* if $\phi(G)$ is large (e.g. $\phi(G) \geq c$ for a constant $0 < c < 1$), and we say that a subset $S \subseteq V$ is a *sparse cut* if $\phi(S)$ is small. Both concepts are very useful.

An expander graph with linear number of edges is an efficient object with diverse applications in theoretical computer science and mathematics; see [HLW06] for an excellent survey. An important implication of Cheeger's inequality is that the second eigenvalue of the normalized Laplacian matrix can be used to certify that a graph is an expander graph, which provides an algebraic way to construct expander graphs that turns out to be very fruitful.

Finding a sparse cut is useful in designing divide-and-conquer algorithms, with applications in image segmentation, data clustering, community detection, VLSI design, among others. The algorithmic implication of Cheeger's inequality is discussed in the following subsection.

Spectral Partitioning Algorithm

A popular heuristic in finding a sparse cut in practice is the following spectral partitioning algorithm.

Algorithm 1 Spectral Partitioning Algorithm

Require: An undirected graph $G = (V, E)$ with $V = [n]$ and $m = |E|$.

1: Compute the second smallest eigenvalue λ_2 of $\mathcal{L}(G)$ and a corresponding eigenvector $x \in \mathbb{R}^n$.
2: Sort the vertices so that $x_1 \geq x_2 \geq \ldots \geq x_n$.
3: Let $S_i = [i]$ if $\text{vol}_G\left([i]\right) \leq m$ and let $S_i = [n] \setminus [i]$ if $\text{vol}_G\left([i]\right) > m$.
4: **return** $\min_{i:1 \leq i \leq n-1}\{\phi(S_i)\}$.

The algorithm is strikingly simple, with only a few lines of code if we use some mathematical software such as MATLAB, which is one reason why this heuristic is popular. The algorithm only checks the linear number of solutions defined by the ordering in a second eigenvector, although there are exponentially many subsets $S \subseteq V$.

There is a near-linear time randomized algorithm to compute an approximate eigenvector of the second eigenvalue, using the power method and a fast Laplacian solver. So, the algorithm is also fast theoretically, but we won't discuss the details in this chapter.

The main reason that it is popular is that it performs very well in various applications, especially in image segmentation and clustering, and it was considered a breakthrough in image segmentation about 20 year ago [SM00].

The proof of Cheeger's inequality will provide a nontrivial performance guarantee of the spectral partitioning algorithm, that it will always output a set S with conductance $\phi(S) \leq \sqrt{2\lambda_2} \leq 2\sqrt{\phi(G)}$. When $\phi(G)$ is a constant, this is a constant factor approximation algorithm. When $\phi(G)$ is small, say $\phi(G) \leq 1/n$, the approximation ratio could be as bad as $\Theta(\sqrt{n})$. It has been an open problem to explain the empirical success rigorously, and we will come back to this in the next chapter.

4.2 Easy Direction

In this section, we prove the easy direction of Cheeger's inequality. For simplicity of exposition, we assume the graph is d-regular, so $\mathcal{L}(G) = \frac{1}{d}L(G)$. We will outline the modifications needed for the analysis of non-regular graphs in a subsection at the end of this chapter, but those are just some additional manipulations where all the main ideas are already in the d-regular case.

First, we start with the nice optimization formulation of the second eigenvalue of the normalized Laplacian matrix using the Rayleigh quotient in Definition 2.9.

Lemma 4.4 (Optimization Formulation for λ_2)**.** *Let $G = (V, E)$ be an undirected d-regular graph with $V = [n]$ and λ_2 be the second smallest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(G)$. Then*

$$\lambda_2 = \min_{x \in \mathbb{R}^n : x \perp \vec{1}} R_{\mathcal{L}}(x) = \min_{x \in \mathbb{R}^n : x \perp \vec{1}} \frac{x^T \mathcal{L} x}{x^T x} = \min_{x \in \mathbb{R}^n : x \perp \vec{1}} \frac{x^T L x}{d \cdot x^T x} = \min_{x \in \mathbb{R}^n : x \perp \vec{1}} \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{d \sum_{i \in V} x(i)^2}.$$

Proof. The first equality is by Lemma 2.11, although it was stated in the maximization form, the same proof works for the minimization form. The last equality is by Lemma 3.17. □

The observation is that the minimization problem of graph conductance can be formulated in a similar way. The following lemma holds for non-regular graphs.

Lemma 4.5 (Optimization Formulation for Graph Conductance)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Then*

$$\phi(G) = \min_{x \in \{0,1\}^n} \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot x(i)^2} \qquad \text{subject to} \qquad \sum_{i \in V} \deg(i) \cdot x(i)^2 \leq |E|.$$

Proof. Each feasible solution $S \subseteq V$ with $\mathrm{vol}(S) \leq |E|$ in the graph conductance problem corresponds to a feasible solution $\chi_S \in \{0,1\}^m$ with $\sum_{i \in V} \deg(i) \cdot x(i)^2 \leq |E|$ in this formulation, and vice versa. Note that the numerator counts the number of edges in $\delta(S)$ and the denominator computes the volume $\mathrm{vol}(S)$ of S. □

Note that for d-regular graphs, the constraint simplifies to $\sum_{i \in V} x(i)^2 \leq n/2$.

Intuition: The main difference between these two formulations is that the former optimizes over the continuous domain $x \in \mathbb{R}^n$, while the latter optimizes over the discrete domain $x \in \{0,1\}^n$. A good way to think of the relation between the two optimization problems is that the former problem is a "relaxation" of the latter problem. This is a common idea in the design of approximation algorithms. The latter problem is an NP-hard optimization problem, because of the discrete domain. The relaxation idea is to optimize over a larger continuous domain, so that the problem can be solved in

polynomial time. Since we optimize over a larger domain, the objective value of the former problem could only be smaller than that of the latter problem, and so we expect that $\lambda_2 \leq \phi(G)$. This is the main intuition for the easy direction. For these two formulations, however, there are also some differences between the constraints, and it only holds that $\lambda_2 \leq 2\phi(G)$.

Proof of the Easy Direction in the d-Regular Case. To upper bound λ_2, we just need to find a vector $x \perp \vec{1}$ and compute its Rayleigh quotient $R_{\mathcal{L}}(x)$. Let $S \subseteq V$ be a subset of vertices with $|S| \leq \frac{|V|}{2}$. Consider the following "binary" solution $x \in \mathbb{R}^n$ with

$$x(i) = \begin{cases} {}^{+1}/_{|S|} & \text{if } i \in S \\ {}^{-1}/_{|V-S|} & \text{if } i \notin S \end{cases}.$$

By construction, $x \perp \vec{1}$, and so

$$\lambda_2 \leq R_{\mathcal{L}}(x) = \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{d \sum_{i \in V} x_i^2} = \frac{|\delta(S)| \cdot \big(\frac{1}{|S|} + \frac{1}{|V-S|}\big)^2}{d\big(|S| \cdot \frac{1}{|S|^2} + |V-S| \cdot \frac{1}{|V-S|^2}\big)} = \frac{|\delta(S)| \cdot |V|}{d \cdot |S| \cdot |V-S|} \leq 2\phi(S),$$

where the last inequality uses the assumption that $|S| \leq \frac{|V|}{2}$ which implies that $\frac{|V|}{|V-S|} \leq 2$. □

4.3 Hard Direction

By optimizing over a larger domain, however, the objective value of the continuous problem will typically be smaller than that of the discrete problem.

Tight Example

Consider the cycle of $4n$ vertices. One can compute the second eigenvector of the cycle exactly, but we don't do it here. Recall that $\lambda_2 = \min_{x \perp \vec{1}} x^T \mathcal{L} x / x^T x$, so to give an upper bound we just need to demonstrate a solution with small objective value. Consider

$$x = \Big(1, 1 - \frac{1}{n}, 1 - \frac{2}{n}, \ldots, \frac{1}{n}, 0, -\frac{1}{n}, \ldots, -1 + \frac{1}{n}, -1, -1 + \frac{1}{n}, \ldots, -\frac{1}{n}, 0, \frac{1}{n}, \ldots, 1 - \frac{1}{n}\Big).$$

Then $x \perp \vec{1}$, and so

$$\lambda_2 \leq \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{2 \sum_{i \in V} x(i)^2} = \Theta\Big(\frac{n\big(\frac{1}{n}\big)^2}{n}\Big) = \Theta\Big(\frac{1}{n^2}\Big).$$

On the other hand, it is easy to verify that the conductance of the cycle of $4n$ vertices is $\Theta(\frac{1}{n})$. This is an example where the hard direction $\phi(G) \leq \sqrt{2\lambda_2}$ is tight up to a constant factor.

Rounding

For the discrete optimization problem of graph conductance, we would like the solution to be a binary solution as in the proof of the easy direction. But once we relax the problem to the continuous domain (so that it becomes polynomial time solvable), the optimal solution x could be very "fractional" or "continuous" as the above example shown. For the hard direction, the task is

to prove that there is always a binary solution z with objective value at most the square root of the objective value of the continuous solution x. A typical way in approximation algorithms to do this task is to "round" the "fractional/continuous" solution x to an "integral/binary" solution z and bound the objective value of z in terms of the objective value of x. This is what we will do.

Intuition

Think of the optimizer x to the optimization problem in Lemma 4.4 as embedding the vertices of the graph into the real line, so that for most edges $|x(i)-x(j)|$ is small. A natural strategy is to do a "threshold rounding", where we pick a threshold t and set $z(i) = 0$ if $x(i) < t$ and $z(i) = 1$ if $x(i) \geq t$. A simple analogy is that if most rows/edges have few nonzeros, then there is a column/threshold with few nonzeros. This intuition can be made precise if the optimization problem is of the form

$$\min_{x \in \mathbb{R}^n_+} \frac{\sum_{ij \in E} |x(i) - x(j)|}{d \sum_{i \in V} x(i)},$$

but the optimization problem in Lemma 4.4 is a sum of quadratic terms and this is basically where the square root loss in the hard direction comes from.

Truncation

The proof of the hard direction has two steps. The first step is a preprocessing step that truncates an optimizer $x \in \mathbb{R}^n$ to the continuous problem in Lemma 4.4 to a vector $y \in \mathbb{R}^n$ with $\mathrm{vol}(\mathrm{supp}(y)) \leq |E|$, where $\mathrm{supp}(y) := \{i \mid y(i) \neq 0\}$. This is to ensure that the solution S produced in the second step satisfies $\mathrm{vol}(S) \leq |E|$, satisfying the constraint in the discrete problem in Lemma 4.5.

There are two ways to do this step. The first way requires that x is indeed an eigenvector, and the proof is shorter and is enough for establishing Cheeger's inequality. The second way only requires that x is perpendicular to the first eigenvector, which is important for algorithmic purpose, but the proof is a bit longer. We will present the proof for the first way and outline the proof for the second way in the problem subsection in the end. The following lemma is from [HLW06] and it holds for general undirected graphs.

Lemma 4.6 (Truncating Eigenvector)**.** *Let $G = (V, E)$ be an undirected graph and $x \in \mathbb{R}^n$ be an eigenvector of $\mathcal{L}(G)$ with eigenvalue λ. Let $x_+ \in \mathbb{R}^n$ be the vector with $x_+(i) = \max\{x(i), 0\}$ for $1 \leq i \leq n$, and $x_- \in \mathbb{R}^n$ be the vector with $x_-(i) = \min\{x(i), 0\}$ for $1 \leq i \leq n$. Then $R_{\mathcal{L}}(x_+) \leq R_{\mathcal{L}}(x) = \lambda$ and $R_{\mathcal{L}}(x_-) \leq R_{\mathcal{L}}(x) = \lambda$.*

Proof. For each vertex $i \in \mathrm{supp}(x_+)$, by the definition of normalized Laplacian matrix in Definition 3.20 and the assumption that x is an eigenvector with eigenvalue λ,

$$(\mathcal{L}x_+)(i) = x_+(i) - \sum_{j:ij \in E} \frac{x_+(j)}{\sqrt{\deg(i)\deg(j)}} \leq x(i) - \sum_{j:ij \in E} \frac{x(j)}{\sqrt{\deg(i)\deg(j)}} = (\mathcal{L}x)(i) = \lambda \cdot x(i) = \lambda \cdot x_+(i).$$

This implies that

$$\langle x_+, \mathcal{L}x_+ \rangle = \sum_{i \in \mathrm{supp}(x_+)} x_+(i) \cdot (\mathcal{L}x_+)(i) \leq \sum_{i \in \mathrm{supp}(x_+)} \lambda \cdot x_+(i)^2 = \lambda \cdot \|x_+\|_2^2.$$

Therefore, $R_{\mathcal{L}}(x_+) = \langle x_+, \mathcal{L}x_+ \rangle / \|x_+\|_2^2 \leq \lambda$, and the proof is the same for x_-. □

Let x be an eigenvector with eigenvalue λ_2. Since $x \neq 0$ and x is perpendicular to the first eigenvector which is a positive vector, both $\mathrm{supp}(x_+)$ and $\mathrm{supp}(x_-)$ are non-empty sets. By choosing either x_+ or x_- that has a smaller volume in its support and taking a proper normalization, we arrive at the following corollary.

Corollary 4.7 (Preprocessed Vector)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$ and λ_2 be the second eigenvalue of $\mathcal{L}(G)$. There exists a vector $y \in \mathbb{R}^n$ that satisfies (i) $y \geq 0$, (ii) $R_{\mathcal{L}}(y) \leq \lambda_2$, (iii) $\mathrm{vol}(\mathrm{supp}(y)) \leq \frac{1}{2}\mathrm{vol}(V) = |E|$, and (iv) $\|y\|_2^2 = 1$.*

For algorithmic purpose, we may not be able to compute an eigenvector exactly, but rather a vector x that is perpendicular to the first eigenvector and $R_{\mathcal{L}}(x) \approx \lambda_2$. In the problem subsection in the end, we describe how to truncate the vector to satisfy Corollary 4.7, which is similar but with an additional shifting/centering step.

In the d-regular case, to summarize, the truncation step transforms a vector x with small Rayleigh quotient that satisfies $x \perp \vec{1}$ in the continuous problem in Lemma 4.4 into a vector y with small Rayleigh quotient that satisfies $|\mathrm{supp}(y)| \leq n/2$ that is required in the discrete problem in Lemma 4.5.

Threshold Rounding

The main step in the hard direction is the threshold rounding step hinted earlier, which takes a vector y in Corollary 4.7 and outputs a set $S \subseteq \mathrm{supp}(y)$ with small conductance $\phi(S)$. As described in the spectral partitioning algorithm, we will only consider those threshold/level sets $S'_t := \{i \in V \mid y(i) \geq t\}$ for $t > 0$, as in every proof of Cheeger's inequality. Our proof will follow that of Trevisan [Tre16], whose idea is to choose a random t and considers the level set $S_t := \{i \in V \mid y(i)^2 \geq t\}$, and to bound the conductance of S_t by computing the expectation of the numerator and the expectation of the denominator separately. The idea of choosing a random t is similar to the idea of randomized rounding in approximation algorithms, and his analysis of computing the expectations separately simplifies the proof.

Lemma 4.8 (Threshold Rounding)**.** *Let $G = (V, E)$ be an undirected d-regular graph with $V = [n]$. Let $y \in \mathbb{R}^n_+$ be a vector with non-negative entries. There exists $t > 0$ such that the threshold set $S_t := \{i \in [n] \mid y(i)^2 \geq t\}$ is nonempty and satisfies $\phi(S_t) \leq \sqrt{2R_{\mathcal{L}}(y)}$.*

Proof. For convenience, we scale y so that $\max_i y(i) = 1$. Let $t \in (0, 1]$ be chosen uniformly at random. Note that the set S_t is nonempty by construction. In the following, we compute the expected value of the numerator and of the denominator in Lemma 4.5 separately.

For an edge $ij \in E$, note that the probability that $ij \in \delta(S_t)$ is $|y(i)^2 - y(j)^2|$, when the random threshold t falls between $y(i)^2$ and $y(j)^2$. By linearity of expectation,

$$\mathbb{E}_t\big[|\delta(S_t)|\big] = \sum_{ij \in E} \Pr(ij \in \delta(S_t)) = \sum_{ij \in E} \big|y(i)^2 - y(j)^2\big| = \sum_{ij \in E} \big|y(i) - y(j)\big| \cdot \big|y(i) + y(j)\big|.$$

To relate this expected value to the numerator of the Rayleigh quotient in Lemma 4.4, the Cauchy-Schwarz inequality is used as in every proof of Cheeger's inequality so that

$$\mathbb{E}_t\big[|\delta(S_t)|\big] \leq \sqrt{\sum_{ij \in E} \big|y(i) - y(j)\big|^2} \sqrt{\sum_{ij \in E} \big|y(i) + y(j)\big|^2} \leq \sqrt{\sum_{ij \in E} \big|y(i) - y(j)\big|^2} \sqrt{2d \cdot \sum_{i \in V} y(i)^2},$$

where the second inequality holds because $\sum_{ij\in E}\big|y(i)+y(j)\big|^2 \le \sum_{ij\in E} 2\big(y(i)^2+y(j)^2\big) = 2d\sum_{i\in V} y(i)^2$ where the assumption that G is d-regular is used.

For a vertex $i \in V$, note that the probability that $i \in S_t$ is $y(i)^2$, when the random threshold t is smaller than $y(i)^2$. By linearity of expectation,

$$\mathbb{E}_t\big[d|S_t|\big] = d\cdot\sum_{i\in V}\Pr(i\in S_t) = d\cdot\sum_{i\in V} y(i)^2.$$

Therefore,

$$\frac{\mathbb{E}_t\big[|\delta(S_t)|\big]}{\mathbb{E}_t\big[d|S_t|\big]} \le \sqrt{\frac{2\sum_{ij\in E}\big|y(i)-y(j)\big|^2}{d\cdot\sum_{i\in V} y(i)^2}} = \sqrt{2R_{\mathcal{L}}(y)}.$$

Note that we cannot conclude from this that $\mathbb{E}_t[\phi(S_t)] = \mathbb{E}_t\big[|\delta(S_t)|/d|S_t|\big] \le \sqrt{2R_{\mathcal{L}}(y)}$, but we can conclude from this that

$$\mathbb{E}_t\Big[|\delta(S_t)| - d|S_t|\sqrt{2R_{\mathcal{L}}(y)}\Big] \le 0 \quad\Longrightarrow\quad \exists t>0 \text{ with } \phi(S_t) = \frac{|\delta(S_t)|}{d|S_t|} \le \sqrt{2R_{\mathcal{L}}(y)}.$$

□

Analysis of the Spectral Partitioning Algorithm

We summarize the proof of the hard direction, which also provides an analysis of the spectral partitioning algorithm.

Proof of the Hard Direction in the d-Regular Case. Let $x \in \mathbb{R}^n$ be an eigenvector of $\mathcal{L}(G)$ with eigenvalue λ_2. First we apply the truncation step in Lemma 4.6 and Corollary 4.7 to obtain a vector $y \in \mathbb{R}^n$ with $R_{\mathcal{L}}(y) \le R_{\mathcal{L}}(x) = \lambda_2$ and $|\operatorname{supp}(y)| \le n/2$. Then we apply the threshold rounding step in Lemma 4.8 on y to obtain a nonempty set $S_t = \{i \in [n] \mid y(i)^2 \ge t\}$ with $t > 0$ and $\phi(S_t) \le \sqrt{2R_{\mathcal{L}}(y)} \le \sqrt{2\lambda_2}$. Since $S_t \subseteq \operatorname{supp}(y)$, it follows that $0 < |S_t| \le |\operatorname{supp}(y)| \le n/2$ and thus $\phi(G) \le \phi(S_t) \le \sqrt{2\lambda_2}$. Finally, note that S_t is a threshold set of y, which is also a threshold set of x by the construction in Lemma 4.6, as y is either x_+ or x_-. This implies that the spectral partitioning algorithm has considered this set, and thus it will output a set S with $\phi(S) \le \sqrt{2\lambda_2}$.
□

4.4 Discussions

We discuss more about the performance of the spectral partitioning algorithm and also outline the modifications needed for general weighted graphs.

More Examples

Both sides of Cheeger's inequality are tight, even the constants are tight. For the easy direction, one can check that it is tight for the hypercubes; see Problem 3.23. For the hard direction, we have already seen that it is tight up to a constant factor for the cycles. It is possible to assign edge weights to the cycle so that even the constant $\sqrt{2}$ is tight, and we leave it as a challenging example to work out.

These are the standard examples to show that both sides of Cheeger's inequality are tight, but they do not yet provide much insights about how the spectral partitioning algorithm only outputs a set S with $\phi(S) \approx \sqrt{\phi(G)}$. For the cycle examples where the $\phi(G) \approx \sqrt{\lambda_2}$, the spectral partitioning actually works perfectly to output a set S with $\phi(S) \approx \phi(G)$, because it outputs a set S with $\phi(S) \approx \sqrt{\lambda_2} \approx \phi(G)$. Indeed, it is a general phenomenon that rounding algorithms work perfectly for the worst integrality gap examples.

So, to find an example where the spectral partitioning algorithm performs poorly, we need to look at the examples where the easy direction is tight but the algorithm outputs a set S with $\phi(S) \approx \sqrt{\lambda_2} \approx \sqrt{\phi(G)}$. Actually, for the hypercube examples where the easy direction is tight, there are vectors in the second eigenspace where the spectral partitioning algorithm performs perfectly and performs poorly.

Problem 4.9 (Spectral Partitioning for Hypercubes)**.** *Let G be the hypercube of dimension d with 2^d vertices and $\mathcal{L}(G)$ be its normalized Laplacian matrix.*

1. *Show that there is an eigenvector vector $x \in \mathbb{R}^{2^d}$ of $\mathcal{L}(G)$ with eigenvalue λ_2 so that the spectral partitioning algorithm applied on x outputs a set S with $\phi(S) = R_{\mathcal{L}}(x) = \frac{1}{2}\lambda_2$.*

2. *Show that there is an eigenvector vector $y \in \mathbb{R}^{2^d}$ of $\mathcal{L}(G)$ with eigenvalue λ_2 so that the spectral partitioning algorithm applied on y outputs a set S with $\phi(S) \approx \sqrt{R_{\mathcal{L}}(y)} = \sqrt{\lambda_2}$.*

Since we do not have control over which eigenvector in the second eigenspace is returned, this provides an example where the spectral partitioning algorithm could perform poorly. But perhaps this example is not so satisfying as we do not see clearly how the spectral partitioning algorithm is fooled.

We construct such an example in the following by tweaking the cycle example. Let G be the weighted graph with vertices $\{v_1, \ldots, v_n, v_{n+1}, \ldots, v_{2n}\}$, and two cycles $(v_1, v_2, \ldots, v_n)$ and $(v_{n+1}, v_{n+2}, \ldots, v_{2n})$ where every edge in these cycles is of weight one, and a "hidden" matching $\{v_1v_{n+1}, v_2v_{n+2}, \ldots, v_nv_{2n}\}$ where every edge in the matching is of weight say $100/n^2$. Then it is easy to see that the set of smallest conductance is the set $S := \{v_1, \ldots, v_n\}$ with $\phi(S) = O(1/n^2)$. However, the edges in the hidden matching are so light that the spectral partitioning algorithm did not "feel" them, and still thinks that the embedding of the cycle is the best embedding of the vertices onto the real line. Indeed, one can verify that the second eigenvector x in this example is still the same as that in the cycle of n vertices, with $x(v_i) = x(v_{n+i})$ for $1 \le i \le n$. Therefore, λ_2 is still $O(1/n^2)$ which is close to $\phi(G)$, but the cut of smallest conductance is completely lost in x and every threshold set has conductance $\Omega(1/n)$. This is a more insightful example to see how the spectral partitioning algorithm is fooled. This example is a weighted graph, but one can also modify this example slightly to keep the same structure while making the graph unweighted.

Cheeger's Inequality for General Weighted Graphs

Once we understand the proof for the d-regular case, it is not difficult to extend it to general weighted graphs.

Let $G = (V, E)$ be an edge weighted graph with a non-negative weight $w(e) \ge 0$ on each edge $e \in E$. The weighted degree of a vertex i is defined as $\deg_w(i) = \sum_{j:ij \in E} w(ij)$, and the diagonal degree matrix is denoted by D_w. The weighted adjacency matrix A_w is defined so that $(A_w)_{i,j} = w(ij)$ for all $i, j \in V$. The weighted Laplacian matrix L_w is defined as $D_w - A_w$, and the weighted

normalized Laplacian matrix $\mathcal{L}_w$ is defined as $D_w^{-\frac{1}{2}} L_w D_w^{-\frac{1}{2}}$. Check that the quadratic form $x^T L_w x = \sum_{ij \in E} w(ij)\big(x(i) - x(j)\big)^2$. The following is a generalization of Lemma 4.4 for λ_2 of $\mathcal{L}_w(G)$, which can be obtained by a change of variable.

Exercise 4.10 (General Optimization Formulation for λ_2)**.** *Let $G = (V, E)$ be a weighted graph with $V = [n]$ and λ_2 be the second smallest eigenvalue of $\mathcal{L}_w(G)$. Show that*

$$\lambda_2 = \min_{y \in \mathbb{R}^n : \sum_{i \in V} \deg_w(i) y(i) = 0} \frac{y^T L_w y}{y^T D_w y} = \min_{y \in \mathbb{R}^n : \sum_{i \in V} \deg_w(i) y(i) = 0} \frac{\sum_{ij \in E} w(ij)\big(y(i) - y(j)\big)^2}{\sum_{i \in V} \deg_w(i) \cdot y(i)^2}.$$

The weighted conductance of a subset is defined naturally as $\phi_w(S) := w(\delta(S))/\operatorname{vol}_w(S)$ where $w(\delta(S)) := \sum_{e \in \delta(S)} w(e)$ and $\operatorname{vol}_w(S) := \sum_{i \in S} \deg_w(i)$, and the weighted conductance of a graph is defined as $\phi_w(G) := \min_{S : \operatorname{vol}_w(S) \leq \frac{1}{2} \operatorname{vol}_w(V)} \phi(S)$. Choosing an appropriate binary solution involving $\operatorname{vol}_w(S)$, the easy direction can be shown similarly as in the d-regular case.

Exercise 4.11 (Easy Direction for General Weighted Graphs)**.** *Let $G = (V, E, w)$ be an edge weighted graph and λ_2 be the second smallest eigenvalue of $\mathcal{L}_w(G)$. Show that $\frac{1}{2}\lambda_2 \leq \phi_w(G)$.*

The main changes are actually in the easy direction. For the hard direction, the proofs are basically the same. The arguments in Lemma 4.6 and Corollary 4.7 work the same way. The analysis of the threshold rounding is very similar to that in Lemma 4.8, but on the formulation in Exercise 4.10.

Exercise 4.12 (Hard Direction for General Weighted Graphs)**.** *Let $G = (V, E, w)$ be an edge weighted graph and λ_2 be the second smallest eigenvalue of $\mathcal{L}_w(G)$. Prove that $\phi_w(G) \leq \sqrt{2\lambda_2}$.*

4.5 Problems

Problem 4.13 (Truncation)**.** *We outline the truncation step which does not require that the vector is an eigenvector. The following statements are for general weighted graphs. You may specialize the problem to the d-regular case.*

Let $G = (V, E)$ be a weighted undirected graph with $V = [n]$ and $y \in \mathbb{R}^n$ be a vector with $\sum_{i \in V} \deg_w(i) y(i) = 0$. Let $R_w(y) := y^T L_w y / y^T D_w y$ be the weighted Rayleigh quotient.

1. *Let c be a value such that $\operatorname{vol}_w(\{i \mid y(i) < c\}) \leq \frac{1}{2}\operatorname{vol}_w(V)$ and $\operatorname{vol}_w(\{i \mid y(i) > c\}) \leq \frac{1}{2}\operatorname{vol}_w(V)$. Let $z := y - c\vec{1}$. Prove that $R_w(z) \leq R_w(y)$. You may need to use the assumption that $\sum_{i \in v} \deg_w(i) y(i) = 0$.*

2. *Let $z \in \mathbb{R}^n$ be the vector obtained in the previous step. Let $z_+ \in \mathbb{R}^n$ be the vector with $z_+(i) := \max\{z(i), 0\}$ for $1 \leq i \leq n$, and $z_- \in \mathbb{R}^n$ be the vector with $z_-(i) := \min\{z(i), 0\}$ for $1 \leq i \leq n$. Prove that $\min\{R_w(z_+), R_w(z_-)\} \leq R_w(z)$.*

3. *Conclude with the suitable generalization of Corollary 4.7 that allows one to continue to prove the hard direction for general weighted graphs.*

4.6 References

[Alo86] Noga Alon. Eigenvalues and expanders. *Comb.*, 6(2):83–96, 1986. 25

[AM85] Noga Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985. 25

[Che70] J. Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems Anal.*, page 195, 1970. 25

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 26, 29, 57, 60, 61, 62, 63, 67, 68, 137

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.*, 82(1):93–133, 1989. 25

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000. 26

[Tre16] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. 2016. 30, 40

Chapter 5

Generalizations of Cheeger's Inequality

We give an overview of some recent generalizations of Cheeger's inequality using other eigenvalues. The presentation follows the chronological order.

5.1 Bipartiteness Ratio and Maximum Cut

Recall from Proposition 3.6 that a graph G is bipartite if and if the spectrum of its adjacency matrix is symmetric around the origin, and from Problem 3.7 that a connected graph G is bipartite if and only if $\alpha_1 = -\alpha_n$ where α_i is the i-th largest eigenvalue of the adjacency matrix. These results are for the spectrum of the adjacency matrix, and the following is a corresponding result for the spectrum of the normalized Laplacian matrix.

Exercise 5.1 (Spectral Characterization of Bipartiteness). *Let $G = (V, E)$ be an undirected graph and λ_n be the largest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(G)$. Then $\lambda_n = 2$ if and only if G has a bipartite component, i.e. a connected component that is a bipartite graph.*

Trevisan [Tre09] proved that λ_n is close to 2 if and only if G is close to having a bipartite component, in the same style as in Cheeger's inequality in Theorem 4.3. To state the result, we write the optimization formulation for $2 - \lambda_n$ and then motivate the corresponding definition of bipartiteness ratio of a subset of vertices S.

Exercise 5.2 (Optimization Formulation for $2 - \lambda_n$). *Let $G = (V, E)$ be an undirected graph and λ_n be the largest eigenvalue of $\mathcal{L}(G)$. Then*

$$2 - \lambda_n = \min_{x \in \mathbb{R}^n} \frac{\sum_{ij \in E} \big(x(i) + x(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot x(i)^2}.$$

Let $S \subseteq V(G)$ be a bipartite component with partition $S = (L, R)$ such that all the edges in S are between L and R. Then the vector $x \in \{-1, 0, 1\}^n$ where

$$x(i) = \begin{cases} +1 & \text{if } i \in L \\ -1 & \text{if } i \in R \\ 0 & \text{otherwise} \end{cases}.$$

is a solution to the above optimization problem with objective value 0. Using this association between a vector $x \in \{-1, 0, 1\}^n$ and a bipartition of a subset $S = (L, R)$, Trevisan considered the

following definition of the bipartiteness ratio of $S = (L, R)$ where each edge within L and each edge within R contributes 2 in the numerator while an edge in $\delta(S)$ contributes 1 in the numerator.

Definition 5.3 (Bipartiteness Ratio)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. The bipartiteness ratio of a vector $x \in \{-1, 0, 1\}^n$ is defined as*

$$\beta(x) := \frac{\sum_{ij \in E} \left|x(i) + x(j)\right|}{\sum_{i \in V} \deg(i) \cdot |x(i)|}.$$

The bipartiteness ratio of a graph G is defined as

$$\beta(G) := \min_{x \in \{-1,0,1\}^n} \beta(x).$$

Trevisan proved the following analog of Cheeger's inequality for $2 - \lambda_n$ and $\beta(G)$.

Theorem 5.4 (Cheeger's Inequality for λ_n [Tre09])**.** *Let $G = (V, E)$ be an undirected graph and λ_n be the largest eigenvalue of $\mathcal{L}(G)$. Then*

$$\frac{1}{2}(2 - \lambda_n) \leq \beta(G) \leq \sqrt{2(2 - \lambda_n)}.$$

The easy direction is left as an exercise and the hard direction is left as a problem. Trevisan used the ideas in his proof of Cheeger's inequality as shown in Section 4.3 to define $L_t = \{i \mid x(i) \geq \sqrt{t}\}$ and $R_t = \{i \mid x(i) \leq -\sqrt{t}\}$ and $S_t = (L_t, R_t)$ for a uniformly random t to prove the hard direction.

Maximum Cut

In the maximum cut problem, we are given an undirected graph $G = (V, E)$ and the task is to find a set $S \subseteq V$ that maximizes $|\delta(S)|$. This is a classical NP-complete problem. It is an exercise that there is always a subset $S \subseteq V$ with $|\delta(S)| \geq \frac{1}{2}|E|$, and this gives a trivial $\frac{1}{2}$-approximation algorithm for the problem. It was not known how to do better until Goemans and Williamson [GW95] introduced semidefinite programming into approximation algorithms and used it to design a 0.878-approximation algorithm for the maximum cut problem. Semidefinite programming was the only approach to do better than $\frac{1}{2}$-approximation for the maximum cut problem until Trevisan used Theorem 5.4 to design a spectral 0.531-approximation algorithm.

The power of the spectral method is that it gives a better upper bound on the optimal value than the trivial upper bound $|E|$. Suppose the maximum cut $(S, V - S)$ cuts at least $1 - \epsilon$ fraction of edges Then check that $\beta(G) \leq \epsilon$, and thus $2 - \lambda_n \leq 2\beta(G) \leq 2\epsilon$ by the easy direction of Theorem 5.4. So, if we compute that $2 - \lambda_n$ is large, then we know that the maximum cut only cuts at most $\frac{1}{2}\lambda_n$ fraction of edges, and thus the trivial approximation algorithm of cutting 50% of edges would be a $1/\lambda_n$-approximation algorithm. To summarize, when λ_n is bounded away from 2, then there is a better than $1/2$-approximation algorithm for the maximum cut problem, using the easy direction of Theorem 5.4.

On the other hand, when $2 - \lambda_n$ is small, by the hard direction of Theorem 5.4, we can find a subset $S = (L, R)$ with small $\beta(S)$, so that most edges with an endpoint in S are between L and R. So we commit on putting vertices in L on one side and vertices in R on the other side. Then we solve the maximum cut problem on $V - S$ recursively. Suppose the returned partition is $V - S = (L', R')$. Then we return the best of $(L' \cup L, R' \cup R)$ and $(L' \cup R, R' \cup L)$ as our solution, to ensure that at

least 50% of edges in $\delta(S)$ are cut. To summarize, when λ_n is close to 2, we can find a set S and ensure that more than 50% of the edges with an endpoint in S will be cut in the returned solution, using the hard direction of Theorem 5.4.

With these ideas, it should be clear that they can be combined to give a better than $1/2$-approximation algorithm for the maximum cut problem. Soto [Sot15] improved the analysis and gave a spectral 0.614-approximation algorithm for the maximum cut problem.

5.2 Small-Set Expansion

A more refined notion of expansion is to study the expansion of sets of different size. We assume the graph is d-regular in this section.

Definition 5.5 (Expansion Profile)**.** *Let $G = (V, E)$ be a d-regular graph. For any $0 < \delta \leq 1/2$, define*

$$\phi_\delta(G) := \min_{S \subseteq V : |S| \leq \delta |V|} \phi(S)$$

to be the δ-small-set expansion of G. The curve $\phi_\delta(G)$ for $0 < \delta \leq 1/2$ is called the expansion profile of the graph G.

The problem of finding small sparse cuts is useful in applications such as community detection in a social network. Also, this problem is of much theoretical interest because of its close connection to the unique games conjecture [RS10]. The small-set expansion conjecture by Steurer and Raghavendra [RS10] states that for any $\epsilon > 0$, there exists $0 < \delta \leq 1/2$ such that it is NP-hard to distinguish between $\phi_\delta(G) \leq \epsilon$ and $\phi_\delta(G) \geq 1 - \epsilon$. This conjecture is still wide open, and if true this would imply optimal inapproximability results for many well-known problems, including Goemans-Williamson 0.878-approximation algorithm for the maximum cut problem!

Motivated by this connection, Arora, Barak and Steurer [ABS10] proved the following Cheeger's inequality for small-set expansion, which roughly says that if λ_k is small for a large enough k, then there is a set S with $|S| \approx |V|/k$ and $\phi(S) \approx \sqrt{\lambda_k}$.

Theorem 5.6 (Cheeger's Inequality for Small-Set Expansion [ABS10])**.** *Let $G = (V, E)$ be a d-regular graph and λ_k be the k-th smallest eigenvalue of $\mathcal{L}(G)$. For $k \geq n^{2\beta}$,*

$$\exists S \subseteq V \text{ with } |S| \lesssim n^{1-\beta} \text{ and } \phi(S) \lesssim \sqrt{\frac{\lambda_k}{\beta}}.$$

They used this theorem to design a sub-exponential time algorithm for the small-set expansion conjecture and the unique games conjecture, together with the ideas of subspace enumeration and graph decomposition. This is an influential paper as it opens up the line of research about higher eigenvalues of graphs.

Analytically Sparse Vectors from Random Walks

Using the threshold rounding in Lemma 4.8, if we could find a vector x with $|\operatorname{supp}(x)| \leq \delta |V|$ and $R_{\mathcal{L}}(x) \lesssim \lambda_k$, then we can find a set S with $|S| \leq \delta |V|$ and $\phi(S) \lesssim \sqrt{\lambda_k}$. This is the starting point.

The constraint $|\operatorname{supp}(x)| \leq \delta |V|$ is combinatorial and not easy to work with directly. Note that any vector satisfying this constraint satisfies the condition $\|x\|_1 \leq \sqrt{\delta |V|} \cdot \|x\|_2$ by Cauchy-Schwarz, an analytical condition more suitable for spectral analysis.

Definition 5.7 (Combinatorial and Analytical Sparse Vectors)**.** *Let $x \in \mathbb{R}^n$ be a vector and $\delta \in [0,1]$. We say x is δ-combinatorially sparse if $|\mathrm{supp}(x)| \leq \delta n$, and x is δ-analytically sparse if $\|x\|_1 \leq \sqrt{\delta n}\|x\|_2$.*

By a truncation argument similar to that in Problem 4.13, we can reduce the problem to finding a δ-analytically sparse vector with small Rayleigh quotient.

Problem 5.8 (Combinatorial Sparse Vector from Analytical Sparse Vector)**.** *Let $x \in \mathbb{R}^n_+$ be a non-negative vector that is δ-analytically sparse. Prove that there exists a non-negative vector $y \in \mathbb{R}^n_+$ that is 4δ-combinatorially sparse with $R_{\mathcal{L}}(y) \leq 2R_{\mathcal{L}}(x)$.*

The main idea in [ABS10] is to find such a vector from random walks, a topic that we will study in the next chapter. Let $W = \frac{1}{2}I + \frac{1}{2}\mathcal{A} = I - \frac{1}{2}\mathcal{L}$ be the lazy random walk matrix. Note that our assumption $\lambda_k \leq \lambda$ translates to $\alpha_k \geq 1 - \frac{\lambda}{2}$ where α_k denotes the k-th largest eigenvalue of W. The main argument using the spectrum is

$$\sum_{i=1}^n \|W^t\chi_i\|_2^2 = \sum_{i=1}^n \chi_i^T W^{2t}\chi_i = \mathrm{Tr}(W^{2t}) = \sum_{i=1}^n \alpha_i^{2t} \geq k\Big(1-\frac{\lambda}{2}\Big)^{2t},$$

where the last equality is by Fact 2.35. Therefore, there exists $i \in [n]$ such that

$$\|W^t\chi_i\|_2^2 \geq \frac{k}{n}\Big(1-\frac{\lambda}{2}\Big)^{2t},$$

and this gives an analytically sparse vector as $\|W^t\chi_i\|_1 = 1$ since it is a probability distribution.

On the other hand, by a relatively standard spectral argument using eigen-decomposition and the power mean inequality, one can prove that the Rayleigh quotient $R(W^t\chi_i)$ is small for every $i \in [n]$.

Problem 5.9 (Rayleigh Quotient of Random Walk Vector)**.** *Let $G = (V,E)$ be a graph with $V = [n]$, $\mathcal{L}$ be its normalized Laplacian matrix, and $W = I - \frac{1}{2}\mathcal{L}$ be its lazy random walk matrix. For any $i \in [n]$,*

$$R_{\mathcal{L}}(W^t\chi_i) \leq 2 - 2\|W^t\chi_i\|_2^{1/t}.$$

These two claims combine to give a vector $W^t\chi_i$ that has small Rayleigh quotient and is analytically sparse. More precisely, by setting $t = \frac{\ln k}{2\lambda}$ and doing some calculations, one can check that there exists i with $W^t\chi_i$ being $\frac{1}{\sqrt{k}}$-analytically sparse and $R(W^t\chi_i) \leq \frac{2\lambda \ln n}{\ln k}$. Then Theorem 5.6 follows when $k \geq n^{2\beta}$.

5.3 Higher-Order Cheeger Inequalities

Recall from Exercise 3.19 that $\lambda_k = 0$ if and only if G has at least k connected components. After seeing Cheeger's inequality in Theorem 4.3 and its analogy for λ_n in Theorem 5.4, we now expect that there is also a robust quantitative generalization of this fact.

Actually, the Cheeger inequality for small-set expansion in Theorem 5.6 can be seen as one such generalization, because when $\lambda_k = 0$ there exists a component of size at most n/k, and Theorem 5.6 proves that there exists a sparse cut of size roughly n/k when k is large enough.

In the following, we see another generalization that λ_k is small if and only if G has at least k disjoint subsets $S_1, \ldots, S_k$ each is close to a connected component.

Definition 5.10 (k-Way Edge Conductance)**.** *Let $G = (V, E)$ be a graph. The k-way edge conductance is defined as*

$$\phi_k(G) = \min_{S_1, S_2, \ldots, S_k \subseteq V} \max_{1 \le i \le k} \phi(S_i),$$

where the minimization is over pairwise disjoint subsets $S_1, \ldots, S_k$ of V.

The following higher-order Cheeger inequalities were obtained independently by two research groups.

Theorem 5.11 (Higher-Order Cheeger Inequalities [LOT14, LRTV12])**.** *Let $G = (V, E)$ be a graph and λ_k be the k-th smallest eigenvalue of its normalized Laplacian matrix. Then*

$$\frac{1}{2}\lambda_k \le \phi_k(G) \lesssim k^2 \cdot \sqrt{\lambda_k}.$$

Moreover,

$$\phi_k(G) \lesssim \sqrt{\log k \cdot \lambda_{2k}}.$$

Note that Theorem 5.11 guarantees that there are disjoint sparse cuts and hence also a sparse cut of size at most n/k, but Theorem 5.6 gives a stronger quantitative bound with no dependency on k which is crucial for the small-set expansion and the unique games conjectures. In short, Theorem 5.11 and Theorem 5.6 are incomparable, and it would be very interesting to obtain a common generalization of these two results. The following is another open question.

Question 5.12. *Is it true that $\phi_k(G) \lesssim \mathrm{polylog}(k) \cdot \sqrt{\lambda_k}$?*

Spectral Embedding

The high level plan is to find k disjoint supported vectors $x_1, \ldots, x_k$ such that each has small Rayleigh quotient $R_{\mathcal{L}}(x_i)$. Then we can apply the threshold rounding in Lemma 4.8 on each x_i to find $S_i \subseteq \mathrm{supp}(x_i)$ with small conductance $\phi(S_i) \lesssim \sqrt{R_{\mathcal{L}}(x_i)}$.

An interesting new idea in [LOT14, LRTV12] is to use the spectral embedding defined by the first k eigenvectors to find the k disjoint sparse cuts.

Definition 5.13 (Spectral Embedding)**.** *Let $G = (V, E)$ be a graph with $V = [n]$, $\lambda_1 \le \ldots \le \lambda_k$ be the k smallest eigenvalues of $\mathcal{L}(G)$, and $v_1, \ldots, v_k \in \mathbb{R}^n$ be corresponding orthonormal eigenvectors. Let $U \in \mathbb{R}^{n \times k}$ be the $n \times k$ matrix where the j-th column is v_j. The spectral embedding $u_i \in \mathbb{R}^k$ of a vertex i is defined as the i-th row of U.*

The spectral embedding is used in practice to find disjoint sparse cuts. A popular heuristic is to apply some well-known geometric clustering algorithms, in particular the k-means algorithm, to partition the point set in the spectral embedding into k groups/clusters, and use this partitioning to cut the graph into k sets. It is still an open problem to analyze this heuristic rigorously.

The proof in [LOT14] analyzed a slightly different algorithm that clusters the points based on directions. As $v_1, \ldots, v_k$ are orthonormal vectors, the matrix U in Definition 5.13 satisfies $U^T U = I_k$, and this implies that the spectral embedding satisfies the following isotropy condition.

Exercise 5.14 (Isotropy Condition)**.** *Let $u_1, \ldots, u_n \in \mathbb{R}^k$ be the spectral embedding of the vertices in Definition 5.13. For any $x \in \mathbb{R}^k$ with $\|x\|_2 = 1$, prove that*

$$\sum_{i=1}^{n} \langle x, u_i \rangle^2 = 1.$$

Note that $\sum_{i=1}^n \|u_i\|^2 = k$ as $U^T U = I_k$. The isotropy condition implies that the points $u_1, \ldots, u_n \in \mathbb{R}^k$ must be "well spread out" in different directions.

Definition 5.15 (Radial Projection Distance)**.** *Let $u_1, \ldots, u_n \in \mathbb{R}^k$ be the spectral embedding of the vertices in Definition 5.13. The radial projection distance between two vertices i and j is defined as*

$$d(i,j) = \left\| \frac{u_i}{\|u_i\|_2} - \frac{u_j}{\|u_j\|_2} \right\|$$

if $\|u_i\| > 0$ and $\|u_j\| > 0$. Otherwise, if $u_i = u_j = 0$ then $d(i,j) := 0$, else $d(i,j) := \infty$.

Problem 5.16 (Spreading Property)**.** *Let $G = (V, E)$ be a graph with $V = [n]$. Let $u_1, \ldots, u_n \in \mathbb{R}^k$ be the spectral embedding of the vertices in Definition 5.13. Let $S \subseteq V$ be such that $d(i,j) \leq \Delta$ for all $i, j \in S$. Then*

$$\sum_{i \in S} \|u_i\|^2 \leq \frac{1}{1 - \Delta^2}.$$

Informally, the spreading property implies that the points cannot be concentrated in less than k directions, as otherwise the spectral embedding cannot identify k clusters.

Suppose there are k clusters $S_1, \ldots, S_k$ such that $\sum_{i \in S_j} \|u_i\|^2 \approx 1$ for $1 \leq i \leq k$ and the pairwise distance $d(S_i, S_j) := \min_{a \in S_i, b \in S_j} d(a,b)$ is large. Then [LOT14] uses an idea called smooth localization to find k disjoint supported vectors $x_1, \ldots, x_k \in \mathbb{R}^n$ each with small Rayleigh quotient.

To achieve this condition, [LOT14] also uses a random partitioning idea to partition $\mathbb{R}^k$ and removes all points close to boundaries so that the distances between different parts are lower bounded. For more details, see L04 in 2019 or the notes by Trevisan [Tre16] or the journal paper [LOT14].

Randomized Rounding Algorithm

The algorithm in [LRTV12] is elegant and simple to describe.

Algorithm 2 Randomized Rounding on Spectral Embedding [LRTV12]

Require: An undirected graph $G = (V, E)$ with $V = [n]$ and $m = |E|$, and a parameter k.

1: Compute the spectral embedding $u_1, \ldots, u_n \in \mathbb{R}^k$ in Definition 5.13.

2: Pick k independent Gaussian vectors $g_1, \ldots, g_k \in N(0,1)^n$. Construct disjointly supported vectors $h_1, \ldots, h_k \in \mathbb{R}^n$ as follows:

$$h_i(j) = \begin{cases} \langle u_j, g_i \rangle & \text{if } i = \text{argmax}_{i \in [k]} \{\langle u_j, g_i \rangle\} \\ 0 & \text{otherwise} \end{cases}.$$

3: Apply the threshold rounding in Lemma 4.8 on each h_i to obtain a set $S_i \subseteq \text{supp}(h_i)$ and $\phi(S_i) \leq \sqrt{2 R_{\mathcal{L}}(h_i)}$.

4: **return** all S_i with $\phi(S_i) \lesssim \sqrt{\log k \cdot \lambda_k}$.

It is proved in [LRTV12] that this algorithm will return $\Omega(k)$ subsets with constant probability. The proof is by computing the expectation and the variance of the numerator and the denominator, using some properties of Gaussian random variables.

5.4 Improved Cheeger Inequalities

Recall that it is an open question to explain the empirical performance of the spectral partitioning algorithm. In practical instances for image segmentation and data clustering, it is reasonable to expect that there are only a few outstanding objects/clusters in the input image/dataset. One way to formalize this is to assume that the k-way conductance $\phi_k(G)$ of the input graph is large for a small constant k. By the higher-order Cheeger inequality in Theorem 5.11, this is qualitatively equivalent to λ_k is large for a small constant k, which is typically satisfied in practical instances of image segmentation. The following improved Cheeger's inequality shows that the spectral partitioning algorithm performs better in these inputs.

Theorem 5.17 (Improved Cheeger's Inequality [KLL+13])**.** *Let $G = (V, E)$ be a graph and λ_k be the k-th smallest eigenvalue of its normalized Laplacian matrix. For any $k \geq 2$,*

$$\frac{\lambda_2}{2} \leq \phi(G) \lesssim \frac{k\lambda_2}{\sqrt{\lambda_k}}.$$

The proof of Theorem 5.17 shows that the spectral partitioning algorithm achieves this guarantee. Note that when $\lambda_k = \Omega(1)$ for a small constant k, Theorem 5.17 implies that the spectral partitioning algorithm is a constant factor approximation algorithm for graph conductance. This provides some rigorous justification of the empirical success of the spectral partitioning algorithm.

Exercise 5.18 (Tight Example for Theorem 5.17)**.** *Check that the improved Cheeger's inequality is tight up to a constant factor for the cycle examples.*

There are also related improved Cheeger's inequalities which work with $\phi_k(G)$ directly and with the robust vertex expansion of the graph [KLL17].

k-Step Functions

To see the main intuition in [KLL+13], consider the simpler scenario when λ_2 is small but λ_3 is big. Since λ_2 is small, the graph has a sparse cut $(S, V - S)$ by Cheeger's inequality in Theorem 4.3. As λ_3 is big, $\phi_3(G)$ is also big by the higher-order Cheeger's inequality in Theorem 5.11. This implies that the induced graph in each S and $V - S$ should be an expander graph, as otherwise there is a good way to cut them into smaller pieces which would contradict that $\phi_3(G)$ is big. Since the induced graphs in S and $V - S$ are expander graphs and $(S, V - S)$ is a sparse cut, we expect that the minimizer for the Rayleigh quotient in Lemma 4.4 should look like a binary solution and thus $\lambda_2 \approx \phi(G)$.

The proof of Theorem 5.17 has two main steps. The first step is to show that if λ_k is large for a small constant k, then any eigenvector of the second eigenvalue should look like a k-step function.

Definition 5.19 (k-Step Function)**.** *Let $G = (V, E)$ be a graph with $V = [n]$. Given $y \in \mathbb{R}^n$ and $1 \leq k \leq n$, we say y is a k-step function if the number of distinct values in $\{y(i)\}_{i \in V}$ is at most k.*

Lemma 5.20 (Constructing k-Step Approximation)**.** *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. For any $x \in \mathbb{R}^n$, there is a $(2k + 1)$-step function y such that*

$$\|x - y\|_2^2 \leq \frac{4\|x\|_2^2 \cdot R_{\mathcal{L}}(x)}{\lambda_k}.$$

The second step is to show that if the second eigenvector is close to a k-step function, then the spectral partitioning algorithm performs better.

Lemma 5.21 (Rounding k-Step Approximation). *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. Let $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$ and let $y \in \mathbb{R}^n$ be a $(2k+1)$-step function. The spectral partitioning algorithm applied on x outputs a set S with $|S| \leq n/2$ and*

$$\phi(S) \leq 4kR_{\mathcal{L}}(x) + 4\sqrt{2}k \cdot \|x - y\|_2 \cdot \sqrt{R_{\mathcal{L}}(x)}.$$

Note that Theorem 5.17 follows immediately from Lemma 5.20 and Lemma 5.21.

To prove Lemma 5.20, the idea is that if x is far from being a k-step function, then x must be "smooth/continuous", and it is possible to decompose x into k disjoint supported vectors $x_1, \ldots, x_k \in \mathbb{R}^n$ such that each has small Rayleigh quotient as shown in the following figure.

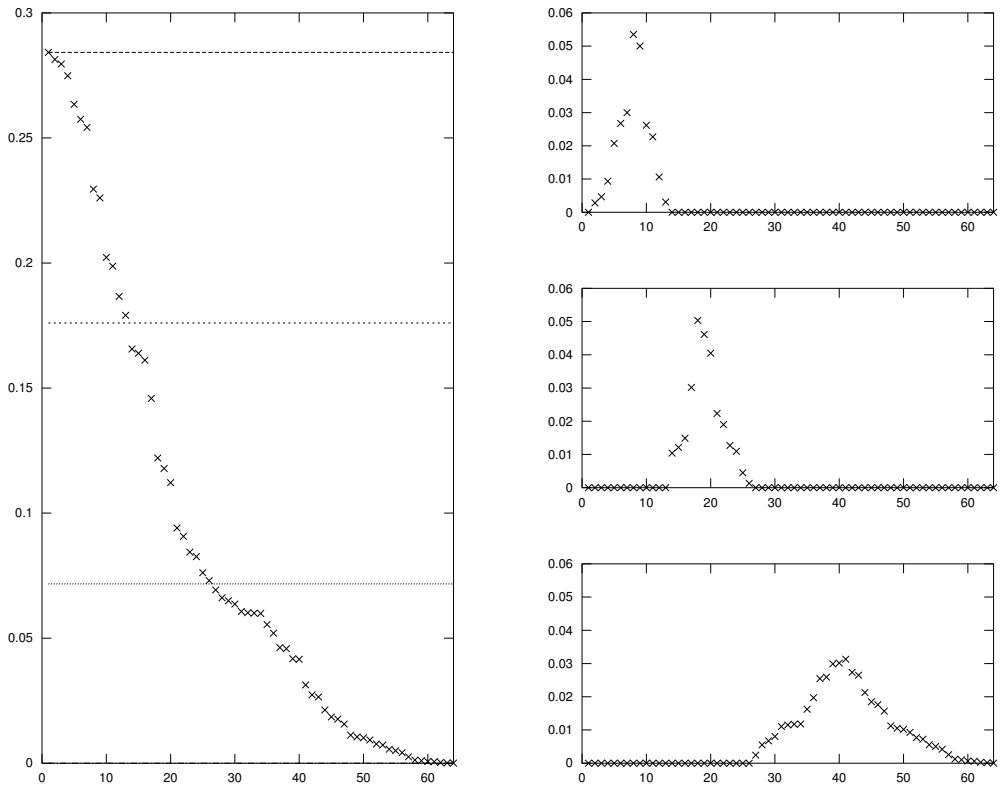


Figure 5.1: The figure on the left is the function x. We cut x into three disjointly supported vectors x_1, x_2, x_3 by setting $t_0 = 0$, $t_1 \approx 0.07$, $t_2 \approx 0.175$, and $t_3 = \max x(i)$. For each $1 \leq i \leq 3$, we define $x_i(j) = \min\{|x(j) - t_{i-1}|, |x(j) - t_i|\}$, if $t_{i-1} \leq x(j) \leq t_i$, and zero otherwise.

For Lemma 5.21, it is instructive to work out the special case when x is exactly a k-step function.

Problem 5.22 (Rounding k-Step Function). *Prove Lemma 5.21 when x is a $(2k+1)$-step function.*

The general idea is to choose a random threshold t with probability proportional to the distance to the nearest step in y. See L05 in 2019 or [KLL+13] for details.

5.5 References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 563–572. IEEE Computer Society, 2010. 37, 38

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. 36

[KLL+13] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 11–20. ACM, 2013. 41, 42

[KLL17] Tsz Chiu Kwok, Lap Chi Lau, and Yin Tat Lee. Improved cheeger's inequality and analysis of local graph partitioning using vertex expansion and expansion profile. *SIAM J. Comput.*, 46(3):890–910, 2017. 41

[LOT14] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6):37:1–37:30, 2014. 39, 40

[LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh S. Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 1131–1140. ACM, 2012. 39, 40

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 755–764. ACM, 2010. 37, 76

[Sot15] José A. Soto. Improved analysis of a max-cut algorithm based on spectral partitioning. *SIAM J. Discret. Math.*, 29(1):259–268, 2015. 37

[Tre09] Luca Trevisan. Max cut and the smallest eigenvalue. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 263–272. ACM, 2009. 35, 36

[Tre16] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. 2016. 30, 40

Chapter 6

Random Walks on Graphs

Given an undirected graph $G = (V, E)$, a random walk is a simple stochastic process where it starts from a vertex, and in each step the walk moves to a uniformly random neighbor of the current vertex. Some of the basic mathematical questions about random walks are:

1. (Stationary Distribution:) Is there a limiting distribution of the random walks?
2. (Mixing Time:) How long does it take for the current distribution to be close to the limiting distribution?
3. (Hitting Time:) Starting from a vertex s, how long does it take to first reach a vertex t?
4. (Cover Time:) How long does it take to reach every vertex in the graph at least once?

There are two main approaches to questions (1) and (2). One is probabilistic and uses the idea of "coupling" two random processes. Another is spectral and uses the eigenvalues of the transition matrix. We study the spectral approach in this chapter and refer the reader to [Häg02, LPW06] for the coupling approach.

Questions (3) and (4) are best answered by viewing the graph as an electrical network. This is a topic that we used to study but probably not in this offering. We refer the reader to [DS84, AF02, LP16, Spi19] for the interesting connection between random walks and electrical networks.

6.1 Markov Chains

In this section, we consider the more general setting of a finite Markov chain and state the fundamental theorem.

A finite Markov chain is defined by a finite state space and a transition matrix.

Definition 6.1 (Transition Matrix)**.** *Let $[n]$ be the state space. A matrix $P \in \mathbb{R}^{n\times n}$ is a probability transition matrix if P is non-negative and $\sum_{j\in[n]} P_{i,j} = 1$ for each $i \in [n]$. For $1 \le i, j \le n$, the entry P_{ij} is the transition probability from state i to state j.*

Definition 6.2 (Markov Chain)**.** *A sequence of random variables $(X_0, X_1, \ldots)$ is a Markov chain with state space $[n]$ and transition matrix $P \in \mathbb{R}^{n\times n}$ if for all $i, j \in [n]$ and $t \ge 1$,*

$$\Pr\left[X_{t+1} = j \mid X_t = i \cap X_{t-1} = i_{t-1} \cap \ldots \cap X_0 = i_0\right] = \Pr\left[X_{t+1} = j \mid X_t = i\right] = P(i, j).$$

The first equality is called the Markov property, which states that the transition probability from i to j is the same regardless of the states $X_0, \ldots, X_{t-1}$ that precedes the current state X_t.

We can simply think of a Markov chain as a random walk on a weighted directed graph $G = ([n], w)$, where the transition probability from state i to state j is proportional to the edge weight $w(i, j)$ such that $P_{i,j} = w(i, j)/\sum_{j \in [n]} w(i, j)$.

Irreducibility and Aperiodicity

The following are two properties that will imply the existence of a unique limiting distribution.

Definition 6.3 (Irreducibility)**.** *A Markov chain defined by transition matrix $P \in \mathbb{R}^{n \times n}$ is called irreducible if for any two states i, j, there exists an integer t such that $\Pr[X_t = j \mid X_0 = i] > 0$.*

An equivalent definition is that the underlying directed graph $G = ([n], E)$ where $E(G) := \{ij \mid P_{i,j} > 0\}$ of P is strongly connected.

This property is called irreducibility, because if it is not satisfied then the Markov chain is reducible to a smaller one for the study of the limiting distribution, as the limiting distribution if exists will only have support on a strongly connected component.

Definition 6.4 (Aperiodicity)**.** *The period of a state i is defined as $\gcd\{t \mid \Pr[X_t = i \mid X_0 = i] > 0\}$, the greatest common divisor of the set of times when it is possible to return to the starting state i. A state i is called aperiodic if its period is equal to 1. A Markov chain is called aperiodic if all states are aperiodic; otherwise it is called periodic.*

For examples, random walks on an undirected bipartite graph is periodic as every state has period 2, and random walks on a directed cycle of length $k > 1$ is periodic as every state has period k. In general, there is no limiting distribution if the Markov chain is periodic.

Irreducibility and aperiodicity together imply the following property that after enough number of steps, the probability to transit from any state to any other state is positive.

Proposition 6.5 (Reachability)**.** *For any finite, irreducible, and aperiodic Markov chain, there exists an integer $\tau < \infty$ such that $\Pr[X_t = j \mid X_0 = i] > 0$ for all i, j and all $t \geq \tau$.*

The proof uses aperiodicity and a simple number-theoretic argument to prove the statement for all $i = j$, and then uses irreducibility to prove the statement for all $i \neq j$. See [Häg02, LPW06] for a proof.

Stationary Distribution and Convergence

Definition 6.6 (Stationary Distribution)**.** *For a Markov chain defined by transition matrix $P \in \mathbb{R}^{n \times n}$, a probability distribution $\vec{\pi} \in \mathbb{R}^n$ is a stationary distribution if $\vec{\pi}P = \vec{\pi}$ when $\vec{\pi}$ is represented as a row vector.*

Informally, a stationary distribution $\vec{\pi}$ is a steady distribution as $\vec{\pi}P^t = \vec{\pi}$ for any $t \geq 1$, and a limiting distribution if exists is a stationary distribution.

We measure how close are two probability distributions by the total variation distance.

Definition 6.7 (Total Variation Distance)**.** *Given two probability distributions $\vec{p}, \vec{q} \in \mathbb{R}^n$, the total variation distance of $\vec{p}$ and $\vec{q}$ is defined as*

$$d_{\mathrm{TV}}(\vec{p}, \vec{q}) := \frac{1}{2} \sum_{i=1}^{n} |p(i) - q(i)| = \frac{1}{2} \|\vec{p} - \vec{q}\|_1.$$

Consider the Markov chain defined by a transition matrix $P \in \mathbb{R}^{n \times n}$. Let $\vec{p}_0 \in \mathbb{R}^n$ be an initial probability distribution on the states. Let $\vec{p}_t \in \mathbb{R}^n$ be the probability distribution $\vec{p}_t := \vec{p}_0 P^t$ after t steps of random walks. We say that $\vec{p}_t$ converges to a probability distribution $\vec{q}$ as $t \to \infty$ if $\lim_{t \to \infty} d_{\mathrm{TV}}(\vec{p}_t, \vec{q}) = 0$.

Fundamental Theorem of Markov Chains

One more definition before we state the fundamental theorem.

Definition 6.8 (Hitting Time)**.** *For a Markov chain defined by transition matrix $P \in \mathbb{R}^{n \times n}$, the hitting time from state i to state j is defined as*

$$H_{i,j} := \min\{t \geq 0 \mid X_t = j, X_0 = i\}.$$

The first return time to state i is defined as $H^+_{i,i} := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$.

Given any finite, irreducible, and aperiodic Markov chain defined by P, after we run it long enough, then it is possible to reach any state from any other state by Proposition 6.5. If two Markov chains $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ of P meet at the same state at some time t such that $X_t = Y_t$, then we cannot distinguish the probability distributions of X_τ and Y_τ for $\tau > t$ anymore as the Markov chains forget about the history. By Proposition 6.5, any two Markov chains of P will eventually meet, and so they will converge to the same distribution as $t \to \infty$, and thus a unique limiting distribution exists. This is the intuition of the coupling proof of the following fundamental theorem.

Theorem 6.9 (Fundamental Theorem of Markov Chains)**.** *Consider the Markov chain defined by a transition matrix $P \in \mathbb{R}^{n \times n}$. Let $\vec{p}_0 \in \mathbb{R}^n$ be an initial probability distribution on the states. Let $\vec{p}_t \in \mathbb{R}^n$ be the probability distribution $\vec{p}_t := \vec{p}_0 P^t$ after t steps. If the Markov chain is finite, irreducible, and aperiodic, then the followings hold.*

1. *There exists a stationary distribution $\vec{\pi}$.*
2. *The distribution $\vec{p}_t$ converges to $\vec{\pi}$ as $t \to \infty$, no matter what is the initial distribution $\vec{p}_0$.*
3. *There is a unique stationary distribution.*
4. *$\pi(i) = \lim_{t \to \infty} (P^t)_{i,i} = (\mathbb{E}[H^+_{i,i}])^{-1}$, the inverse of the expected first return time to i.*

We will see a spectral proof of this theorem in the special case of random walks on undirected graphs. For the general result, see [Häg02] for a probabilistic proof using coupling, [LPW06] for a probabilistic and algebraic proof, and [HJ13] for a purely algebraic proof related to the Perron-Frobenius Theorem 2.16.

6.2 Random Walks on Undirected Graphs

We consider random walks on an unweighted undirected graph $G = (V, E)$ with $V = [n]$, where in each step the walk moves to a uniformly random neighbor of the current vertex. The fundamental theorem becomes simpler and easier in this special case, as there are simple characterizations of irreducibility and aperiodicity and also the limiting distribution.

Matrix Formulation: The transition probability P_{ij} from a vertex i to a vertex j is simply $1/\deg(i)$, and so the transition matrix is $P = D^{-1}A$ where D is the diagonal degree matrix in Definition 3.13 and A is the adjacency matrix in Definition 3.1. Let $p_0 : V \to \mathbb{R}$ be an initial probability distribution, and p_t be the probability distribution after t steps of random walks. By definition, $p_{t+1}(i) := \sum_{j:ij\in E} p_t(j)/\deg(j)$ for all $1 \le i \le n$. Note that these equations can be written compactly as $p_{t+1} = P^T p_t = (AD^{-1})p_t$, and by induction $p_t = (AD^{-1})^t p_0$. We remark that it is common to write a probability distribution as a row vector, but in these notes we use the convention that p_t is a column vector.

Stationary Distribution: Recall that a probability distribution $\pi : V \to \mathbb{R}$ is a stationary distribution of P if $P^T\pi = \pi$. It is equivalent to saying that π is an eigenvector of P^T with eigenvalue 1. Given that $P = D^{-1}A$ for random walks on undirected graphs, it is not difficult to identify one such eigenvector.

Lemma 6.10 (Stationary Distribution of Undirected Graphs)**.** *Let $G = (V, E)$ be an undirected graph and $P = D^{-1}A$ be its transition matrix. The distribution $\pi : V \to \mathbb{R}$ with*

$$\pi(i) = \frac{\deg(i)}{\sum_{j\in V}\deg(j)} = \frac{\deg(i)}{2|E|}$$

for all $i \in V$ is a stationary distribution of P.

Irreducibility: Is π in Lemma 6.10 the unique stationary distribution? Not necessarily. For example, if the graph is disconnected, then the distribution after many steps depends on the initial distribution (e.g. which component does the starting vertex belongs to). This corresponds to the irreducibility condition in the fundamental theorem. For undirected graphs, the irreducibility condition is simply equivalent to the condition that the graph is connected.

Aperiodicity: Even if the graph is connected, a limiting distribution may not exist. For example, in a connected bipartite graph, if the initial distribution p_0 is on a single vertex, then the distribution p_t depends on the parity of t, as the support of p_t oscillates between the two sides of the bipartite graph. This corresponds to the aperiodicity condition in the fundamental theorem. For connected undirected graphs, observe that the aperiodicity condition is equivalent to the condition that the graph is non-bipartite.

Fundamental Theorem: Given the simple characterizations of the conditions in the fundamental theorem, it reduces to the following statement for undirected graphs.

Theorem 6.11 (Fundamental Theorem for Undirected Graphs)**.** *Let G be a connected, non-bipartite undirected graph. Let $P = D^{-1}A$ be the transition matrix of random walks on G. The distribution π in Lemma 6.10 is the unique stationary distribution. Furthermore, $p_t := (P^T)^t p_0$ converges to π as $t \to \infty$ regardless of the initial distribution p_0.*

Lazy Random Walks: The non-bipartiteness condition is to ensure that the Markov chain is aperiodic. There is a simple modification of the random walks so that this assumption can be removed, by adding self-loops in the graph.

Definition 6.12 (Lazy Random Walks)**.** *Let G be an undirected graph. The transition matrix W of the lazy random walks is defined as $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$. In words, the lazy random walks stay at the current vertex with probability $1/2$ and moves to a uniform random neighbor of the current vertex with probability $1/2$.*

By doing the lazy random walks, we make the graph non-bipartite and obtain the following corollary of Theorem 6.11.

Corollary 6.13 (Fundamental Theorem for Lazy Undirected Graphs)**.** *Let G be a connected undirected graph. Let $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$ be the transition matrix of lazy random walks on G. The distribution π in Lemma 6.10 is the unique stationary distribution. Furthermore, $p_t := (W^T)^t p_0$ converges to π as $t \to \infty$ regardless of the initial distribution p_0.*

It will be clear from the spectral analysis why the constant $1/2$ is used.

6.3 Spectral Analysis of Mixing Time for Undirected Graphs

In this section, we will prove the fundamental theorem for undirected graphs in Theorem 6.11 using a spectral analysis. Besides that the spectral analysis is elegant, it has the very important feature that it can also be used to analyze the mixing time, which is the rate of convergence to the unique stationary distribution.

As in Chapter 4 for Cheeger's inequality, we will first assume the given undirected graph is d-regular and prove Theorem 6.11 and then define and bound the mixing time. After that, we outline the modifications needed for general undirected graphs.

Spectrum for Regular Graphs

For d-regular graphs, the transition matrix P for random walks is simply $P = D^{-1}A = \frac{1}{d}A = \mathcal{A}$, the normalized adjacency matrix. And the transition matrix W for lazy random walks is $W = \frac{1}{2}I + \frac{1}{2}\mathcal{A}$. This is the main simplification from the d-regular assumption, as the matrices P and W are still real symmetric. Another simplification is that for d-regular graphs the stationary distribution π in Lemma 6.10 is simply the uniform distribution $\vec{1}/n$.

Our goal is to prove that $\lim_{t\to\infty} P^t p_0 = \vec{1}/n$ regardless of the initial distribution p_0, as long as the graph is connected and non-bipartite. And, similarly, $\lim_{t\to\infty} W^t p_0 = \vec{1}/n$, as long as the graph is connected.

To compute $P^t p_0$ and $W^t p_0$, a repeated application of the same operator, it is very helpful to know the spectrum of the matrices P and W as discussed in Chapter 2. Let $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be the eigenvalues of $\mathcal{A}$ and $v_1, \ldots, v_n$ be the corresponding orthonormal eigenvectors. Recall that

- $\alpha_1 = 1$ and $v_1 = \vec{1}/\sqrt{n}$ from Lemma 3.21,
- $\alpha_2 < 1$ if and only if G is connected from Proposition 3.18,
- and when G is connected, $\alpha_n > -1$ if and only if G is non-bipartite from Problem 3.7.

For the lazy random walk matrix W, the spectrum is $\frac{1}{2}(1+\alpha_1) \geq \frac{1}{2}(1+\alpha_2) \geq \ldots \geq \frac{1}{2}(1+\alpha_n)$, which implies that the smallest eigenvalue is at least $0 > -1$. This is why the non-bipartitness assumption can be removed when we consider lazy random walks.

Limiting Distribution

We translated the combinatorial conditions in the fundamental theorem into spectral conditions, and the fundamental Theorem 6.11 for d-regular undirected graphs can be restated as follows and the proof is relatively straightforward.

Proposition 6.14 (Limiting Distribution for Regular Graphs). *Let $G = (V, E)$ be a d-regular undirected graph with $V = [n]$. Let $P = \mathcal{A}$ be the transition matrix of random walks on G and $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be its eigenvalues. If $\alpha_2 < 1$ and $\alpha_n > -1$, then $\lim_{t\to\infty} P^t p_0 = \vec{1}/n$.*

Proof. Let $v_1, v_2, \ldots, v_n$ be the corresponding orthonormal eigenvectors. For any initial distribution p_0, as $v_1, \ldots, v_n$ forms an orthonormal basis, we can write $p_0 = c_1 v_1 + \ldots + c_n v_n$ as a linear combination of $v_1, \ldots, v_n$, where $c_i = \langle p_0, v_i \rangle$ for $1 \leq i \leq n$. Then,

$$P^t p_0 = \mathcal{A}^t\Big(\sum_{i=1}^n c_i v_i\Big) = \sum_{i=1}^n c_i \mathcal{A}^t v_i = \sum_{i=1}^n c_i \alpha_i^t v_i.$$

The assumptions $\alpha_2 < 1$ and $\alpha_n > -1$ imply that $|\alpha_i| < 1$ for $2 \leq i \leq n$. Hence,

$$\lim_{t\to\infty} P^t p_0 = \lim_{t\to\infty} \sum_{i=1}^n c_i \alpha_i^t v_i = c_1 v_1,$$

as all but the first term go to zero as $t \to \infty$. Recall that in the d-regular case, $v_1 = \vec{1}/\sqrt{n}$ and thus $c_1 = \langle p_0, \vec{1}/\sqrt{n} \rangle = 1/\sqrt{n}$ as p_0 is a probability distribution. Therefore, we conclude that

$$\lim_{t\to\infty} P^t p_0 = c_1 v_1 = \frac{1}{\sqrt{n}} \cdot \frac{\vec{1}}{\sqrt{n}} = \frac{\vec{1}}{n}.$$

□

Basically, the proof says that under the assumption $|\alpha_i| < 1$ for $2 \leq i \leq n$, $P^t p_0$ converges to the first eigenvector which is proportional to the all-one vector.

Check that the fundamental theorem for lazy random walks in Corollary 6.13 hold for d-regular graphs as well.

Mixing Time

The mixing time is to quantify how fast p_t converges to the limiting distribution. The following definition is for general Markov chains.

Definition 6.15 (Mixing Time)**.** *Consider the Markov chain defined by a transition matrix $P \in \mathbb{R}^{n\times n}$. Let $p_0 \in \mathbb{R}^n$ be an initial probability distribution on the states. Let $p_t \in \mathbb{R}^n$ be the probability distribution $p_t := (P^T)^t p_0$ after t steps. Suppose that the limiting distribution $\pi = \lim_{t\to\infty} p_t$ exists. For any $0 < \epsilon \leq 1$, the ϵ-mixing time $\tau_\epsilon(P)$ of P is defined as the smallest t such that $d_{\mathrm{TV}}(p_t, \pi) \leq \epsilon$ for any initial distribution p_0, where d_{TV} is the total variation distance in Definition 6.7.*

To bound the mixing time, we use the same approach and assume that α_2 and $|\alpha_n|$ are bounded away from one for α_i^t to converge to zero quickly for $2 \leq i \leq n$.

Theorem 6.16 (Bounding Mixing Time by Spectral Gap)**.** *Let $G = (V, E)$ be a d-regular undirected graph with $V = [n]$. Let $P = \mathcal{A}$ be the transition matrix of random walks on G and $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be its eigenvalues. Let $g := \min\{1 - \alpha_2, 1 - |\alpha_n|\}$ be the spectral gap. Then the ϵ-mixing time of P is*

$$\tau_\epsilon(P) \lesssim \frac{1}{g} \ln\left(\frac{n}{\epsilon}\right).$$

Proof. Recall from Proposition 6.14 that $P^t p_0 = \frac{\vec{1}}{n} + \sum_{i=2}^n c_i \alpha_i^t v_i$ where $v_1, \ldots, v_n$ are the orthonormal eigenvectors, and the limiting distribution is $\pi = \vec{1}/n$. So,

$$d_{\mathrm{TV}}(p_t, \pi) = d_{\mathrm{TV}}(P^t p_0, \pi) = \frac{1}{2}\left\| P^t p_0 - \frac{\vec{1}}{n} \right\|_1 = \frac{1}{2}\left\| \sum_{i=2}^n c_i \alpha_i^t v_i \right\|_1 \lesssim \sqrt{n} \left\| \sum_{i=2}^n c_i \alpha_i^t v_i \right\|_2.$$

Since $v_1, \ldots, v_n$ are orthonormal, it follows that from the spectral gap assumption that

$$\left\| \sum_{i=2}^n c_i \alpha_i^t v_i \right\|_2^2 = \sum_{i=2}^n c_i^2 \alpha_i^{2t} \leq (1-g)^{2t} \sum_{i=2}^n c_i^2.$$

Note that $\sum_{i=2}^n c_i^2 \leq \sum_{i=1}^n c_i^2 = \|p_0\|_2^2 \leq \|p_0\|_1^2 = 1$, and thus

$$d_{\mathrm{TV}}(p_t, \pi) \lesssim \sqrt{n} \left\| \sum_{i=2}^n c_i \alpha_i^t v_i \right\|_2 \leq \sqrt{n(1-g)^{2t} \sum_{i=2}^n c_i^2} \leq \sqrt{n}(1-g)^t \leq \sqrt{n} e^{-gt}.$$

Therefore, by setting $t \gtrsim \frac{1}{g} \ln(\frac{n}{\epsilon})$, we have $d_{\mathrm{TV}}(p_t, \pi) \leq \epsilon$ for any initial distribution p_0. □

For the lazy random walk matrix W, recall that the smallest eigenvalue is at least 0, and so the spectral gap for W is simply $g = \frac{1}{2}(1 - \alpha_2)$. The following is an important corollary from Cheeger's inequality in Theorem 4.3 that $1 - \alpha_2 \gtrsim \phi(G)^2$

Corollary 6.17 (Bounding Mixing Time by Conductance)**.** *Let $G = (V, E)$ be a d-regular undirected graph with $V = [n]$. Let $W = \frac{1}{2}I + \frac{1}{2}\mathcal{A}$ be the transition matrix of lazy random walks on G. Then the ϵ-mixing time of W is*

$$\tau_\epsilon(W) \lesssim \frac{1}{\phi(G)^2} \ln\left(\frac{n}{\epsilon}\right).$$

This provides a combinatorial condition that guarantees fast mixing. In particular, for expander graph where $\phi(G) = \Omega(1)$, the mixing time of lazy random walks is $O(\ln n)$, such that the random walk reaches an almost uniformly random vertex after $O(\ln n)$ steps. For many applications, it is important that the mixing time is polylogarithmic in the graph size, as we will discuss some examples soon.

Corollary 6.17 is very useful in designing random sampling algorithms. For the purpose of uniform random sampling, the analysis for regular graphs is usually enough, as we can design the Markov chain (possibly by adding self-loops) so that the underlying graph is regular.

Spectrum for General Graphs

The random walk matrix for general graphs is $P = D^{-1}A$ and the lazy random walk matrix is $W = \frac{1}{2}I + \frac{1}{2}P$. The main difference is that these matrices are not necessarily symmetric, and so we cannot directy apply the spectral theorem in Theorem 2.5 to reason about their eigenvalues and eigenvectors. A simple but important observation is that P and W are similar to a real symmetric matrix (see Definition 2.3 for the definition of similar matrices), and so the eignevalues of P and W are still all real numbers.

Lemma 6.18 (Spectrum of Random Walk Matrices)**.** *Let $G = (V, E)$ be a connected undirected graph with $V = [n]$ and $\mathcal{A}$ be its normalized adjacency matrix. Let the eigenvalues of $\mathcal{A}$ be $\alpha_1 > \alpha_2 \geq \ldots \geq \alpha_n$ and $v_1, v_2, \ldots, v_n$ be the corresponding orthonormal basis of eigenvectors. Then the eigenvalues of the random walk matrix $P = D^{-1}A$ are $\alpha_1 > \alpha_2 \geq \ldots \geq \alpha_n$, and the corresponding eigenvectors of $P^T = AD^{-1}$ are $D^{\frac{1}{2}}v_1, D^{\frac{1}{2}}v_2, \ldots, D^{\frac{1}{2}}v_n$. Also, the eigenvalues of the lazy random walk matrix $W = \frac{1}{2}I + \frac{1}{2}P$ are $\frac{1}{2}(1+\alpha_1) > \frac{1}{2}(1+\alpha_2) \geq \ldots \geq \frac{1}{2}(1+\alpha_n)$, and the corresponding eigenvectors of W^T are $D^{\frac{1}{2}}v_1, D^{\frac{1}{2}}v_2, \ldots, D^{\frac{1}{2}}v_n$.*

Proof. Note that $P = D^{-1}A = D^{-\frac{1}{2}}(D^{-\frac{1}{2}}AD^{-\frac{1}{2}})D^{\frac{1}{2}} = D^{-\frac{1}{2}}\mathcal{A}D^{\frac{1}{2}}$, and so P is similar to $\mathcal{A}$, as D is non-singular when the graph is connected. By the same argument, W is similar to $\frac{1}{2}I + \frac{1}{2}\mathcal{A}$. So, by Fact 2.4, P and $\mathcal{A}$ have the same spectrum, and W and $\frac{1}{2}I + \frac{1}{2}\mathcal{A}$ have the same spectrum. Note that $D^{\frac{1}{2}}v_i$ is an eigenvector of P^T with eigenvalue α_i, as $P^T\big(D^{\frac{1}{2}}v_i\big) = \big(D^{\frac{1}{2}}\mathcal{A}D^{-\frac{1}{2}}\big)\big(D^{\frac{1}{2}}v_i\big) = D^{\frac{1}{2}}\mathcal{A}v_i = \alpha_i\big(D^{\frac{1}{2}}v_i\big)$, and similarly $D^{\frac{1}{2}}v_i$ is an eigenvector of W_T with eigenvalue $\frac{1}{2}(1+\alpha_i)$. □

Note that $D^{\frac{1}{2}}v_1, \ldots, D^{\frac{1}{2}}v_n$ are linearly independent as D is non-singular for a connected graph. Therefore, any initial distribution p_0 can be written as $\sum_{i=1}^{n} c_i D^{\frac{1}{2}}v_i$, a linear combination of the eigenvectors of P^T and W^T. With this setup, we can adapt the proof in Proposition 6.14 to prove the following equivalent form of the fundamental theorem for undirected graphs in Theorem 6.11.

Problem 6.19 (Limiting Distribution for Undirected Graphs)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Let $P = D^{-1}A$ be the transition matrix of random walks on G and $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be its eigenvalues. If $\alpha_2 < 1$ and $\alpha_n > -1$, then $\lim_{t\to\infty}(P^T)^t p_0 = \frac{\vec{d}}{2|E|}$ where $\vec{d}$ is the degree vector with $\vec{d}(i) = \deg(i)$ for $1 \leq i \leq n$.*

Then the same approach as in Theorem 6.16 works to bound the mixing time for general undirected graphs. To bound $\|(P^T)^t p_0 - \pi\|_2$, it will be more convenient to bound $\|D^{-\frac{1}{2}}\big((P^T)^t p_0 - \pi\big)\|_2^2$, so as to take advantage of the orthonormality of $v_1, \ldots, v_n$. Then, one can bound $\|(P^T)^t p_0 - \pi\|_2 \leq \sqrt{n}\sqrt{\frac{d_{\max}}{d_{\min}}}(1-g)^t$ where g is the spectral gap, and extend the results for d-regular graphs to general undirected graphs.

Problem 6.20 (Bounding Mixing Time)**.** *Prove Theorem 6.16 and Corollary 6.17 for general undirected graphs.*

The same arguments work for weighted undirected graphs basically without modifications, by generalizing the definitions as in Section 4.4.

We remark that this spectral approach can be extended to prove the fundamental theorem for directed graphs as well, but it is more involved and requires the Perron-Frobenius theorem and the Jordan normal form (see [HJ13] for proofs).

6.4 Random Sampling

An important application of random walks is in random sampling. Consider the following algorithm for generating a random spanning tree of an undirected graph.

Algorithm 3 Random Exchange Algorithm for Sampling Random Spanning Trees

Require: An undirected graph $G = (V, E)$.

Compute an arbitrary spanning tree T_0 of the graph.
for $1 \leq t \leq \tau$ **do**
  Choose a uniform random edge $e \in E - T_{t-1}$ to T_{t-1}.
  Choose a uniform random edge f in the unique cycle in $T_{t-1} + e$.
  Set $T_t := T_{t-1} + e - f$.
end for
return T_τ.

To analyze this algorithm, it is equivalent to analyzing the random walks on a huge "spanning tree exchange graph" H, in which there is a vertex in H for each spanning tree of the graph, and two vertices in H have an edge if the corresponding spanning trees T and T' can be obtained from one step of the algorithm (i.e. $T' = T + e - f$ for some edges e, f in the input graph).

Note that if the original graph G has n vertices, then this exchange graph H could have as many as $\Omega(n^{n-2})$ vertices. So, to prove that $\tau \lesssim \text{poly}(n)$ would work to return an almost uniform random spanning tree, we must prove that the random walks in the exchange graph mix in polylogarithmic time in the size of H. In other words, we need to prove that the spanning tree exchange graph H is an expander graph.

This is usually a difficult task. There are different approaches to prove fast mixing of Markov chains. One is called the coupling method, which is the most common and versatile probabilistic technique in bounding mixing time (see [LPW06]). Another is called the canonical path method, which is based on using multicommodity flow to lower bound the graph conductance so as to use Corollary 6.17 to upper bound the mixing time. A very important application of the canonical path method is to approximate the permanent of a non-negative matrix [JSV04], which is equivalent to counting the number of perfect matchings in a bipartite graph.

We won't discuss these methods in this course, but we will see how to analyze the random exchange algorithm for sampling random spanning trees using the new techniques from high dimensional expanders in the third part of the course.

6.5 Local Graph Partitioning

Another useful application of random walks is in graph partitioning. As we have seen in the small-set expansion problem in Section 5.2, the random walk distribution $W^t\chi_i$ for some starting vertex i can be used to find a small sparse cut of the graph. This idea was originally proposed by Spielman and Teng [ST13]. They proved that the performance of the random walk algorithm for graph partitioning is similar to that of the spectral partitioning algorithm in Chapter 4. Furthermore, the random walk algorithm has the important advantage that it can be implemented *locally*, such that the running time depends only on output size but not on the original graph size, and this provides a sublinear time algorithm for graph partitioning for some instances. This idea can also be used to design approximation algorithms for the small-set expansion problem [KL12]. Local graph partitioning is an active research topic on its own, and there are several other algorithms such as using PageRank vector [ACL06] and evolving sets [AOPT16]. We won't discuss these results further in this course. This is a good project topic for those who are interested in these algorithms.

6.6 Problems

Problem 6.21 (Lower Bounding Mixing Time)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Let $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$ be the transition matrix of lazy random walks on G. Prove that the ϵ-mixing time of W is*

$$\tau_\epsilon(W) \gtrsim \frac{1}{1-\alpha_2} \ln\left(\frac{1}{\epsilon}\right),$$

where α_2 is the second largest eigenvalue of the normalized adjacency matrix $\mathcal{A}(G)$. A simpler problem is to prove that

$$\tau_\epsilon(W) \gtrsim \frac{1}{\phi(G)} \ln\left(\frac{1}{\epsilon}\right),$$

where $\phi(G)$ is the edge conductance of G. You may also consider the special case when G is d-regular.

Problem 6.22 (Page Ranking)**.** *Suppose someone searches a keyword (e.g. "car") and we would like to identify the webpages that are the most relevant for this keyword and the webpages that are the most reliable sources for this keyword (a page is a reliable source if it points to many most relevant pages). First we identify the pages with this keyword and ignore all other pages. Then we run the following ranking algorithm on the remaining pages. Each vertex corresponds to a remaining page, and there is a directed edge from page i to page j if there is a link from page i to page j. Call this directed graph $G = (V, E)$. For each vertex i, we have two values $s(i)$ and $r(i)$, where intentionally $r(i)$ represents how relevant is this page and $s(i)$ represents how reliable it is as a source (the larger the values the better). We start from some arbitrary initial values, say $s(i) = 1/|V|$ for all i, as we have no ideas at the beginning. At each step, we update s and r (where s and r are vectors of $s(i)$ and $r(i)$ values) as follows: First we update $r(i) = \sum_{j:ji\in E} s(j)$ for all i, as a page is more relevant if it is linked by many reliable sources. Then we update $s(i) = \sum_{j:ij\in E} r(j)$ for all i (using the just updated values $r(j)$), as a page is a more reliable source if it points to many relevant pages. To keep the values small, we let $R = \sum_{i=1}^{|V|} r(i)$ and $S = \sum_{i=1}^{|V|} s(i)$, and divide each $s(i)$ by S and divide each $r(i)$ by R. We repeat this step for many times to refine the values.*

Let $s, r \in \mathbb{R}^{|V|}$ be the vectors of the s and r values. Give a matrix formulation for computing s and r. Suppose G is weakly connected (when we ignore the direction of the edges the underlying undirected graph is connected) and there is a self-loop at each vertex. Prove that there is a unique

limiting s and a unique limiting r for any initial s as long as $s \geq 0$ and $s \neq 0$. You may use the Perror-Frobenius Theorem 2.16 to solve this problem.

6.7 References

[ACL06] Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. Local graph partitioning using pagerank vectors. In *FOCS*, pages 475–486. IEEE Computer Society, 2006. 54

[AF02] David Aldous and James Allen Fill. *Reversible Markov Chains and Random Walks on Graphs*. 2002. http://www.stat.berkeley.edu/~aldous/RWG/book.html. 45

[AOPT16] Reid Andersen, Shayan Oveis Gharan, Yuval Peres, and Luca Trevisan. Almost optimal local graph clustering using evolving sets. *J. ACM*, 63(2):15:1–15:31, 2016. 54

[DS84] Peter G. Doyle and J. Laurie Snell. *Random Walks and Electric Networks*. Number Book 22 in Carus Mathematical Monographs. Mathematical Association of America, Washington, DC, 1984. 45

[Häg02] Olle Häggström. *Finite Markov Chains and Algorithmic Applications*. London Mathematical Society Student Texts. Cambridge University Press, 2002. 45, 46, 47

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge; New York, 2nd edition, 2013. 12, 13, 47, 53

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004. 53, 172

[KL12] Tsz Chiu Kwok and Lap Chi Lau. Finding small sparse cuts by random walk. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 15th International Workshop, APPROX 2012*, volume 7408 of *Lecture Notes in Computer Science*, pages 615–626. Springer, 2012. 54

[LP16] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016. Available at https://rdlyons.pages.iu.edu/. 45

[LPW06] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006. 45, 46, 47, 53

[Spi19] Daniel A. Spielman. *Spectral and Algebraic Graph Theory*. 2019. 17, 45

[ST13] Daniel A. Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM J. Comput.*, 42(1):1–26, 2013. 54

Chapter 7

Expander Graphs

In this chapter, we first define expander graphs and see some of their properties. Then, we study a deterministic combinatorial construction of expander graphs, called the zig-zag products. Then, we discuss various interesting and important applications of expander graphs. Most of the material in this chapter is extracted from the excellent survey by Hoory, Linial, and Wigderson [HLW06].

There are several possible ways to define regular expander graphs.

1. Combinatorically, expander graphs are graphs with very good "connectivity", e.g. graphs with good edge expansion or vertex expansion.

2. Probabilistically, expander graphs are graphs in which random walks mix rapidly.

3. Algebraically, expander graphs are graphs with a large spectral gap $\alpha_1 - \alpha_2$.

We have already seen in Chapter 4 and Chapter 6 that these definitions are closely related. Cheeger's inequality in Theorem 4.3 states that a graph has a large spectral gap if and only if its edge expansion is large. The spectral analysis in Corollary 6.17 and Problem 6.21 show that lazy random walks mix quickly if and only if the spectral gap is large.

Note that complete graphs are the best expander graphs in each of the above definitions, but we are interested in sparse expander graphs with linear number of edges, that is, d-regular expander graphs with constant d. In constructions of expander graphs, the spectral definition is the most convenient, and we will use the following stronger spectral defintion that also bounds the last eigenvalue.

Definition 7.1 (Spectral Expanders)**.** *Let G be a d-regular graph and let $d = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq -d$ be the spectrum of its adjacency matrix. We say that G is an (n, d, ϵ)-graph if it has n vertices, is d-regular, and with $\max\{\alpha_2, |\alpha_n|\} \leq \epsilon d$. The quantity $\alpha := \max\{\alpha_2, |\alpha_n|\}$ is called the spectral radius of the graph.*

The smaller is the spectral radius, the stronger the graph is as a spectral expander. Probabilistically, the spectral radius is small if and only if the non-lazy random walks mix rapidly, as shown in Theorem 6.16. Combinatorically, $|\alpha_n|$ is small if and only if there is no nearly bipartite component, as shown in Theorem 5.4.

7.1 Properties of Expander Graphs

We collect more combinatorial and probabilistic properties of a spectral expander in this section.

Expander Mixing Lemma

A well-known and useful property of expander graphs is that it behaves as a random d-regular graph. Consider the number of edges between two subsets S, T of vertices.

Definition 7.2 (Induced Edges)**.** *Given a graph $G = (V, E)$ and $S, T \subseteq V$, define $E(S,T) := \{(u,v) \mid u \in S, v \in T, uv \in E\}$ be the set of ordered pairs where $u \in S$ and $v \in T$. Note that an edge with $u \in S \cap T$ and $v \in S \cap T$ is counted twice, as both (u,v) and (v,u) are in $E(S,T)$.*

In a random graph where every pair of vertices has an edge with probability $\frac{d}{n}$, we expect that $|E(S,T)|$ is close to $\frac{d}{n}|S||T|$. The expander mixing lemma says that in a spectral expander $|E(S,T)|$ is close to this expectation.

Theorem 7.3 (Expander Mixing Lemma)**.** *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. If the spectral radius of G is α, then for every $S \subseteq V$ and $T \subseteq V$,*

$$\left| \left|E(S,T)\right| - \frac{d|S||T|}{n} \right| \leq \alpha\sqrt{|S||T|}.$$

Proof. First, we write $|E(S,T)|$ as an algebraic expression. Let χ_S and χ_T be the characteristic vectors of S and T, such that $\chi_S(i) = 1$ if $i \in S$ and $\chi_S(i) = 0$ if $i \notin S$. Notice that $|E(S,T)| = \chi_S^T A \chi_T$, where A is the adjacency matrix of G.

Then, we use eigen-decompositions of χ_S and χ_T to relate $|E(S,T)|$ to the eigenvalues of A. Let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors of A. Recall that $\alpha_1 = d$ and $v_1 = \frac{1}{\sqrt{n}}\vec{1}$. Write $\chi_S = \sum_{i=1}^n a_i v_i$ and $\chi_T = \sum_{i=1}^n b_i v_i$ as linear combination of the eigenvectors. So, $a_1 = \langle \chi_S, v_1 \rangle = \frac{|S|}{\sqrt{n}}$ and $b_1 = \langle \chi_T, v_1 \rangle = \frac{|T|}{\sqrt{n}}$. Then, by orthonormality of $v_1, \ldots, v_n$,

$$\left|E(S,T)\right| = \chi_S^T A \chi_T = \sum_{i=1}^n \alpha_i a_i b_i = \frac{d|S||T|}{n} + \sum_{i=2}^n \alpha_i a_i b_i.$$

Therefore, by the definition of spectral radius and an application of the Cauchy-Schwarz inequality,

$$\left| \left|E(S,T)\right| - \frac{d|S||T|}{n} \right| \leq \left| \sum_{i=2}^n \alpha_i a_i b_i \right| \leq \alpha \sum_{i=2}^n |a_i||b_i| \leq \alpha \|\vec{a}\|_2 \|\vec{b}\|_2 = \alpha \|\chi_S\|_2 \|\chi_T\|_2 = \alpha\sqrt{|S||T|},$$

where $\vec{a} = (a_1, \ldots, a_n)$ and $\vec{b} = (b_1, \ldots, b_n)$. □

The following is a consequence of the expander mixing lemma.

Exercise 7.4 (Maximum Independent Set of Spectral Expanders)**.** *Let $G = (V, E)$ be a d-regular graph with $V = [n]$ with spectral radius α. Show that the size of a maximum independent set is at most $\frac{\alpha n}{d}$. Conclude that an (n, d, ϵ)-graph has chromatic number at least $\frac{1}{\epsilon}$.*

Converse of Expander Mixing Lemma

Interestingly, Bilu and Linial [BL06] proved a converse of the expander mixing lemma, showing that it comes close in characterizing the spectral radius of a graph.

Theorem 7.5 (Converse of Expander Mixing Lemma [BL06]). *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. Suppose that for any subsets $S, T \subseteq V$ with $S \cap T \neq \emptyset$, it holds that*

$$\left| |E(S,T)| - \frac{d|S||T|}{n} \right| \leq \alpha \sqrt{|S||T|}.$$

Then all but the largest eigenvalue of $A(G)$ are bounded in absolute value by $O\big(\alpha(1 + \log \frac{d}{\alpha})\big)$.

The proof of Theorem 7.5 is based on the following linear algebraic lemma.

Lemma 7.6 (Bounding Spectral Radius [BL06]). *Let A be an $n \times n$ real symmetric matrix such that the ℓ_1-norm of each row of A is at most d, and all diagonal entries of A are with absolute value $O\big(\alpha \log(\frac{d}{\alpha}) + 1\big)$. Suppose that for any two vectors $u, v \in \{0,1\}^n$ with $\mathrm{supp}(u) \cap \mathrm{supp}(v) = \emptyset$, it holds that $|u^T A v| \leq \alpha \|u\|_2 \|v\|_2$. Then the spectral radius of A is $O\big(\alpha \log(\frac{d}{\alpha}) + 1\big)$.*

The proof of Lemma 7.6 is based on linear programming duality and is not quite intuitive. It would be very interesting if there is a proof of Theorem 7.5 which is of a similar style of Trevisan's proof of Cheeger's inequality in Theorem 4.3.

Vertex Expansion

Cheeger's inequality proves that a d-regular spectral expander has large edge expansion. One could lower bound the vertex expansion of a d-regular graph through edge expansion, but with a factor d loss. Tanner's theorem proves a much stronger lower bound than that followed from edge expansion.

Definition 7.7 (Vertex Boundary). *Let $G = (V, E)$ be an undirected graph. For $S \subseteq V$, the open vertex boundary of S is defined as $\partial(S) := \{v \in V - S \mid \exists u \in S \text{ with } uv \in E\}$, and the closed vertex boundary of S is defined as $\partial[S] := S \cup \partial(S)$.*

Definition 7.8 (Vertex Expansion). *Let $G = (V, E)$ be an undirected graph. The vertex expansion of a subset $S \subseteq V$ and of a graph are defined as*

$$\psi(S) := \frac{|\partial(S)|}{|S|} \quad \text{and} \quad \psi(G) := \min_{S : |S| \leq |V|/2} \psi(S).$$

Theorem 7.9 (Tanner's Theorem). *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. Suppose the spectral radius of G is at most ϵd for some $0 < \epsilon < 1$. Then, for any $0 < \delta \leq 1/2$, for any subset $S \subseteq V$ with $|S| = \delta n$,*

$$\psi(S) \geq \frac{1}{\delta(1 - \epsilon^2) + \epsilon^2} - 1.$$

Proof. The key is to consider the quantity $\|A\chi_S\|_2^2$, where A is the adjacency matrix and χ_S is the characteristic vector of $S \subseteq V$. For a vertex $v \in V$, let $\deg_S(v) := |\{u \in S \mid uv \in E\}|$ be the number of neighbors of v in S. On one hand,

$$\|A\chi_S\|_2^2 = \sum_{v \in V} \deg_S(v)^2 = \sum_{v \in \partial[S]} \deg_S(v)^2 \geq \frac{\big(\sum_{v \in \partial[S]} \deg_S(v)\big)^2}{|\partial[S]|} = \frac{(d|S|)^2}{|\partial[S]|}.$$

On the other hand, we write $\chi_S = \sum_{i=1}^n c_i v_i$ as a linear combination of the orthonormal eigenvectors of A, with $v_1 = \vec{1}/\sqrt{n}$ and $c_1 = \langle \chi_S, v_1 \rangle = |S|/\sqrt{n}$. Then

$$\|A\chi_S\|_2^2 = \Big\|\sum_{i=1}^n c_i \alpha_i v_i\Big\|_2^2 = \sum_{i=1}^n c_i^2 \alpha_i^2 \le \frac{d^2|S|^2}{n} + (\epsilon d)^2\big(\|\chi_S\|^2 - c_1^2\big) = d^2|S|\big(\delta + \epsilon^2(1-\delta)\big),$$

where the second equality is by orthonormality of $v_1, \dots, v_n$, the inequality is by the assumption of the spectral radius and $\sum_{i=1}^n c_i^2 = \|\chi_S\|^2$, and the final equality is by plugging in $|S| = \delta n$. Combining the inequalities yields the theorem. □

Note that when $\delta \ll \epsilon^2$, Tanner's theorem gives $\psi(S) \gtrsim 1/\epsilon^2$, which implies that $|\partial(S)|$ is much larger than $|S|$ when $|S|$ is small enough. Check that a straightforward application of Cheeger's inequality only gives $\psi(S) \ge \frac{1}{2}(1-\epsilon)$ for $|S| \le |V|/2$.

Alon-Boppana Bound

How small can the spectral radius be? There are graphs, called Ramanujan graphs, with spectral radius $2\sqrt{d-1}$. This is essentially tight, as the following theorem by Alon and Boppana showed.

Theorem 7.10 (Alon-Boppana Bound)**.** *Let $G = (V, E)$ be a d-regular graph and α_2 be the second largest eigenvalue of its adjacency matrix. Then*

$$\alpha_2 \ge 2\sqrt{d-1} - \frac{2\sqrt{d-1} - 1}{\lfloor \mathrm{diag}(G)/2 \rfloor},$$

where $\mathrm{diag}(G)$ *denotes the diameter of the graph G.*

Note that the theorem implies that if we have an infinite family of d-regular expander graphs each has spectral radius at most α, then $\alpha \ge 2\sqrt{d-1}$ as the diameter goes to infinity as the size of the graph grows.

There are two different proofs of this result. One is by the trace method, which computes the number of closed walks that starts and ends at some given vertex. Another is by constructing a function with small Rayleigh quotient. These two methods can also be used to solve Problem 3.10, which is closely related to the spectral radius of Ramanujan graphs as we will see later in the course.

We will not prove Theorem 7.10 and refer the reader to [HLW06] or Trevisan's blog posts for proofs. We just present an easy proof that the spectral radius is at least $\sqrt{d}\big(1 - o(1)\big)$, using a very simple trace argument.

Claim 7.11 (Easy Lower Bound on Spectral Radius)**.** *Let $G = (V, E)$ be a d-regular graph with $V = [n]$. Then its spectral radius α is at least $\sqrt{d}\sqrt{\frac{n-d}{n-1}}$.*

Proof. Note that $\mathrm{Tr}(A^2) \ge nd$, as each edge uv contributes one length-two walk from u to u and one length-two walk from v to v. On the other hand, by Fact 2.35, $\mathrm{Tr}(A^2) = \sum_{i=1}^n \alpha_i^2 \le d^2 + (n-1)\alpha^2$. Combining the two inequalities gives the claim. □

Random Walks on Expander Graphs

We know from Theorem 6.16 that random walks on a (n, d, ϵ)-graph converge to the uniform distribution in $O(\log n/(1-\epsilon))$ steps. Interestingly, random walks on expander graphs not only give good randomness properties for the final vertex in the walk, but also for the sequence of vertices traversed in the walk. In some applications, the sequence of vertices of a walk can be used to replace a sequence of indpendent uniform random variables.

The following result is not of the most general form, but it will be enough for the application of probability amplication that we will see in Section 7.3. See [HLW06, Vad12] for more general statements. To get an intuition, it is instructive to compare the probability bound below with the probability bound when each X_i is an independent uniform random sample.

Theorem 7.12 (Concentration Property of Random Walks on Spectral Expanders)**.** *Let $G = (V, E)$ be a d-regular graph with spectral radius ϵd for some $\epsilon \leq 1/10$. Let $B \subseteq V$ with $|B| \leq \frac{1}{100}|V|$. Let X_0 be a uniform random vertex, and $X_1, \ldots, X_t$ be the vertices produced by t steps of a random walk. Let $S = \{i \mid X_i \in B\}$ be the set of times when the random walk is in B. Then*

$$\Pr\left(|S| > \frac{t}{2}\right) \leq \left(\frac{2}{\sqrt{5}}\right)^{t+1}.$$

Proof. We first set up the matrix formulation of the problem. Let $n = |V|$. The initial distribution $p_0 = \vec{1}/n$ of X_0 is the uniform distribution. Let χ_B and $\chi_{\overline{B}}$ be the characteristic vectors of B and $\overline{B}$ respectively, where $\overline{B} = V - B$. Let I_B be the diagonal matrix with a 1 in the i-th diagonal entry if $i \in B$ and zero otherwise, and similarly $I_{\overline{B}}$. Let p be a probability vector, i.e. p is non-negative and the sum of its entries is at most one. Then $I_B \cdot p$ is the probability vector that is the restriction of p on B, such that $(I_B \cdot p)(i) = p(i)$ if $i \in B$ and $(I_B \cdot p)(i) = 0$ if $i \notin B$. Check that the probability the random walk is in B at precisely the time steps in S is

$$p_S := \vec{1}^T (I_{Z_t}\mathcal{A})(I_{Z_{t-1}}\mathcal{A})(I_{Z_{t-2}}\mathcal{A}) \ldots (I_{Z_2}\mathcal{A})(I_{Z_1}\mathcal{A})p_0,$$

where $Z_i = B$ if $i \in S$ and $Z_i = \overline{B}$ if $i \notin S$, and $\mathcal{A}$ is the normalized adjacency matrix which is the probability transition matrix of the random walks. We will prove that $p_S \leq (\frac{1}{5})^{|S|}$. The theorem will then follow by a union bound as

$$\Pr\left(|S| > \frac{t}{2}\right) \leq \sum_{S:|S|>t/2} p_S \leq \sum_{S:|S|>t/2} \left(\frac{1}{5}\right)^{|S|} \leq \sum_{S:|S|>t/2} \left(\frac{1}{5}\right)^{\frac{t+1}{2}} \leq 2^{t+1}\left(\frac{1}{5}\right)^{\frac{t+1}{2}} = \left(\frac{2}{\sqrt{5}}\right)^{t+1}.$$

To prove $p_S \leq (\frac{1}{5})^{|S|}$, we use the concept of operator norm in Definition 2.17. Check that $\|I_B\|_{\text{op}} = \|I_{\overline{B}}\|_{\text{op}} = \|\mathcal{A}\|_{\text{op}} = 1$. We will prove that $\|I_B\mathcal{A}\|_{\text{op}} \leq \frac{1}{5}$, and this would imply that $p_S \leq (\frac{1}{5})^{|S|}$ because

$$\begin{aligned}
p_S &= \vec{1}^T (I_{Z_t}\mathcal{A})(I_{Z_{t-1}}\mathcal{A})(I_{Z_{t-2}}\mathcal{A}) \ldots (I_{Z_2}\mathcal{A})(I_{Z_1}\mathcal{A})p_0 \\
&\leq \|\vec{1}\|_2 \cdot \left\|(I_{Z_t}\mathcal{A})(I_{Z_{t-1}}\mathcal{A})(I_{Z_{t-2}}\mathcal{A}) \ldots (I_{Z_2}\mathcal{A})(I_{Z_1}\mathcal{A})p_0\right\|_2 && \text{by Cauchy-Schwarz} \\
&\leq \|\vec{1}\|_2 \cdot \left(\prod_{i=1}^{t} \|I_{Z_i}\mathcal{A}\|_{\text{op}}\right) \cdot \|p_0\|_2 && \text{by Fact 2.19} \\
&\leq \|\vec{1}\|_2 \cdot \left(\frac{1}{5}\right)^{|S|} \cdot \|p_0\|_2 && \text{as } \|I_B\mathcal{A}\|_{\text{op}} \leq \frac{1}{5} \text{ and } \|I_{\overline{B}}\mathcal{A}\|_{\text{op}} \leq 1 \\
&= \left(\frac{1}{5}\right)^{|S|} && \text{as } \|1\|_2 = \sqrt{n} \text{ and } \|p_0\|_2 = \frac{1}{\sqrt{n}}.
\end{aligned}$$

It remains to prove that $\|I_B\mathcal{A}\|_{\text{op}} \le \frac{1}{5}$. Let x be any nonzero vector. Write $x = c_1v_1 + \ldots + c_nv_n$, where $v_1, \ldots, v_n$ are the orthonormal eigenvectors of $\mathcal{A}$ with eigenvalues $\alpha_1 \ge \ldots \ge \alpha_n$. Then

$$\|I_B\mathcal{A}x\|_2^2 = \|I_B\mathcal{A}(c_1v_1 + \ldots + c_nv_n)\|_2^2 = \left\|I_B\sum_{i=1}^n c_i\alpha_iv_i\right\|_2^2 \le 2\|I_Bc_1\alpha_1v_1\|_2^2 + 2\left\|I_B\sum_{i=2}^n c_i\alpha_iv_i\right\|_2^2,$$

where the inequality is by $\|x+y\|_2^2 \le 2\|x\|_2^2 + 2\|y\|_2^2$. Recall that $\alpha_1 = 1$, $v_1 = \vec{1}/\sqrt{n}$ and $c_1 = \langle x, v_1\rangle = \frac{1}{\sqrt{n}}\cdot\sum_{i=1}^n x(i)$. So, the first term on the RHS is

$$2\|I_Bc_1\alpha_1v_1\|_2^2 = 2\left\|\frac{1}{n}\Big(\sum_{i=1}^n x(i)\Big)I_B\vec{1}\right\|_2^2 = 2|B|\left(\frac{\sum_{i=1}^n x(i)}{n}\right)^2 \le 2|B|\cdot\frac{\|x\|_2^2}{n} \le \frac{1}{50}\|x\|_2^2,$$

where the first inequality is by Cauchy-Schwarz and the second inequality is by the assumption that $|B| \le \frac{n}{100}$. The second term on the RHS is

$$2\left\|I_B\sum_{i=2}^n c_i\alpha_iv_i\right\|_2^2 \le 2\|I_B\|_{\text{op}}^2\cdot\left\|\sum_{i=2}^n c_i\alpha_iv_i\right\|_2^2 = 2\sum_{i=2}^n c_i^2\alpha_i^2 \le 2\epsilon^2\sum_{i=2}^n c_i^2 \le 2\epsilon^2\|x\|_2^2 \le \frac{1}{50}\|x\|_2^2,$$

where the equality is by orthnormality of $v_1, \ldots, v_n$, the second inequality is by the assumption that the spectral radius of the adjacency matrix A is at most ϵd and so the eigenvalues of $\mathcal{A} = A/d$ satisfies $\max_{2\le i\le n}\{|\alpha_i|\} \le \epsilon$, and the last inequality is by the assumption that $\epsilon \le \frac{1}{10}$. Combining the two terms,

$$\|I_B\mathcal{A}x\|_2^2 \le \frac{1}{25}\|x\|_2^2 \quad\Longrightarrow\quad \|I_B\mathcal{A}\|_{\text{op}} \le \frac{1}{5}.$$

□

See [Gil98] for a well-known Chernoff bound for spectral expanders, and [GLSS18] for a recent generalization to the matrix setting. See [CPT21] for an interesting recent paper showing that many functions are fooled by expander random walks, in that they cannot distinguish independent random samples from those obtained by expander random walks.

7.2 Constructions of Expander Graphs

It can be shown that a random d-regular graph is an expander graph with high probability using the combinatorial definitions, by standard techniques using Chernoff bound and union bound. It is a good problem to work out the details; see [HLW06, Vad12] for the precise statements and proofs.

Perhaps surprisingly, while almost every d-regular graph is an expander graph, it is very difficult to come up with a deterministic construction of expander graphs. One possible explanation is that random graphs have high descriptive complexity, while in deterministic constructions the infinite family of expander graphs can be described in a succinct way.

There are explicit constructions of d-regular expander graphs, most of them are algebraic constructions.

- A family of 8-regular graphs G_m for every integer m. The vertex set is $V = \mathbb{Z}_m \times \mathbb{Z}_m$. The neighbors of vertex (x, y) is $(x, y\pm x), (x\pm y, y), (x, y+1\pm x), (x+1\pm y, y)$, where all additions are mod m. Note that this family is very explicit, meaning that the neighbors of a vertex can

be computed in $O(\log m)$ time, which is very useful for some applications such as probability amplification as we will see. This construction is due to Margulis whose proof did not give any explicit bound. Gabber and Galil proved that its spectral radius is at most $5\sqrt{2} < 8$. Their proof uses Fourier analysis; see [HLW06].

- A family of 3-regular p-vertex graph for every prime number p. The vertex set is $\mathbb{Z}_p$, and a vertex is connected to $x+1, x-1$ and its multiplicative inverse x^{-1} (for vertex 0 its inverse is 0), where the additions are mod p. The proof uses some deep results in number theory.
- The main source of explicit deterministic construction is from Cayley graphs, which are graphs defined by groups. Some of the stronger expanders, the Ramanujan graphs with spectral radius $2\sqrt{d-1}$, are from Cayley graphs and the proofs require sophisticated mathematical tools.

In the second part of the course, we will see a new way to show the existence of "bipartite" Ramanujan graphs using combinatorial and probabilistic methods, through interlacing family of polynomials.

In the following, we will study a combinatorial construction of expander graphs, known as the zig-zag product, whose proof is more elementary and intuitive, although the bound is not as sharp and the construction is not as explicit.

Combinatorial Constructions

The general idea of the combinatorial constructions is to construct bigger expander graphs from smaller expander graphs.

The base case could simply be a constant size complete graph. Let G be an (n, k, ϵ_1)-graph and H be an (k, d, ϵ_2)-graph. A natural product of G and H is to replace each vertex v in G by a copy of H, so that each edge incident on v is incident on a different vertex of H. This is called the replacement product of G and H.

Definition 7.13 (Replacement Product)**.** *Let G be a k-regular graph on n vertices and H be a d-regular graph on k vertices. The replacement product $G \textcircled{r} H$ is a graph where the vertex set is the Cartesian product $[n] \times [k]$ of the vertex set of G and H, and two vertices (u, i) and (v, j) have an edge if and only if (1) $u = v$ and $ij \in E(H)$ or (2) $vu \in E(G)$ and v is the i-th neighbor of u in G and u is the j-th neighbor of v in G.*

Intuitively, $G \textcircled{r} H$ is a combinatorial expander if G and H are combinatorial expanders. Consider a set $S \subseteq V(G \textcircled{r} H)$. If S has either large of small intersection with each "cloud" (copy of H), then S should have large expansion because of the large expansion of G as S is basically a set of vertices in G. If S has medium intersections with many clouds, then S should have large expansion because of the large expansion of H as there are many crossing edges within each such cloud. However, it is not clear how to make this intuition precise, as there seems to be no clean way to decompose a subset's contribution into its contribution from G and its contribution from H. In a way, the spectral proof that we are going to see soon can be thought of as a linear algebraic approach to carry out this idea in a more general setting.

Zig-Zag Product

The actual construction by Reingold, Vadhan and Wigderson [RVW02] that we will analyze is slightly more complicated.

Definition 7.14 (Zig-Zag Product). *Let G be a k-regular graph on n vertices and H be a d-regular graph on k vertices. The zig-zag product $G \textcircled{z} H$ is a graph with the same vertex set $[n] \times [k]$ as the replacement product, and two vertices (u, i) and (v, j) have an edge if and only if $u \neq v$ and there exists $a \in [k]$ such that (u, i)–(u, a), (u, a)–(v, b), and (v, b)–(v, j) are all edges in the replacement product $G \textcircled{r} H$, where (u, a)–(v, b) is the unique edge incident on (u, a) with $v \neq u$ (i.e. the unique edge incident on (u, a) that leaves the cloud of u in the replacement product).*

In words, each edge in the zig-zag product $G \textcircled{z} H$ corresponds to a length three walk in the replacement product $G \textcircled{r} H$, where the first step is within a cloud, the second step is the unique way to leave a cloud, and the third step is within the other cloud.

The intuition that the zig-zag product is a spectral expander comes from random walks. Edge edge in $G \textcircled{z} H$ corresponds to a random step in H, a deterministic step in G, and a random step in H. We should think of the first two steps as going to random neighboring cloud, and the third step corresponds to moving to a random neighbor within the neighboring cloud. Since both G and H are spectral expanders and thus have the fast mixing property, after not many steps of random walks, we won't know which cloud we are in and the location within the cloud, and so $G \textcircled{z} H$ also has the fast mixing property and hence a spectral expander.

Theorem 7.15 (Zig-Zag Theorem). *Let G be an (n, k, ϵ_1)-graph and H be an (k, d, ϵ_2)-graph. Then $G \textcircled{z} H$ is an $(nk, d^2, \epsilon_1 + \epsilon_2 + \epsilon_2^2)$-graph.*

We will prove the theorem in the next subsection. Let us first see how zig-zag product can be used to construct bigger and bigger constant degree expander graphs. The idea is to combine with the following standard operation that decreases the spectral radius.

Definition 7.16 (Graph Power). *Let G be a graph with adjacency matrix A. The k-th power G^k is the graph with the same vertex set as G and with (weighted) adjacency matrix A^k.*

In words, the number of parallel edges between u and v in G^k is equal to the number of length k walks between u and v in G. Note that the spectral radius of G^k has improved significantly, but the degree of G^k has also improved significantly.

Exercise 7.17 (Spectral Radius of Graph Power). *If G is an (n, d, ϵ)-graph, then G^k is an (n, d^k, ϵ^k)-graph.*

The idea of the combinatorial construction is to use graph power to decrease the spectral radius, and then use zig-zag product to decrease the degree while not increasing the spectral radius too much.

Theorem 7.18 (Expanders from Zig-Zag Product). *For large enough constant d, there is an infinite family of d^2-regular with spectral radius at most $\frac{1}{4}d^2$.*

Proof. Let H be a $(d^4, d, 1/16)$-graph. We can prove its existence by a probabilistic argument when d is a large enough constant. Since d is a constant, one can find it by an exhaustive search in constant time.

Using the building block H, we inductively define G_i by $G_1 = H^2$ and $G_{i+1} = G_i^2 \textcircled{z} H$. We claim that G_i is a $(d^{4i}, d^2, 1/4)$-graph for all $i \geq 1$. The base case is clearly true by Exercise 7.17. Assume G_i is a $(d^{4i}, d^2, 1/4)$-graph. Then G_i^2 is a $(d^{4i}, d^4, 1/16)$-graph by Exercise 7.17. And $G_i^2 \textcircled{z} H$ is a $(d^{4(i+1)}, d^2, 1/4)$-graph by Theorem 7.15. □

Proof of the Zig-Zag Theorem

Check that $G \textcircled{z} H$ has nk vertices and is d^2-regular. We bound the spectral radius of $G \textcircled{z} H$ in the rest of this subsection.

Matrix Formulation: The first step is to write down the walk matrix Z of the zig-zag product $G \textcircled{z} H$. Let $W(H)$ be the $k \times k$ walk matrix of H, which is simply $\frac{1}{k}A(H)$ where $A(H)$ is the adjacency matrix of H. Let W be the $nk \times nk$ matrix with n copies of W_H on the diagonal, which is the transition matrix of one step of random walk within the clouds in $G \textcircled{z} H$. The steps between clouds are deterministic: the walk moves from a vertex (u, i) to a unique vertex (v, j) with $v \neq u$. The transition matrix for this deterministic step is thus a permutation matrix P with $P_{(u,i),(v,j)} = 1$ for each inter-cloud edge and zero otherwise. It follows from the definition of the zig-zag product that

$$Z = WPW.$$

So the random walk matrix of $G \textcircled{z} H$ has a very nice form, which should be the reason for the definition of zig-zag product in Definition 7.14.

The graph $G \textcircled{z} H$ is a regular graph, and so $\vec{1}_{nk}$ is an eigenvector of Z with eigenvalue 1. To prove the zig-zag product theorem, we will prove that for all $f \perp \vec{1}_{nk}$ the Rayleigh quotient

$$R_Z(f) = \frac{|f^T Z f|}{\|f\|_2^2} \leq \epsilon_1 + \epsilon_2 + \epsilon_2^2,$$

and this will imply that the spectral radius of Z is at most $\epsilon_1+\epsilon_2+\epsilon_2^2$ by the optimization formulation of the second eigenvalue in Lemma 2.11 and an analogous formulation for the last eigenvalue.

Vector Decomposition For any $f \perp \vec{1}_{nk}$, we decompose f to two vectors to apply the results in G and in H. This is where the power of linear algebra comes from, as in the larger domain $\mathbb{R}^{nk}$ there is a natural way to decompose the vector, while in the combinatorial setting it is not clear how to decompose a set of vertices in $G \textcircled{z} H$ into a set of vertices in G and a set of vertices in H to apply the expansion properties of G and of H as we discussed before.

Define f_G as the average of f on clouds, such that $f_G(u, i) = \frac{1}{k}\sum_{j=1}^{k} f(u, j)$ for all $(u, i) \in V(G \textcircled{z} H)$, so that two vertices in the same cloud have the same value in f_G. Define $f_H = f - f_G$. Note that f_H sums to zero in each cloud, such that $\sum_{j=1}^{k} f_H(u, j) = 0$ for each $u \in G$. Using triangle inequality,

$$|f^T Z f| = |f^T WPWf| = |(f_G+f_H)^T WPW(f_G+f_H)| \leq |f_G^T WPWf_G|+2|f_G^T WPWf_H|+|f_H^T WPWf_H|.$$

Since $W(H) \cdot \vec{1}_k = \vec{1}_k$ as H is a regular graph, it follows that $Wf_G = f_G$ as vertices in the same cloud have the same value in f_G. Therefore,

$$|f^T Z f| \leq |f_G^T P f_G| + 2|f_G^T PWf_H| + |f_H^T WPWf_H|.$$

We will use the spectral expansion of G to prove $|f_G^T P f_G| \leq \epsilon_1 \|f_G\|_2^2$ in Claim 7.21, the spectral expansion of H to prove $|f_H^T WPWf_H| \leq \epsilon_2^2 \|f_H\|_2^2$ in Claim 7.19, and a simple argument to bound $2|f_G^T PWf_H| \leq 2\epsilon_2 \|f_G\|_2 \|f_H\|_2$ in Claim 7.20. Assuming these claims, then

$$\begin{aligned} |f^T Z f| &\leq \epsilon_1 \|f_G\|_2^2 + 2\epsilon_2 \|f_G\|_2 \|f_H\|_2 + \epsilon_2^2 \|f_H\|_2^2 \\ &\leq \epsilon_1 \|f_G\|_2^2 + \epsilon_2 \big(\|f_G\|_2^2 + \|f_H\|_2^2\big) + \epsilon_2^2 \|f_H\|_2^2 \\ &\leq (\epsilon_1 + \epsilon_2 + \epsilon_2^2)\|f\|_2^2, \end{aligned}$$

where the last inequality holds because $f_G \perp f_H$ and so $\|f\|_2^2 = \|f_G\|_2^2 + \|f_H\|_2^2$ and also $\|f_G\|_2 \leq \|f\|_2$ and $\|f_H\|_2 \leq \|f\|_2$. This will complete the proof of Theorem 7.15 and so it remains to prove the three claims.

Spectral Expansion The following claim uses the spectral expansion of H and that f_H sums to zero in each cloud.

Claim 7.19 (Quadratic Term of H)**.** $|f_H^T W P W f_H| \leq \epsilon_2^2 \|f_H\|_2^2$

Proof. As the spectral radius of $W(H)$ is ϵ_2, we claim that $\|W(H) \cdot x\|_2 \leq \epsilon_2 \|x\|_2$ for any $x \perp \vec{1}_k$. To see this, let $x = \sum_{i=1}^k c_i v_i$ where $v_1, \ldots, v_n$ is an orthonormal basis of eigenvectors of $W(H)$ with eigenvalues $\alpha_1, \ldots, \alpha_k$. Note that $c_1 = 0$ as $v_1 = \vec{1}/\sqrt{k}$ and $x \perp \vec{1}$. Then

$$\|W(H) \cdot x\|_2^2 = \left\| W(H) \cdot \left(\sum_{i=2}^k c_i v_i \right) \right\|_2^2 = \left\| \sum_{i=2}^k c_i \alpha_i v_i \right\|_2^2 = \sum_{i=2}^k c_i^2 \alpha_i^2 \leq \epsilon_2^2 \sum_{i=2}^k c_i^2 \leq \epsilon_2^2 \|x\|_2^2,$$

where the first inequality is by the spectral radius of $W(H)$. This implies that $\|W f_H\|_2 \leq \epsilon_2 \|f_H\|$ as the sum of the entries in each cloud is zero in f_H as we argued earlier. Therefore,

$$|f_H^T W P W f_H| \leq \|W f_H\|_2 \cdot \|P W f_H\|_2 = \|W f_H\|_2^2 \leq \epsilon_2^2 \|f_H\|_2^2,$$

where the first inequality is by Cauchy-Schwarz and the equality is because P is a permutation matrix. □

The second claim is straightforward.

Claim 7.20 (Cross Term)**.** $|f_G^T P W f_H| \leq \epsilon_2 \|f_G\|_2 \|f_H\|_2$.

Proof. By Cauchy-Schwarz,

$$|f_G^T P W f_H| \leq \|f_G\|_2 \cdot \|P W f_H\|_2 = \|f_G\|_2 \cdot \|W f_H\|_2 \leq \epsilon_2 \|f_G\|_2 \|f_H\|_2,$$

where the last inequality was established in the proof of Claim 7.19. □

The final claim uses the spectral expansion of G and that $f \perp \vec{1}_{nk}$.

Claim 7.21 (Quadratic Term of G)**.** $|f_G^T P f_G| \leq \epsilon_1 \|f_G\|_2^2$.

Proof. The main point is to see that the LHS is equal to a corresponding quadratic form of the walk matrix of G. To see this, we "contract" each cloud to a single vertex. Define $g : V(G) \to \mathbb{R}$ as $g(v) = \sqrt{k} \cdot f_G(v, i)$. Note that $\|g\|_2^2 = \|f_G\|_2^2$. Note also that $f_G^T P f_G = g^T W(G) g$, where $W(G)$ is the random walk matrix of G, as each edge (u, i)-(v, j) in $G \textcircled{z} H$ contributes $f_G(u, i) \cdot f_G(v, j)$ to $f_G^T P f_G$ while the corresponding edge $uv \in G$ contributes $\left(\sqrt{k} f_G(u, i)\right)\left(\frac{1}{k}\right)\left(\sqrt{k} f_G(v, j)\right) = f_G(u, i) \cdot f_G(v, j)$ to $g^T W g$. Therefore,

$$\frac{f_G^T P f_G}{\|f_G\|_2^2} = \frac{g^T W g}{\|g\|^2}.$$

Since $f \perp \vec{1}$, it follows that $f_G \perp \vec{1}$ and thus $g \perp \vec{1}$. As G is an (n, k, ϵ_1)-graph, we conclude that

$$\frac{f_G^T P f_G}{\|f_G\|_2^2} = \frac{g^T W g}{\|g\|^2} \leq \epsilon_1.$$

□

This concludes the proof of Theorem 7.15. The idea of decomposing a vector into different components is useful in many proofs. We will use it again when we study high dimensional expanders in the third part of the course.

7.3 Applications of Expander Graphs

We discuss some of the many interesting applications of expander graphs in this section, with more details on expander codes as they are the basics of the recent breakthroughs in designing asymptotically good codes that are locally testable [DEL+21, PK21].

Probability Amplification

Suppose we have a randomized algorithm with error probability $1/100$ requiring n random bits. To decrease the failure probability, a standard way is to run the randomized algorithm independently k times, and then take the majority answer as the output. By a standard Chernoff bound argument, this decreases the failure probability to δ^k for some small constant δ. The number of random bits used is kn.

We show how to achieve exponentially small error probability while using only $n + ck$ bits where c is a constant. First, we see the above analysis in a slightly different perspective. Let V be the set of all n-bit strings. The randomized algorithm has error probability at most $1/100$ is equivalent in saying that among the 2^n n-bit strings, at most $2^n/100$ of them are "bad" strings. Denote this set of bad strings by $B \subseteq V$. The standard algorithm of taking the majority answer would fail if and only if we choose more than $k/2$ random strings from B, which is highly unlikely as $|B| \leq \frac{1}{100}|V|$. We can interpret the standard algorithm as doing a random walk of length k on the complete graph on V, and use the corresponding bit strings of the vertices $X_1, \ldots, X_k$ on this walk.

The idea is to replace a random walk on the complete graph on V by a random walk on a constant degree expander graph on V. Construct a d-regular expander graph G with 2^n vertices with spectral radius ϵd where d is a constant and $\epsilon \leq 1/100$. This can be done, say, by taking a large enough constant power of a Margulis expander. In the first step of the random walk, we use an n-bit random string, with error probability at most $1/100$. In the subsequent steps, instead of using n random bits to find the next n-bit string, we just choose a random neighbor of the current string in G and use the corresponding string in this random neighbor. Since G is a d-regular graph, we just need to use $\lceil \log_2 d \rceil$ random bits to choose a random neighbor in each subsequent step. Thus, the total number of bits used is $n + (k-1) \cdot \lceil \log_2 d \rceil$. Note that it is important that the neighbors of a Margulis expander can be computed quickly, so that we can find out the corresponding strings in this random walk quickly.

What is the error probability of this expander walk algorithm? This is exactly what Theorem 7.12 is formulated for, which shows that the error probability of taking the majority answer of a random walk of length k on a spectral expander with $\epsilon \leq 1/100$ is at most $(2/\sqrt{5})^k$.

This is just one example of using expander graphs in derandomization; see [HLW06, Vad12, AB06] for many more. The expander mixing lemma in Theorem 7.3 is very useful in derandomization.

Constructing Efficient Objects

We can think of a d-regular expander graph as a very efficient, as it only has a linear number of edges and it achieves very high connectivity. It should not be surprising that expander graphs are

useful in constructing efficient networks.

One interesting example is the construction of *superconcentrators*, which are directed graphs with n input nodes and n output nodes, satisfying the strong connectivity property that for any $k \leq n$ there are k vertex disjoint paths between any k input nodes and any k output nodes. For instance, the complete bipartite graph $K_{n,n}$ satisfies this property, but it has $\Theta(n^2)$ edges. Valiant conjectured that there is no superconcentrator with $O(n)$ edges, in an attempt to prove circuit lower bound. Later, he found a recursive construction of superconcentrator with $O(n)$ edges using expander graphs as building blocks. See [HLW06] for details.

Superconcentrators and expander graphs can be used to design efficient algorithms as well. One application is in designing fast algorithms for computing matrix rank [CKL13], where an expander graph or a superconcentrator is used to "compress" a rectangular matrix $A \in \mathbb{F}^{m \times n}$ with $n \gg m$ into a square matrix $B \in \mathbb{F}^{m \times m}$ in linear time such that $\operatorname{rank}(A) = \operatorname{rank}(B)$ with high probability.

A famous classical example of using expander graphs is to construct optimal sorting networks [AKS83], with $O(n \log n)$ edges and depth $O(\log n)$.

Undirected Connectivity in Log-Space

A striking application of the zig-zag product in Definition 7.14 is to solve the s-t connectivity problem in an undirected graph in logarithmic space. If we are allow to use randomized algorithms, then there is a very simple algorithm to solve the s-t connectivity problem in log-space, simply running a random walk for $O(n^3)$ steps would do, as it is well-known that the expected cover time for any undirected graph is at most $O(n^3)$. There is a deterministic algorithm by Savitch that solves the more general problem of s-t connectivity in *directed* graphs in $O(\log^2 n)$ space, by recursively guessing the midpoint of a directed s-t path. It has been a long standing and important open problem whether directed s-t connectivity can be solved in log-space. If such an algorithm exists, then this would imply that $NL = L$, the complexity classes of non-deterministic log-space problems and deterministic log-space problems are the same.

Reingold [Rei08] discovered a determinstic $O(\log n)$ space algorithm for s-t connectivity in undirected graphs using zig-zag products. Suppose the input graph G is a d-regular expander graph for a constant d. Then it can be shown that G has diameter $O(\log n)$. Then one can enumerate all paths of length $O(\log n)$ in $O(\log n)$ space, since each neighbor can be described in $\lceil \log_2 d \rceil$ space as we have seen in the probability amplification application above. Reingold's idea is to transform any graph G into a d-regular expander graph H such that s, t are connected in G if and only if s, t are connected in H. First, one can reduce G into a d-regular graph with constant d by replacing each vertex of high degree by a constant degree expander graph (and adding self-loops to each low degree vertex), similar to what was done in the replacement product in Definition 7.13. To improve the expansion, one can construct the graph $(G \textcircled{z} C)^8$, where C is a $(d, d^{1/16}, 1/2)$-graph. Using a variant of the zig-zag theorem in Theorem 7.15, it is possible to prove that the spectral gap doubles in the resulting graph. Then, one just needs to repeat this construction $O(\log n)$ times to get a graph H with constant spectral gap, as the initial spectral gap is at least $\Omega(1/n^2)$ for any connected undirected graph. Note that the size of H is at most a polynomial factor larger than the size of G, and s, t are connected in G if and only if s, t are connected in H.

A technical difficulty in carrying out this approach is to compute a neighbor of a vertex in H in log-space. The hope is that there are only $O(\log n)$ recursion levels for the zig-zag construction, and in each level we only need constant space, as there are only three steps and the degree is constant. Reingold proved that this can indeed be done; see [Rei08, Vad12] for details.

Hardness Amplification

Random walks on expander graphs can also be used for hardness amplifications, to take instances that are hard to approximate and construct instances that are even harder to approximate. See for example Chapter 22 of [AB06] for a simple application of expander random walks in proving hardness of approximating maximum independent sets.

Dinur [Din07] found an amazing proof of the very important PCP theorem using expander random walks. Her proof was inspired by Reingold's result, which involves many iterations of "powering" and "degree reduction", that makes the underlying constraint satisfaction problem harder and harder to approximate. See [AB06] for a good exposition of the PCP theorem. This is a great project topic especially for those who are interested in complexity theory.

Expander Codes

A main motivation for early developments in expander graphs is from coding theory.

A code $C \subseteq \{0,1\}^n$ of length n is a subset of n-bit strings, where each string in C is called a codeword. To design a good error correcting code, we would like to choose codewords that are far from each other so as to correct more errors, but at the same time choose as many codewords as possible so as to maximize the information rate. This can be thought of as a sphere packing problem, where the objective is to fit in as many disjoint spheres of a certain radius as possible in $\mathbb{F}_2^n$.

Definition 7.22 (Distance of Code). *Given $C \subseteq \{0,1\}^n$, the distance of C is defined as $\operatorname{dist}(C) := \min_{x \neq y \in C} d_H(x,y)$, where $d_H(x,y)$ is the Hamming distance between two codewords x and y. The relative distance of C is defined as $\operatorname{dist}(C)/n$.*

Definition 7.23 (Rate of Code). *Given $C \subseteq \{0,1\}^n$, the rate of C is defined as $\log|C|/n$, where $\log|C|$ can be thought of as the number of bits of information sent.*

Definition 7.24 (Asymptotically Good Code). *A family $C_n \in \{0,1\}^n$ of codes is asymptotically good if there are constants $r > 0$ and $\delta > 0$ such that for all n both the relative distance of C_n is at least δ and the rate of C_n is at least r.*

The existence of an asymptotically good code can be proved a standard probabilistic method. For the codes to be useful in practice, we would also like that encoding and decoding can be done in polynomial time in n (and ideally linear time in n), but this makes the problem much more challenging.

A common class of codes is the class of linear codes, where C is a linear subspace of $\mathbb{F}_2^n$. Linear codes have the advantage that they can be described by a basis and so encoding can be done in $O(n^2)$ time. Also, a simple but useful property of linear codes is that the minimum distance of the code is equal to the minimum ℓ_1-norm of a non-zero codeword, because $d_H(x,y) = \|x-y\|_1$ and $x - y$ is a codeword. The natural decoding strategy is to find the nearest codeword of a received word, but this is an NP-complete problem even for linear codes.

Low Density Parity Check Codes The idea of constructing codes from graphs was first suggested by Gallager, who uses sparse bipartite graphs to design low-density parity check codes (LDPC codes).

Let A be a parity check matrix for code C, such that $C = \{x \mid Ax = 0\}$ where $A \in \{0,1\}^{m \times n}$ with $m < n$. Each row i of A is a parity-check constraint, requiring $\sum_{j=1}^{n} A_{ij} \cdot x(j) = 0$ where the addition is mod 2. Note that the rate of this code is $1 - m/n$, so we want m/n to be bounded away from 1.

The matrix A can be viewed as a bipartite graph $G = (L, R; E)$ with $L = [n]$ and $R = [m]$ between the variables and the constraints, where there is a vertex in L for each variable and a vertex in R for each constraint, and variable i and constraint j has an edge if and only if $A_{ij} = 1$. We will see that good expansion of G yields good LDPC codes.

Definition 7.25 (Left Small-Set Vertex Expansion)**.** *Let $G = (L, R; E)$ be a bipartite graph with $|L| = n$ and $|R| = m$ and $m < n$. For any $0 < \delta < 1$, define the left δ-small-set vertex expansion of G as*

$$\psi_\delta^L(G) := \min_{S \subseteq L : |S| \leq \delta n} \frac{|\partial(S)|}{|S|},$$

where $\partial(S)$ is the vertex boundary in Definition 7.7.

Note that $\psi_\delta^L(G) \leq k$ for any k-left-regular bipartite graph G and any δ. Kahale proved that a Ramanujan graph satisfies $\psi_\delta^L(G) \approx \frac{1}{2}k$ for some constant $\delta > 0$ and this bound cannot be improved. In the following, we will need a stronger requirement that $\psi_\delta^L(G) \geq \frac{3}{4}k$, which is satisfied in a random k-left-regular bipartite graph with high probability. Capalbo, Reingold, Vadhan, Wigderson gave deterministic constructions of these "lossless expanders" satisfying $\psi_\delta^L(G) \geq 0.99k$ for some $\delta > 0$ and $m/n < 0.99$ using some variant of the zig-zag product.

First we see that the relative distance of a lossless expander code is a constant. The proof uses the unique neighbor property of a lossless expander.

Theorem 7.26 (Distance of Expander Code [SS96])**.** *Let $G = (L, R; E)$ be a left k-regular bipartite graph with $\psi_\delta^L(G) > \frac{1}{2}k$. Then the parity check code $C(G)$ defined by G has relative distance greater than δ.*

Proof. Let $S \subseteq L$ be a subset of left vertices with $|S| \leq \delta n$. Then $|\partial(S)| > \frac{k}{2}|S|$ by the left small-set vertex expansion assumption of G. A simple counting argument shows that that there exists a vertex $v \in \partial(S) \subseteq R$ with only one neighbor in S. Let us call such a vertex a unique neighbor of S.

To lower bound the minimum distance, recall that it is equivalent to lower bounding the ℓ_1-norm/support-size of a codeword $x \in \{0,1\}^n$. Let S be the support of x. If $|S| \leq \delta n$, by the previous paragraph, there exists a unique neighbor $v \in R$ of S. This implies that the parity constraint on v is not satisfied by x, and thus x is not a codeword of the parity check code defined by G. Therefore, any codeword of this parity check code must have support size greater than δn, and thus the minimum distance of this code is greater than δn. □

The key feature of the LPDC codes defined by expander graphs is that there is a surprisingly simple and efficient decoding algorithm.

Algorithm 4 Flip Algorithm for Expander Code

Require: A parity check matrix $A \in \{0,1\}^{m \times n}$ and a bit string $x \in \{0,1\}^n$.

1: Let $x^{(0)} := x$ and $t = 0$.
2: **while** there is an unsatisfied parity check constraint **do**
3: Find a bit i such that flipping it decreases the number of unsatisfied parity constraints. That is, an $i \in [n]$ such that $\left\|A\left(x^{(t)} + \chi_i\right)\right\|_1 < \left\|Ax^{(t)}\right\|_1$, where χ_i is the characteristic vector of i and the addition is under arithmetic mod 2. Set $x^{(t+1)} := x^{(t)} + \chi_i$ and $t \leftarrow t + 1$.
4: **end while**
5: **return** $x^{(t)}$.

The analysis of the flip algorithm uses a stronger assumption about the left small-set vertex expansion than than in Theorem 7.26.

Theorem 7.27 (Efficient Decoding of Expander Code [SS96])**.** *Let $G = (L, R; E)$ be a left k-regular bipartite graph with $L = [n]$ and $R = [m]$ and $\psi_\delta^L(G) > \frac{3}{4}k$. Let x be an n-bit string whose distance from a codeword y is at most $\frac{1}{2}\delta n$. Then 4 will return y in at most m iterations.*

Proof. Let $\Delta^{(t)} := \{i \in [n] \mid x^{(t)}(i) \neq y(i)\}$ be the set of errors at the t-th iteration. The plan is to argue that as long as $\mathrm{dist}_H\left(x^{(t)}, y\right) = \left|\Delta^{(t)}\right| \leq \delta n$, there exists a bit i such that flipping it decreases the number of unsatisfied constraints, and also argue that $\mathrm{dist}_H\left(x^{(t)}, y\right) \leq \delta n$ for all t if $\mathrm{dist}_H\left(x^{(0)}, y\right) \leq \frac{1}{2}\delta n$. These would imply that after at most $\tau \leq m$ iterations, there will be no unsatisfied constraints and so $x^{(\tau)}$ is a codeword, and thus $x^{(\tau)}$ must be equal to y as $\mathrm{dist}_H(x^{(\tau)}, y) \leq \delta n$ while the distance between y and other codewords is strictly bigger than δn.

For ease of notation, let $\Delta := \Delta^{(t)}$ be the set of error at some iteration t. Assume that $0 < |\Delta| \leq \delta n$, we would like to argue that there is a bit i that flipping it decreases the number of unsatisfied constraints. Partition $\partial(\Delta)$ into the set of satisfied neighbors $\partial_+(\Delta)$ of Δ and the set of unsatisfied neighbors $\partial_-(\Delta)$ of Δ. On one hand, since $|\Delta| \leq \delta n$, by the left small-set vertex expansion of Δ,

$$|\partial_+(\Delta)| + |\partial_-(\Delta)| = |\partial(\Delta)| > \frac{3}{4}k|\Delta|.$$

On the other hand, when we consider the $k|\Delta|$ number of edges between Δ and $\partial(\Delta)$, observe that each vertex in $\partial_+(\Delta)$ has at least two such edges while each vertex in $\partial_-(\Delta)$ has at least one such edge, and so

$$2|\partial_+(\Delta)| + |\partial_-(\Delta)| \leq k|\Delta|.$$

Combining these two inequalities gives that $|\partial_-(\Delta)| > \frac{1}{2}k|\Delta|$. This implies that there must exist a vertex $i \in \Delta$ with strictly more unsatisfied neighbors than satisfied neighbors. Therefore, as long as $|\Delta| \leq \delta n$, there must exist a bit i such that flipping it decreases the number of unsatisfied constraints.

To complete the proof, we argue that $|\Delta| \leq \delta n$ in any iteration. Suppose this is not true, then since $|\Delta|$ changes by one in each iteration, there is an (earliest) iteration such that $|\Delta| = \delta n$. Then, by the argument in the previous paragraph, there are strictly more than $\frac{1}{2}k|\Delta| = \frac{1}{2}k\delta n$ unsatisfied constraints in that iteration. However, since $|\Delta^{(0)}| \leq \frac{1}{2}\delta n$, the number of unsatisfied constraints in the beginning is at most $\frac{1}{2}k\delta n$. This contradicts with the previous paragraph that the number of unsatisfied constraints is decreasing when $|\Delta| \leq \delta n$. □

Spielman showed that it is possible to use expander codes to obtain asymptotically good codes that are linear time encodable and decodable!

Tanner Codes Tanner code is a generalization of LDPC code in which the "base code" can be more general than just checking parity. Let $C_0 \subseteq \{0,1\}^k$ be the base code. Let $G = (V, E)$ be a k-regular graph with $V = [n]$ and $E = [m]$. The Tanner code is defined as $C(G) := \{y \in \{0,1\}^m \mid y_{|\delta(i)} \in C_0 \ \forall i \in [n]\}$, where $y_{|\delta(i)}$ is the vector y restricted on the k edges in $\delta(i)$ for a vertex $i \in V$. That is, each bit $y(j)$ of a codeword is on an edge $j \in E$ of G, and a binary string y is a codeword if $y_{|\delta(i)}$ is a codeword of the base code C_0 for every vertex $i \in V$ of G.

The advantage of using Tanner code is that we could use a stronger base code with larger minimum distance, rather than just the parity check code with minimum distance only two. With a base code C_0 of minimum distance d_0, the requirement on the vertex expansion of G can be relaxed to k/d_0 to achieve the same distance as that of the corresponding LDPC code. In particular, because of Tanner's theorem in Theorem 7.9, one can simply use a spectral expander as G to design asymptotically good codes that are linear time encodable and decodable, without using lossless expanders. The decoding algorithm is still an iterative "fixing" algorithm where we replace an invalid codeword on a vertex by its nearest codeword. The analysis has a similar flavor that if the decoding algorithm fails, then one argues that there must be a "denser" subgraph than what is allowed by the expander mixing lemma.

The recent breakthroughs [DEL+21, PK21] in designing asymptotically good codes that are also locally testable is a generalization of Tanner codes on 2-dimensional expanders (where graphs are 1-dimensional expanders). Hope we will have some time to discuss it in the third part of the course when we study high-dimensional expanders.

7.4 References

[AB06] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2006. 67, 69

[AKS83] Miklós Ajtai, János Komlós, and Endre Szemerédi. An o(n log n) sorting network. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 25-27 April, 1983, Boston, Massachusetts, USA*, pages 1–9. ACM, 1983. 68

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Comb.*, 26(5):495–519, 2006. 59, 137, 138

[CKL13] Ho Yee Cheung, Tsz Chiu Kwok, and Lap Chi Lau. Fast matrix rank algorithms and applications. *J. ACM*, 60(5):31:1–31:25, 2013. 68

[CPT21] Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: a fourier-analytic approach. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1643–1655. ACM, 2021. 62

[DEL+21] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. *Electron. Colloquium Comput. Complex.*, page 151, 2021. 67, 72

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007. 69

[Gil98] David Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, 1998. 62

[GLSS18] Ankit Garg, Yin Tat Lee, Zhao Song, and Nikhil Srivastava. A matrix expander Chernoff bound. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1102–1114. ACM, 2018. 62

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 26, 29, 57, 60, 61, 62, 63, 67, 68, 137

[PK21] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. *CoRR*, abs/2111.03654, 2021. 67, 72

[Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):17:1–17:24, 2008. 68

[RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, 155(1):pp. 157–187, 2002. 63

[SS96] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inf. Theory*, 42(6):1710–1722, 1996. 70, 71

[Vad12] Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012. 61, 62, 67, 68

Chapter 8

Fastest Mixing and Vertex Expansion

We study a very recent result by Oleskar-Taylor and Zanetti [OZ21] relating the fastest mixing time to the vertex expansion of a graph, proving a surprising and beautiful Cheeger inequality for vertex expansion.

The fastest mixing time problem was proposed by Boyd, Diaconis and Xiao [BDX04]. In this problem, we are given an undirected graph $G = (V, E)$ and a target probability distribution $\pi : V \to \mathbb{R}$. The task is to assign a transition probability $P(u, v)$ on each edge $uv \in E(G)$, so that the stationary distribution of random walks with transition matrix P is π. The objective is to find a transition matrix P that minimizes the mixing time to π, among all transition matrices with stationary distribution π. We know from Chapter 6 that the mixing time to the stationary distribution is approximately inversely proportional to the spectral gap $1 - \alpha_2(P)$ of the transition matrix P, where $1 = \alpha_1(P) \geq \alpha_2(P) \geq \cdots \geq \alpha_{|V|}(P) \geq -1$ are the eigenvalues of P. The fastest mixing time problem is thus formulated in [BDX04] by the maximum spectral gap achievable through a "reweighting" P of the adjacency matrix of G.

Definition 8.1 (Maximum Reweighted Spectral Gap [BDX04])**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted spectral gap is defined as*

$$\begin{aligned} \lambda_2^*(G) \;:=\; \max_{P \geq 0} \quad & 1 - \alpha_2(P) & \\ \text{subject to} \quad & P(u, v) = 0 & \forall uv \notin E \\ & \sum_{v \in V} P(u, v) = 1 & \forall u \in V \\ & \pi(u)P(u, v) = \pi(v)P(v, u) & \forall uv \in E. \end{aligned}$$

The last constraint is called the time reversible condition, which is to ensure that the stationary distribution of P is π. Note that $\lambda_2^*(G) = \max_{P \geq 0}(1 - \alpha_2(P)) = \max_{P \geq 0} \lambda_2(I - P)$, which is the maximum reweighted second smallest eigenvalue of the normalized Laplacian matrix of G subject to the above constraints.

Boyd, Diaconis and Xiao showed that this optimization problem can be written as a semidefinite program and thus $\lambda_2^*(G)$ can be computed in polynomial time. Subsequently, the fastest mixing time problem has been studied in various work (see the references in [OZ21]), but no general characterization was known. Roch [Roc05] showed that the vertex expansion $\psi(G)$ defined in Definition 7.8 is an upper bound on the optimal spectral gap $\lambda_2^*(G)$. Very recently, Olesker-Taylor

and Zanetti [OZ21] proved that small vertex expansion is qualitatively the only obstruction for fastest mixing time to the uniform distribution.

Theorem 8.2 (Cheeger Inequality for Vertex Expansion [OZ21]). *For any undirected graph $G = (V, E)$ and the uniform distribution $\pi = \vec{1}/|V|$,*

$$\frac{\psi(G)^2}{\log |V|} \lesssim \lambda_2^*(G) \lesssim \psi(G).$$

In terms of the fastest mixing time $\tau^*(G)$ to the uniform distribution,

$$\frac{1}{\psi(G)} \lesssim \tau^*(G) \lesssim \frac{\log^2 |V|}{\psi^2(G)}.$$

Note the analogy to the Cheeger's inequality in Theorem 4.3, where spectral gap is replaced by maximum reweighted spectral gap and edge conductance is replaced by vertex expansion.

Unlike Cheeger's inequality for edge conductance where $\phi(G)^2 \lesssim \lambda_2(G) \lesssim \phi(G)$, it is noted in [OZ21] that the $\log |V|$ term may not be completely removed: Louis, Raghavendra and Vempala [LRV13] proved that there is no polynomial time algorithm that can distinguish between $\psi(G) \leq \epsilon$ and $\psi(G) \gtrsim \sqrt{\epsilon \log d}$ for every $\epsilon > 0$ where d is the maximum degree of the graph G, assuming the small-set expansion conjecture of Raghavendra and Steurer [RS10]. So, if the $\log |V|$ factor in Theorem 8.2 can be completely removed, then $\lambda_2^*(G)$ is a polynomial time computable quantity that can be used to distinguish between the two cases, disproving the small-set expansion conjecture.

Remark 8.3 (Uniform Distribution and Self-Loops). *We will make two assumptions about the problem. One is that the target distribution is the uniform distribution. Another is that the graph has a self-loop on each vertex, so that the problem in Definition 8.1 is always feasible. In the context of Markov chains, this corresponds to allowing a non-negative holding probability on each vertex.*

8.1 Dual Program for Fastest Mixing

To prove Theorem 8.2, Oleskar-Taylor and Zanetti use a dual minimization program obtained by Roch [Roc05] of the primal maximization program in Definition 8.1, to relate $\lambda_2^*(G)$ to the minimum vertex expansion of the input graph. We will use Von Neumann's minimax theorem to derive Roch's dual program.

Theorem 8.4 (Von Neumann's Minimax Theorem). *Let X, Y be compact convex sets. If f is a real-valued continuous function on $X \times Y$ with $f(x, \cdot)$ concave on Y for all $x \in X$ and $f(\cdot, y)$ convex on X for all $y \in Y$, then*

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y).$$

In the proof, we will use the following semidefinite program for computing the second eigenvalue, which is an extension of the spectral program in Lemma 4.4 to higher dimension but with the same optimal value.

Lemma 8.5 (Semidefinite Program for the Second Eigenvalue). *Let $P \in \mathbb{R}^{n \times n}$ be a reweighted matrix of a graph $G = (V, E)$ satisfying the constraints in Definition 8.1. Then*

$$1 - \alpha_2(P) = \min_{f: V \to \mathbb{R}^n,\ \sum_{v \in V} f(v) = 0} \frac{\sum_{uv \in E} \|f(u) - f(v)\|^2 \cdot P(u, v)}{\sum_{v \in V} \|f(v)\|^2}.$$

Proof. As explained in Definition 8.1, $1-\alpha_2(P) = \lambda_2(I-P)$ where $I-P$ is the normalized Laplacian matrix of the weighted graph P with weighted degree one for each vertex. By Lemma 4.4,

$$1-\alpha_2(P) = \min_{f:V\to\mathbb{R},\ \sum_{v\in V} f(v)=0} \frac{\sum_{uv\in E} |f(u)-f(v)|^2 \cdot P(u,v)}{\sum_{v\in V} f(v)^2},$$

which is almost the same as in the statement, except that $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$ as in the statement. Clearly, by considering all $f : V \to \mathbb{R}^n$, the feasible set could only be bigger and so the optimal value could only be smaller. On the other hand, given a solution $f : V \to \mathbb{R}^n$, by using the inequality

$$\min_{1\le i\le n} \frac{a_i}{b_i} \le \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$$

on the coordinates of $f : V \to \mathbb{R}^n$, we see that the best coordinate gives a one-dimensional solution $f : V \to \mathbb{R}$ with objective value as good as that of the n-dimensional solution $f : V \to \mathbb{R}^n$. To summarize, the relaxation from $f : V \to \mathbb{R}$ to $f : V \to \mathbb{R}^n$ is an exact relaxation. □

To see that it is a semidefinite program, recall that a positive semidefinite matrix Y can be written as F^TF where $F \in \mathbb{R}^{n\times n}$ by Fact 2.7. We associate each column v of F to $f(v)$, so that $Y_{u,v} = \langle f(u), f(v)\rangle$ for all $u, v \in V$. Then the above program can be rewritten as

$$\begin{aligned} \min \quad & \sum_{uv\in E} (Y_{u,u} - 2Y_{u,v} + Y_{v,v}) \cdot P(u,v) \\ \text{subject to} \quad & \sum_{v\in V} Y_{v,v} = 1 \\ & \sum_{u,v\in V} Y_{u,v} = 0 \\ & Y \succcurlyeq 0, \end{aligned}$$

where the objective function is the numerator in Lemma 8.5, the first constraint is normalizing the denominator in Lemma 8.5 to one, the second constraint is equivalent to the constraint $\sum_{v\in V} f(v) = 0$, and the last constraint is to ensure the correspondence $Y_{u,v} = \langle f(u), f(v)\rangle$ for all $u, v \in V$. So, the program in Lemma 8.5 can be written as optimizing a linear function with linear constraints on the entries of a positive semidefinite matrix Y, and this is a semidefinite program that can be solved in polynomial time.

The reason that we use the above semidefinite program for the second eigenvalue instead of the spectral program is that the set of feasible solutions is a convex set (while it is not the case for the spectral program), and this would allow us to apply the Von-Neumann minimax theorem to derive the following dual program by Roch.

Proposition 8.6 (Dual Program for Fastest Mixing [Roc05, OZ21])**.** *Given an undirected graph $G = (V, E)$ with a self-loop on each vertex and the uniform distribution $\pi = \vec{1}/|V|$ on V, the following semidefinite program is dual to the primal program in Definition 8.1 with strong duality*

$\lambda_2^*(G) = \gamma(G)$ *where*

$$\begin{aligned}
\gamma(G) := \min_{f:V\to\mathbb{R}^n,\ g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{v\in V} g(v) \\
\text{subject to} \quad & \sum_{v\in V} \|f(v)\|^2 = 1 \\
& \sum_{v\in V} f(v) = \vec{0} \\
& g(u) + g(v) \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E.
\end{aligned}$$

Proof. For a fixed P, by Lemma 8.5,

$$1 - \alpha_2(P) = \min_{f:V\to\mathbb{R}^n,\ \sum_{v\in V} f(v)=0} \frac{\sum_{uv\in E}\|f(u) - f(v)\|^2 \cdot P(u,v)}{\sum_{v\in V}\|f(v)\|^2}.$$

The maximum reweighted spectral gap in Definition 8.1 can thus be formulated as

$$\begin{aligned}
\lambda_2^*(G) = \max_{P\geq 0}(1-\alpha_2(P)) \quad = \quad & \max_{P\geq 0} \min_{f:V\to\mathbb{R}^n,\ \sum_{v\in V} f(v)=0} \frac{\sum_{uv\in E}\|f(u) - f(v)\|^2 \cdot P(u,v)}{\sum_{v\in V}\|f(v)\|^2} \\
\text{subject to} \quad & P(u,v) = 0 \quad \forall uv \notin E \\
& \sum_{v\in V} P(u,v) = 1 \quad \forall u \in V \\
& P = P^T.
\end{aligned}$$

Check that the assumptions in the Von Neumann minimax Theorem 8.4 are satisfied, and so we can switch the order of the max and the min and obtain the dual program

$$\gamma(G) \quad := \quad \min_{f:V\to\mathbb{R}^n,\ \sum_{v\in V} f(v)=0} \ \max_{P\geq 0} \frac{\sum_{uv\in E}\|f(u) - f(v)\|^2 \cdot P(u,v)}{\sum_{v\in V}\|f(v)\|^2},$$

subjected to the same constraints on P as above.

For a fixed $f : V \to \mathbb{R}^n$, note that the inner maximization problem is a linear program over the entries of P, and so we can reformuate it using LP duality to obtain

$$\begin{aligned}
\gamma(G) \quad = \quad & \min_{f:V\to\mathbb{R}^n,\ \sum_{v\in V} f(v)=0} \ \min_{g\geq 0} \sum_{v\in V} g(v) \\
\text{subject to} \quad & g(u) + g(v) \geq \frac{\|f(u) - f(v)\|^2}{\sum_{v\in V}\|f(v)\|^2} \quad \forall uv \in E,
\end{aligned}$$

where $g(u)$ is a dual variable for the constraint $\sum_{v\in V} P(u,v) = 1$. Note that the constraint $g \geq 0$ is from the assumption that there is a self-loop at each vertex. Normalizing so that $\sum_{v\in V}\|f(v)\|^2 = 1$ gives the statement. □

We remark that the self-loop assumption is to ensure that the dual program has the inequality $g \geq 0$. This is a crucial but subtle condition that will be used only once, and we will point it out when it is used.

One-Dimensional Dual Program and Random Projection

The first step in the proof of Theorem 8.2 is to project the solution $f : V \to \mathbb{R}^n$ to $\gamma(G)$ into a 1-dimensional solution $f : V \to \mathbb{R}$ as follows.

Definition 8.7 (One-Dimensional Dual Program for Fastest Mixing [OZ21])**.** *Given an undirected graph $G = (V, E)$, $\gamma^{(1)}(G)$ is defined to be the program:*

$$\begin{aligned} \gamma^{(1)}(G) := \min_{f:V\to\mathbb{R},\ g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{v\in V} g(v) \\ \text{subject to} \quad & \sum_{v\in V} f(v)^2 = 1 \\ & \sum_{v\in V} f(v) = 0 \\ & g(u) + g(v) \geq (f(u) - f(v))^2 \qquad \forall uv \in E. \end{aligned}$$

A very important result in metric embedding is the dimension reduction theorem by Johnson and Lindenstrauss, which says that n high-dimensional vectors can be projected into $O(\log n)$-dimensional vectors so that the pairwise Euclidean distances are approximately preserved.

Theorem 8.8 (Johnson-Lindenstrauss Lemma)**.** *Given $0 < \epsilon < 1$, a set X of n points in $\mathbb{R}^m$, there is a linear map $A : \mathbb{R}^m \to \mathbb{R}^k$ for $k \lesssim \ln(n)/\epsilon^2$ such that for all $u, v \in X$ it holds that*

$$(1-\epsilon)\|u-v\|_2^2 \leq \|Au - Av\|_2^2 \leq (1+\epsilon)\|u-v\|_2^2.$$

Apply the Johnson-Lindenstrauss lemma to the n-dimensional solution f in Proposition 8.6 to obtain a $O(\log n)$-dimensional solution f' with only constant distortion, and then use the "best" coordinate in f' as a solution to Definition 8.7, one can prove the following bounds between the two programs. Note that the $\log |V|$ factor in Theorem 8.2 is from this dimension reduction step.

Problem 8.9 (Dimension Reduction [OZ21])**.** *For any undirected graph G,*

$$\gamma(G) \leq \gamma^{(1)}(G) \lesssim \log |V(G)| \cdot \gamma(G).$$

The main step in the proof of Theorem 8.2 is the following Cheeger-type inequality between the 1-dimensional program in Definition 8.7 and the vertex expansion of the graph.

Theorem 8.10 (Cheeger Inequality for Vertex Expansion [OZ21])**.** *For any undirected graph G,*

$$\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \psi(G).$$

Combining Proposition 8.6 and Problem 8.9 and Theorem 8.10 gives

$$\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \log |V| \cdot \gamma(G) = \log |V| \cdot \lambda_2^*(G) \quad \text{and} \quad \lambda_2^*(G) = \gamma(G) \leq \gamma^{(1)}(G) \lesssim \psi(G),$$

proving Theorem 8.2. Henceforth, our goal is to prove Theorem 8.10, although we will need to do one more transformation described in the next section before getting to the main proof.

8.2 Matching Expansion

Instead of reasoning about the vertex expansion directly, Oleskar-Taylor and Zanetti defined an interesting new concept called the matching expansion, and showed that it is closely related to the vertex expansion and is easier to relate to the 1-dimensional dual program in Definition 8.7.

Definition 8.11 (Matching Expansion [OZ21])**.** *Let $G = (V, E)$ be an undirected weighted graph with a weight $w(e)$ on each edge $e \in E$. Given a subset of edges $F \subseteq E$, let the weight of a maximum matching in F be*

$$\nu(F) = \max_{\text{matching } M \subseteq F} \sum_{e \in M} w(e).$$

Define the matching expansion of a subset $S \subseteq V$ and of the graph as

$$\psi_\nu(S) = \frac{\nu(\delta(S))}{|S|} \quad \text{and} \quad \psi_\nu(G) = \min_{S : 0 < |S| \le |V|/2} \psi_\nu(S).$$

Note that while vertex expansion of a set in Definition 7.8 could be much larger than one, the matching expansion of a set is always at most one (in the case when $w(e) = 1$ for all $e \in E$), as is the edge conductance of a set in Definition 4.2. However, it can be shown that the vertex expansion of a graph is about the same as the matching expansion of a graph.

Problem 8.12 (Matching Expansion and Vertex Expansion)**.** *Let G be an undirected graph where every edge is of weight one. Then*

$$\psi_\nu(G) \le \psi(G) \le 4\psi_\nu(G).$$

The main technical work in [OZ21] is in proving the following Cheeger inequality for matching expansion.

Theorem 8.13 (Cheeger Inequality for Matching Expansion [OZ21])**.** *For any undirected graph G where every edge is of weight one,*

$$\psi_\nu(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \psi_\nu(G).$$

It should be clear that Problem 8.12 and Theorem 8.13 imply Theorem 8.10, which in turn implies Theorem 8.2. Our goal is then to prove Theorem 8.13.

Maximum Matching, Auxiliary Directed Graphs, and Directed Matching

The intuition that matching is relevant to the problem is from the constraints in the 1-dimensional dual program in Definition 8.7. The following lemma follows from weak duality of linear programming and is easy to see directly. This is the only place that the constraint $g \ge 0$ is used crucially, so pay attention when the following lemma is used in the main proof.

Lemma 8.14 (Matching and Vertex Cover)**.** *Let $G = (V, E)$ be an undirected graph with a weight $w(e)$ on each edge $e \in E$. The weighted matching number $\nu(E)$ is upper bounded by*

$$\begin{aligned} & \sum_{v \in V} g(v) & \\ \text{subject to} \quad & g(u) + g(v) \ge w(u, v) & \forall uv \in E \\ & g(v) \ge 0 & \forall v \in V. \end{aligned}$$

From this perspective, a good way to interpret the solution to Definition 8.7 is that g is a weighted fractional vertex cover when each edge has weight $(f(u) - f(v))^2$. The following auxiliary graphs will be used in the main proof.

Definition 8.15 (Auxiliary Graphs). *Let $G = (V, E)$ be an undirected graph and f, g be a solution to the program in Definition 8.7. Define G_f to be the weighted undirected graph where each edge uv in E has weight $|f(u)^2 - f(v)^2|$. Define $\overrightarrow{G_f}$ to be the orientation of G_f where there is a directed edge uv with weight $f(u)^2 - f(v)^2$ if and only if $uv \in E(G)$ and $f(u) > f(v)$.*

A matching in an undirected graph is a subgraph in which every vertex is of degree at most one. The following analog of directed matching will be used in the proof.

Definition 8.16 (Directed Matching). *Given a directed graph $\overrightarrow{G} = (V, \overrightarrow{E})$ with a weight $w(e)$ on each edge $e \in \overrightarrow{E}$, a subset of edges $\overrightarrow{F} \subseteq \overrightarrow{E}$ is a directed matching if the indegree and the outdegree of each vertex in $\overrightarrow{F}$ is at most one. Let $\nu(\overrightarrow{E})$ be the maximum weight of a directed matching in $\overrightarrow{E}$.*

A simple combinatorial argument shows that the maximum weight of an undirected matching is within a constant factor of the maximum weight of a directed matching.

Exercise 8.17. *For any edge-weighted graph $G = (V, E)$ and any orientation $\overrightarrow{G} = (V, \overrightarrow{E})$,*

$$\nu(E) \leq \nu(\overrightarrow{E}) \leq 4\nu(E).$$

8.3 Cheeger Inequality for Matching Expansion

As in the proof of Cheeger's inequality in Theorem 4.3, one direction is the easy direction where we see that $\gamma^{(1)}(G)$ is a relaxation for the matching expansion $\psi_\nu(G)$, and another direction is the hard direction where we round a fractional solution to $\gamma^{(1)}(G)$ to obtain an integral solution to $\psi_\nu(G)$.

Easy Direction

There are two ways to see the easy direction. One way is to plug in a binary solution defined by a set S minimizing the matching expansion to upper bound $\gamma^{(1)}(G)$.

Proposition 8.18 (Easy Direction for Matching Expansion [OZ21]). *For any undirected graph G where every edge is of weight one,*

$$\gamma^{(1)}(G) \lesssim \psi_\nu(G).$$

Proof. Given $S \subseteq V$, plug in

$$f(v) = \frac{1}{|S|} \frac{\sqrt{|S||V-S|}}{\sqrt{|S| + |V-S|}} \text{ for } v \in S, \quad \text{and } f(v) = -\frac{1}{|V-S|} \frac{\sqrt{|S||V-S|}}{\sqrt{|S| + |V-S|}} \text{ otherwise.}$$

Let M be a maximum matching in $\delta(S)$. Set $g(v) = 2/|S|$ for $v \in M$ and $g(v) = 0$ otherwise. Check that it is a feasible solution to $\gamma^{(1)}(G)$ in Definition 8.7 with objective value at most $4\psi_\nu(S)$. □

Another way is to understand the easy direction of Theorem 8.2 directly, as to use the edge conductance of a reweighted graph H of G to certify the vertex expansion of the input graph G.

Proposition 8.19 (Vertex Expansion through Edge Expansion)**.** *Let H be an edge-reweighted graph of $G = (V, E)$ with weighted adjacency matrix P satisfying the constraints in the primal program in Definition 8.1. Then $\phi(H) \leq \psi(G)$ where $\phi(H)$ is the weighted edge conductance of H.*

Proof. As the reweighted matrix P satisfies the constraints in Definition 8.1, the graph H is a weighted 1-regular graph and so its weighted edge conductance is simply

$$\phi(H) = \min_{S:0<\mathrm{vol}_w(S)\leq\frac{1}{2}\mathrm{vol}_w(V)} \frac{w(\delta(S))}{\mathrm{vol}_w(S)} = \min_{S:0<|S|\leq\frac{1}{2}|V|} \frac{w(\delta(S))}{|S|},$$

where we denote $w(u,v) = P(u,v)$ as the weight of an edge and $w(\delta(S)) = \sum_{e\in\delta(S)} w(e)$. Observe the important point that $|\partial(S)| \geq w(\delta(S))$, because each edge in $\delta(S)$ has an endpoint in $\partial(S)$ and each vertex in $\partial(S)$ has weighted degree one, and so $|\partial(S)| = \sum_{v\in\partial(S)} \deg_w(v) \geq w(\delta(S))$. Therefore,

$$\phi(H) = \min_{S:0<|S|\leq\frac{1}{2}|V|} \frac{w(\delta(S))}{|S|} \leq \min_{S:0<|S|\leq\frac{1}{2}|V|} \frac{|\partial(S)|}{|S|} = \psi(G).$$

□

By Proposition 8.19, the edge conductance of any edge reweighted graph H of G satisfying the constraints in Definition 8.1 is a lower bound on the vertex expansion of G. To prove the best lower bound on the vertex expansion of G, we thus maximum the edge conductance of an edge reweighted graph H. Note that the edge conductance of the reweighted graph H is lower bounded by the spectral gap of the reweighted matrix P by the easy direction of Cheeger's inequality in Theorem 4.3. Therefore,

$$\lambda_2^*(G) = \max_{H:H \text{ is a reweighting of } G} \lambda_2(H) \leq \max_{H:H \text{ is a reweighting of } G} 2\phi(H) \leq 2\psi(G).$$

To summarize, a good way to understand the easy direction of the new Cheeger inequality for vertex expansion in Theorem 8.2 is that it is a way to certify the vertex expansion of a graph through a reduction to the edge conductance and spectral gap of a reweighted graph. Very interestingly, the hard direction proves that there is always a reweighted graph so that this reduction works well to certify the vertex expansion.

Hard Direction

The structure of the hard direction is similar to that for Cheeger's inequality in Chapter 4. The first step is a truncation step to ensure that the output set S satisfies $|S| \leq |V|/2$. Again, as in the truncation step in Lemma 4.6 in the hard direction of Cheeger's inequality, the condition $\sum_{v\in V} f(v) = 0$ is used to trade for the non-negativity condition and the support-size condition.

Problem 8.20 (Truncation)**.** *Let $G = (V, E)$ be an undirected graph and $\pi = \vec{1}/|V|$ be the uniform distribution. Given a solution f, g to $\gamma^{(1)}(G)$ in Definition 8.7, there is a solution x, y with $x \geq 0$ and $y \geq 0$ and $|\mathrm{supp}(x)| \leq |V|/2$ such that*

$$\sum_{v\in V} y(v) \lesssim \gamma^{(1)}(G)$$

$$\sum_{v\in V} x(v)^2 = 1$$

$$y(u) + y(v) \geq (x(u) - x(v))^2 \qquad \forall uv \in E.$$

Again, the main step is to apply threshold rounding on the solution in Problem 8.20 to find a set S with small matching expansion.

Proposition 8.21 (Hard Direction for Matching Expansion [OZ21]). *Let $G = (V, E)$ be an undirected graph and $\pi = \vec{1}/|V|$ be the uniform distribution. Given a solution x, y satisfying the conditions in Problem 8.20, there is a set $S \subseteq \mathrm{supp}(x)$ with $\psi_\nu(S) \lesssim \sqrt{\gamma^{(1)}(G)}$.*

Proof. Let $S_t := \{v \in V \mid x(v)^2 > t\}$ be a level set for $t \geq 0$. Choose t uniform randomly, Trevisan's argument implies that

$$\min_t \psi_\nu(S_t) \leq \frac{\int_0^\infty \nu(\delta(S_t))dt}{\int_0^\infty |S_t|dt},$$

so that we can compute the numerator and the denominator separately.

The denominator is

$$\int_0^\infty |S_t|dt = \int_0^\infty \sum_{v\in V} \mathbb{1}(v \in S_t)dt = \sum_{v\in V} \int_0^\infty \mathbb{1}(x(v)^2 > t)dt = \sum_{v\in V} x(v)^2 = 1.$$

To bound the numerator, Oleskar-Taylor and Zanetti consider the auxiliary graphs G_x and $\overrightarrow{G_x}$ in Definition 8.15. Using some combinatorial arguments about matchings, they proved a key lemma in Lemma 8.22 that

$$\int_0^\infty \nu(\delta(S_t))dt \leq 8\nu(G_x).$$

Assuming Lemma 8.22, let M be a maximum weighted matching in G_x, we further bound $\nu(G_x)$ by standard Cauchy-Schwarz manipulation so that

$$\begin{aligned}
\nu(G_x) &= \sum_{uv\in M} \big|x(u)^2 - x(v)^2\big| \\
&= \sum_{uv\in M} \big|x(u) - x(v)\big| \cdot \big|x(u) + x(v)\big| \\
&\leq \sqrt{\sum_{uv\in M} \big(x(u) - x(v)\big)^2} \sqrt{\sum_{uv\in M} \big(x(u) + x(v)\big)^2} \\
&\leq \sqrt{\sum_{uv\in M} \big(x(u) - x(v)\big)^2} \sqrt{\sum_{v\in V} 2x(v)^2} \\
&= \sqrt{2\sum_{uv\in M} \big(x(u) - x(v)\big)^2}.
\end{aligned}$$

where the last inequality holds because M is a matching so that each vertex is of degree one in M.

Next, we use the weak duality between matching and vertex cover stated in Lemma 8.14 to relate the RHS to the solution y in Problem 8.20, so that

$$\sum_{uv\in M} \big(x(u) - x(v)\big)^2 \leq \sum_{v\in V} y(v) \lesssim \gamma^{(1)}(G),$$

where the first inequality is where the constraint $y \geq 0$ is crucially used. Therefore, we conclude that

$$\min_t \psi_\nu(S_t) \leq \frac{\int_0^\infty \nu(\delta(S_t))dt}{\int_0^\infty |S_t|dt} \leq 8\nu(G_x) \leq 8\sqrt{2\sum_{uv\in M} \big(x(u) - x(v)\big)^2} \lesssim \sqrt{\gamma^{(1)}(G)}.$$

□

It remains to prove the key lemma about the numerator. Let M_t be a maximum matching in $\delta(S_t)$. Then $\int_0^\infty \nu(\delta(S_t))dt = \int_0^\infty |M_t|dt$. The main idea is to prove that a greedy directed matching $\overrightarrow{M}$ in $\overrightarrow{G_x}$ satisfies $|\overrightarrow{M} \cap \delta(S_t)| \geq \frac{1}{2}|M_t|$ for every t. That is, the fixed matching $\overrightarrow{M}$ is almost as good as the maximum matching in every threshold set S_t.

Lemma 8.22. *Let $G = (V, E)$ be an undirected graph and $\pi = \vec{1}/|V|$ be the uniform distribution. Given a solution x, y satisfying the conditions in Problem 8.20, let $S_t := \{v \in V \mid x(v)^2 > t\}$ for $t \geq 0$, then*

$$\int_0^\infty \nu(\delta(S_t))dt \leq 8\nu(G_x).$$

Proof. Using the definitions of G_x and $\overrightarrow{G_x}$ in Definition 8.15, we will prove that $\int_0^\infty \nu(\delta(S_t))dt \leq 2\nu(\overrightarrow{G_x})$, and then the lemma follows from Exercise 8.17.

To prove $\int_0^\infty \nu(\delta(S_t))dt \leq 2\nu(\overrightarrow{G_x})$, we consider a greedy directed matching $\overrightarrow{M}$ in $\overrightarrow{G_f}$, which is obtained by sorting the directed edges in non-increasing order of weights and greedily adding edges to the directed matching whenever possible.

Let M_t be a maximum matching in $\delta(S_t)$. Note that there could be a different maximum matching for each t. The key observation is that the fixed greedy matching $\overrightarrow{M}$ satisfies $|\overrightarrow{M} \cap \delta(S_t)| \geq |M_t|/2$ for each t. To see this, let uv be an edge in M_t with $x(u) > x(v)$. Suppose $uv \notin \overrightarrow{M}$. Since $\overrightarrow{M}$ is a greedy directed matching, when uv was considered and was not added to $\overrightarrow{M}$, then either u has outdegree one or v has indegree one at that time, as otherwise we could add the edge uv to $\overrightarrow{M}$. In either case, say $uw \in \overrightarrow{M}$, then it must hold that $x(u)^2 - x(w)^2 \geq x(u)^2 - x(v)^2$, as the edges are considered in a non-increasing order of weights. Since uw is at least as long as uv, the edge uw is in every threshold cut that uv is in, and so $uw \in \delta(S_t) \cap \overrightarrow{M}$. Using this argument, we can map each edge $uv \in M_t$ to some other edge in $\delta(S_t) \cap \overrightarrow{M}$ sharing an endpoint with uv. Crucially, since M_t is a matching, each edge in $\overrightarrow{M}$ is mapped by at most two edges in M_t, one for each endpoint. This establishes the claim that $|\overrightarrow{M} \cap \delta(S_t)| \geq |M_t|/2$. Then, we can conclude that

$$\int_0^\infty \nu(\delta(S_t))dt = \int_0^\infty |M_t|dt \leq 2\int_0^\infty |\overrightarrow{M} \cap \delta(S_t)|dt = 2\sum_{uv \in \overrightarrow{M}} \left(x(u)^2 - x(v)^2\right) \leq 2\nu(\overrightarrow{G_x}).$$

□

Summary and Discussions

Starting from the primal program $\lambda_2^*(G)$ in Definition 8.1, we construct the dual program $\gamma(G)$ in Proposition 8.6 using von-Neumann minimax theorem. Then we use the Johnson-Lindenstrass lemma to reduce an n-dimensional solution to $\gamma(G)$ to a 1-dimensional solution to $\gamma^{(1)}(G)$ in Definition 8.7, where the $\log |V|$ factor in Theorem 8.2 is from this step. Then we consider the matching expansion $\psi_\nu(G)$ in Definition 8.11 as a proxy to the vertex expansion $\psi(G)$ in Definition 7.8, and reduce the Cheeger inequality for vertex expansion in Theorem 8.2 to the Cheeger inequality for matching expansion in Theorem 8.13. The easy direction of Theorem 8.13 can be proved by plugging in a binary solution from matching expansion. Also, there is a good way to understand the easy direction of Theorem 8.2 as a reduction from vertex expansion to the edge conductance of the

best reweighted graph. The hard direction of Theorem 8.13 is proved by a truncation step and a threshold rounding step as in the proof for Cheeger's inequality in Theorem 4.3. A key Lemma 8.22 in the hard direction is proved by a combinatorial argument about greedy directed matching. After finding a set of small matching expansion, we can use Problem 8.12 to find a set of small vertex expansion.

Oleskar-Taylor and Zanetti left open the problem of reducing the $\log |V|$ factor in Theorem 8.2 to $\log d$ where d is the maximum degree of the input graph, and the problem of generalizing Theorem 8.2 to arbitrary target probability distribution π. It would also be interesting to construct an example where Theorem 8.2 is nearly tight.

8.4 Problems

Problem 8.23 (λ_∞ and Symmetric Vertex Expansion [BHT00])**.** *Bobkov, Houdré and Tetali defined an interesting quantity*

$$\lambda_\infty(G) := \min_{x:V\to\mathbb{R},\ x\perp\vec{1}} \frac{\sum_{u\in V} \max_{v:(v,u)\in E} (x(u)-x(v))^2}{\sum_{u\in V} x(u)^2}$$

and prove an analog of Cheeger's inequality that

$$\Phi^V(G)^2 \lesssim \lambda_\infty(G) \lesssim \Phi^V(G),$$

where

$$\Phi^V(S) := |V| \cdot \frac{|\partial(S)\cup\partial(V-S)|}{|S|\cdot|V-S|} \quad \text{and} \quad \Phi^V(G) := \min_{S\subset V} \Phi^V(S)$$

is called the symmetric vertex expansion of the graph. Give a proof of their theorem.

8.5 References

[BDX04] Stephen P. Boyd, Persi Diaconis, and Lin Xiao. Fastest mixing markov chain on a graph. *SIAM Rev.*, 46(4):667–689, 2004. 75

[BHT00] Sergey G. Bobkov, Christian Houdré, and Prasad Tetali. λ_∞ vertex isoperimetry and concentration. *Comb.*, 20(2):153–172, 2000. 85

[LRV13] Anand Louis, Prasad Raghavendra, and Santosh S. Vempala. The complexity of approximating vertex expansion. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 360–369. IEEE Computer Society, 2013. 76

[OZ21] Sam Olesker-Taylor and Luca Zanetti. Geometric bounds on the fastest mixing markov chain. *CoRR*, abs/2111.05816, 2021. 75, 76, 77, 79, 80, 81, 83

[Roc05] Sébastien Roch. Bounding Fastest Mixing. *Electronic Communications in Probability*, 10(none):282 – 296, 2005. 75, 76, 77

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 755–764. ACM, 2010. 37, 76

Chapter 9

Spectral Sparsification

In this chapter, we introduce the spectral sparsification problem formulated by Spielman and Teng [ST11], which is a generalization of the graph sparsification problem formulated by Karger [Kar99]. Then we will see a random sampling algorithm to solve the problem by Spielman and Srivastava [SS11], matching the result for the graph sparsification problem by Benczur and Karger [BK15].

As we will see in later chapters, the study of spectral sparsification has led to major breakthroughs, and this is a striking example of using linear algebraic techniques to solve combinatorial problems.

9.1 Graph Sparsification

The graph sparsification problem is to find a sparse graph which approximates all cut values of a given graph.

Definition 9.1 (Cut Approximator [Kar99])**.** *Let $G = (V, E)$ be an undirected graph with a weight $w_G(e)$ on each edge $e \in E$, and $H = (V, F)$ be an undirected graph on the same vertex set with a weight $w_H(e)$ on each edge $e \in F$. For $0 \leq \epsilon \leq 1$, we say H is a $(1 \pm \epsilon)$-cut approximator of G if for all $S \subseteq V$,*

$$(1 - \epsilon) \cdot w_G\big(\delta_G(S)\big) \leq w_H\big(\delta_H(S)\big) \leq (1 + \epsilon) \cdot w_G\big(\delta_G(S)\big).$$

This problem was formulated by Karger [Kar99], and the goal is to find a sparse graph H that is a good cut approximator of the input graph G. Note that this definition does not require that H is a subgraph of G (that is, $F \subseteq E$), but all constructions that we will see satisfy this property which is useful in some applications.

Uniform Sampling

A first example to think about is when G is a complete graph. We know from Chapter 7 that a random sparse graph H is an expander graph, which is a good approximation to the complete graph. So it is a natural strategy to construct a sparsifier H by sampling a uniform random subgraph of G. Karger considered the following simple uniform random sampling algorithm, where the idea is that the expected weight of each edge e in H is the same as the weight of e in G.

Algorithm 5 Uniform Sampling Algorithm for Graph Sparsification

Require: An unweighted undirected graph $G = (V, E)$.

1: Set a sampling probability p. For each $e \in E$, with probability p, add e in F with weight $1/p$.
2: **return** $H = (V, F)$.

Karger proved that the uniform sampling algorithm would work to sparsify the input graph G when the minimum cut value of G is $\Omega(\log n)$.

Theorem 9.2 (Uniform Sampling for Graph Sparsification [Kar99]). *Let $G = (V, E)$ be an unweighted undirected graph with $V = [n]$ and minimum cut value c. Set the sampling probability $p = \frac{9 \ln n}{\epsilon^2 c}$. Then H produced by Algorithm 5 is a $(1 \pm \epsilon)$-cut approximator of G with $O(p \cdot |E|)$ edges with probability at least $1 - \frac{4}{n}$.*

The well-known Chernoff bound is used to analyze the success probability.

Theorem 9.3 (Chernoff Bound for Heterogeneous Coin Flips). *Let $X_1, X_2, \ldots, X_n$ be independent random variables with $X_i = 1$ with probability p_i and $X_i = 0$ otherwise. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p_i$ be the expected value of X. Then, for any $0 < \delta < 1$,*

$$\Pr\left(|X - \mu| \geq \delta\mu\right) \leq 2e^{-\delta^2 \mu/3}.$$

The proof outline of Theorem 9.2 is as follows. With the assumption that the minimum cut value is $\Omega(\log n)$, Chernoff bound can be used to show that the probability that $w_H(\delta_H(S))$ is not a $(1 \pm \epsilon)$-approximation of $w_G(\delta_G(S))$ for a particular subset $S \subseteq V$ is at most $1/\operatorname{poly}(n)$. While this probability is quite small, this is not nearly small enough to apply a union bound on the exponential number of subsets directly. Karger's observation is that there are only a polynomial number of small cuts as stated below, and so a more careful union bound based on the cut value can be used to prove Theorem 9.2.

Proposition 9.4 (Number of Approximate Minimum Cuts). *Let $G = (V, E)$ be an unweighted undirected graph with $V = [n]$ and minimum cut value c. For any $\alpha \geq 1$, the number of subsets S with $|\delta(S)| \leq \alpha c$ is at most $n^{\lceil 2\alpha \rceil}$.*

An interesting way to prove Proposition 9.4 is to use Karger's random contraction algorithm for solving the minimum cut problem. See [Kar99] or L03/L04 of CS761 for proofs of Theorem 9.3 and Proposition 9.4.

Non-Uniform Sampling

Without the minimum cut assumption, then it is easy to see that the uniform sampling algorithm could fail. For example, consider the dumbbell graph where there is a bridge connecting two complete graphs.

In 1996, Benczur and Karger [BK15] designed a very clever non-uniform sampling algorithm, where the sampling probability p_e for each edge $e = uv$ is proportional to the "connectivity" of u and v. The idea is that edges with low connectivity are in H with higher probability p_e and smaller weight $1/p_e$ because they are crucial and so we basically just keep them (with the right expectation), while edges with high connectivity are in H with lower probability and larger weight to sparsify the graph. They defined a notion called "strong connectivity" for the non-uniform sampling algorithm and proved that every graph has a cut approximator with only $O(n \log n)$ edges.

Theorem 9.5 (Benczur-Karger Cut Sparsification [BK15]). *For any edge-weighted undirected graph $G = (V, E)$ and any $0 < \epsilon < 1$, there is a reweighted subgraph $H = (V, F)$ on the same vertex set with at most $O(\frac{n \log n}{\epsilon^2})$ edges such that H is a $(1 \pm \epsilon)$-cut approximator of G. Furthermore, H can be computed in nearly linear time $\tilde{O}(|E|)$.*

The definition of "strong connectivity" is a bit unnatural, and Benczur and Karger conjecured that it can be replaced by the more natural edge-connectivity between u and v. This conjecture is proved by Fung, Hariharan, Harvey and Panigrahi in 2011 [FHHP19].

Applications of Graph Sparsifications

An important feature of Theorem 9.2 and Theorem 9.5 is that they provide a near-linear time algorithm to find a cut sparsifer, and they become an important primitive in designing fast graph algorithms. For example, suppose we would like to solve the minimum s-t cut problem in a graph G. Standard algorithms have their time complexity depending on the number of edges in G, so when G is dense with $\Omega(n^2)$ edges the algorithms are quite slow. To design a fast approximation algorithm, we can first use Theorem 9.5 to obtain a $(1 \pm \epsilon)$-cut approximator H of G with only $O(n \log n/\epsilon^2)$ edges. Then, we just run the standard algorithms on H to find an optimal s-t cut in H, and it can be shown that this is a $(1 + 3\epsilon)$-approximate minimum s-t cut in G. More generally, with these sparsification algorithms, for many graph problems about cuts (e.g. graph conductance), one could trade a small loss in the optimality of the solutions for a time complexity that is faster by at least one order of n.

Truly remarkably, Karger [Kar00] used the uniform sampling algorithm in Theorem 9.2 to design a near-linear time algorithm to solve the minimum cut problem *optimally*. It is actually crucial that the sparsifier is unweighted for this application, so for example the stronger Theorem 9.5 cannot be used. It took more than 20 years for researchers to finally find a deterministic near-linear time algorithm for the minimum cut problem [KT19].

9.2 Spectral Sparsification

On their way of designing a near-linear time algorithm for solving Laplacian systems of linear equations, Spielman and Teng [ST11] defined the following stronger notion of spectral sparsification for Laplacian matrices.

Definition 9.6 (Spectral Approximator). *Let $G = (V, E)$ be a weighted undirected graph and $H = (V, F)$ be a weighted undirected graph on the same vertex set. For $0 \leq \epsilon \leq 1$, we say H is a $(1 \pm \epsilon)$-spectral approximator of G if for all $x : V \to \mathbb{R}$,*

$$(1 - \epsilon) \cdot x^T L_G x \leq x^T L_H x \leq (1 + \epsilon) \cdot x^T L_G x,$$

where L_G and L_H are the weighted Laplacian matrices of G and H respectively. Equivalently, H is a $(1 \pm \epsilon)$-spectral approximator of G if

$$(1 - \epsilon) L_G \preccurlyeq L_H \preccurlyeq (1 + \epsilon) L_G.$$

Exercise 9.7 (Spectrum of Spectral Approximator). *Let G and H be weighted undirected graphs with Laplacian spectrums $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ and $\gamma_1 \leq \gamma_2 \leq \ldots \leq \gamma_n$ respectively. Prove that if H is a $(1 \pm \epsilon)$-spectral approximator of G, then $(1 - \epsilon)\lambda_i \leq \gamma_i \leq (1 + \epsilon)\lambda_i$ for every $1 \leq i \leq n$.*

Again, the goal is to find a sparse graph H that is a good spectral approximator of the input graph G. Their original motivation is to use L_H as a "preconditioner" for solving the equations $L_G \cdot z = b$. Their definition is inspired by the result of Benczur and Karger [BK15] and is indeed a more demanding one.

Lemma 9.8 (Spectral Approximator is Cut Approximator). *If H is a $(1 \pm \epsilon)$-spectral approximator of G, then H is a $(1 \pm \epsilon)$-cut approximator of G.*

Proof. Let S be a subset of vertices and χ_S be the characteristic vector of S. Then, by Lemma 3.17,

$$\chi_S^T L_G \chi_S = \sum_{ij \in E(G)} w(i,j)\big(\chi_S(i) - \chi_S(j)\big)^2 = w_G\big(\delta_G(S)\big),$$

and similarly $\chi_S^T L_H \chi_S = w_H\big(\delta_H(S)\big)$. Since H is a $(1 \pm \epsilon)$-spectral approximator of G, it follows that

$$(1-\epsilon)\cdot\chi_S^T L_G \chi_S \le \chi_S^T L_H \chi_S \le (1+\epsilon)\cdot\chi_S^T L_G \chi_S \implies (1-\epsilon)\cdot w_G\big(\delta_G(S)\big) \le w_H\big(\delta_H(S)\big) \le (1+\epsilon)\cdot w_G\big(\delta_G(S)\big).$$

As this holds for any subset $S \subseteq V$, H is a $(1 \pm \epsilon)$-cut approximator of G. □

Since spectral sparsification is a strictly stronger requirement than cut sparsification, one would expect that it is a strictly harder problem to solve. Initially, Spielman and Teng [ST11] proved that there is always a $(1 \pm \epsilon)$-spectral sparsifier with $O(n \operatorname{polylog}(n)/\epsilon^2)$ edges and gave a fast algorithm for constructing such sparsifiers. This is enough for their grand goal of designing a nearly-linear time algorithm for solving Laplacian equations, which has become the engine for a new generation of fast algorithms for graph problems.

Reduction

In 2008, Spielman and Srivastava [SS11] revisited the spectral sparsification problem and proved that there is always a $(1 \pm \epsilon)$-spectral approximator with $O(n \log n/\epsilon^2)$ edges, thus by Lemma 9.8 generalizing the result of Benczur and Karger in Theorem 9.5 for cut sparsification. They reduced it to the following simpler statement where the objective is to bound only the maximum eigenvalue and the minimum eigenvalue.

Theorem 9.9 (Sparse Spectral Approximator of Identity Matrix [SS11]). *For any m vectors $u_1, \ldots, u_m \in \mathbb{R}^n$ satisfying $\sum_{i=1}^m u_i u_i^T = I_n$, there always exist scalars $s_1, \ldots, s_m$ with at most $O(\frac{n \log n}{\epsilon^2})$ nonzeros such that*

$$(1-\epsilon) I_n \preccurlyeq \sum_{i=1}^m s_i u_i u_i^T \preccurlyeq (1+\epsilon) I_n.$$

For the reduction, we need the concept of pseudoinverse of a matrix in Definition 2.23.

Definition 9.10 (Pseudoinverse of Laplacian Matrix). *Let G be a connected graph. Let $0 = \lambda_1 < \lambda_2 \le \ldots \le \lambda_n$ be the eigenvalues of $L(G)$ and $u_1, \ldots, u_n$ be the corresponding eigenvectors. Then the pseudoinverse of $L(G)$ is $L_G^\dagger = \sum_{i=2}^n \frac{1}{\lambda_i} u_i u_i^T$. And the square root of $L_G^\dagger$ is $L_G^{\dagger/2} = \sum_{i=2}^n \frac{1}{\sqrt{\lambda_i}} u_i u_i^T$*

Lemma 9.11 (Reduction to Identity). *Suppose for any m vectors with $\sum_{i=1}^m v_i v_i^T = I_n$ there are always scalars with at most $O(n \log n/\epsilon^2)$ nonzeros such that $(1-\epsilon)I_n \preccurlyeq \sum_{i=1}^m s_i v_i v_i^T \preccurlyeq (1+\epsilon)I_n$. Then for any graph $G = (V, E)$ with n vertices and m edges, there is always a graph H with $O(n \log n/\epsilon^2)$ edges such that H is a $(1 \pm \epsilon)$-spectral approximator of G.*

Proof. We assume without loss of generality that G is a connected graph. Let $L_G = \sum_{e \in E} b_e b_e^T$ be the Laplacian matrix of G as written in Definition 3.15. Define $v_e = U^T L_G^{\dagger/2} b_e$ where $L_G^{\dagger/2}$ is from Definition 9.10 and U is the $n \times (n-1)$ matrix where the i-th column is the $(i+1)$-th eigenvector u_{i+1} of $L(G)$ for $1 \le i \le n-1$. Then

$$\sum_{e \in E} v_e v_e^T = \sum_{e \in E} U^T L_G^{\dagger/2} b_e b_e^T L_G^{\dagger/2} U = U^T L_G^{\dagger/2} L_G L_G^{\dagger/2} U = U^T \left(\sum_{i=2}^n u_i u_i^T \right) U = I_{n-1}.$$

By assumption, there are scalars $s_1, \ldots, s_m$ with at most $O(n \log n/\epsilon^2)$ nonzeros such that

$$(1-\epsilon) I_{n-1} \preccurlyeq \sum_{e \in E} s_e v_e v_e^T \preccurlyeq (1+\epsilon) I_{n-1}.$$

Multiplying $L_G^{1/2} U$ on the left and $U^T L_G^{1/2}$ on the right of the above inequalities, then $L_G^{1/2} U U^T L_G^{1/2} = L_G$ and thus

$$(1-\epsilon) L_G \preccurlyeq \sum_{e \in E} s_e b_e b_e^T \preccurlyeq (1+\epsilon) L_G.$$

Let H be the graph with weight s_e on edge e. Then $L_H = \sum_{e \in E} s_e b_e b_e^T$, and thus H is a $(1 \pm \epsilon)$-spectral approximator of G with $O(n \log n/\epsilon^2)$ edges. □

Random Sampling Algorithm

Isotropy Condition: We first get some intuition about the condition $\sum_{i=1}^m v_i v_i^T = I_n$. We have discussed this isotropy condition before in Exercise 5.14 when we studied the higher-order Cheeger's inequality. When $m = n$, then $v_1, \ldots, v_n$ must be an orthonormal basis. When $m > n$, we can think of $v_1, \ldots, v_m$ as an "overcomplete" basis, as we can write any $x \in \mathbb{R}^n$ as $x = I_n x = \left(\sum_{i=1}^m v_i v_i^T\right) x = \sum_{i=1}^m \langle x, v_i \rangle v_i$, which has applications in communication theory. Also, as stated in Exercise 5.14, for any unit vector $y \in \mathbb{R}^n$, it holds that $\sum_{i=1}^m \langle y, v_i \rangle^2 = 1$. Informally, the vectors are "evenly spread out" so that the projections of these vectors to any direction y are the same. Given $\sum_{i=1} v_i v_i^T = I_n$, we would like to find a small subset of vectors $S \subseteq \{1, \ldots, m\}$ and some scaling factors so that $\sum_{i \in S} s_i v_i v_i^T \approx I_n$, and thus $\sum_{i \in S} s_i \langle y, v_i \rangle^2 \approx 1$. So, the subset should still be "evenly spread out", with the contribution in each direction about the same.

Idea: As in the cut sparsification case, uniform sampling may not work. For example, if some vector v_j has $\|v_j\| = 1$, then we must include v_j in the solution, as otherwise that direction will not be covered in the solution and so it won't be a spectral sparsifier. The analogy in the cut sparsification result is that a cut edge must be included in any cut sparsifier. So, as in the cut sparsification case, we need to do non-uniform sampling if we do random sampling.

The idea is similar and very natural. For longer vectors, we should set the sampling probability p_e to be higher because they are crucial and so we basically just keep them. For shorter vectors, we can afford to set the sampling probability p_e to be lower and the weight $1/p_e$ to be larger in order

to reduce the number of vectors. Concretely, we sample each vector v_i with probability $\|v_i\|_2^2$, and if it is chosen, we set the scalar $s_i = \frac{1}{\|v_i\|_2^2}$, so that the expected contribution is

$$\mathbb{E}\left[s_i v_i v_i^T\right] = \frac{v_i v_i^T}{\|v_i\|^2} \cdot \Pr[v_i \text{ is chosen}] = \frac{v_i v_i^T}{\|v_i\|^2} \cdot \|v_i\|^2 = v_i v_i^T.$$

Algorithm: The actual algorithm is basically the same, but we need to repeat the experiment $\Theta(\log n)$ times and take the average, so that we can prove concentration.

Algorithm 6 Random Sampling Algorithm for Spectral Sparsification

Require: Vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ satisfying $\sum_{i=1}^m v_i v_i^T = I_n$.
1: Initialization: $\vec{s} \leftarrow \vec{0}$ and $\tau = \frac{6 \ln n}{\epsilon^2}$.
2: **for** $1 \le i \le m$ **do**
3: **for** $1 \le t \le \tau$ **do**
4: Update $s_i \leftarrow s_i + \frac{1}{\tau p_i}$ with probability $p_i = \|v_i\|_2^2$.
5: **end for**
6: **end for**
7: **return** $\sum_{i=1}^m s_i v_i v_i^T$.

There are two steps in the analysis. One is to show that there are $O(n \log n/\epsilon^2)$ non-zero scalars. Another is to show that the returned solution is a $(1 \pm \epsilon)$-spectral approximator to the identity matrix.

Expectation

We bound the number of non-zero scalars by computing its expected value and using Markov's inequality. The sampling probability is used to bound the expected value.

Lemma 9.12 (Number of Nonzeros)**.** *Let $\vec{s}$ be the output of Algorithm 6 and $S = \text{supp}(\vec{s})$ be the set of vectors with non-zero scalars. Then $|S| = O(n \log n/\epsilon^2)$ with probability at least 0.9.*

Proof. The expected value is

$$\mathbb{E}\left[|S|\right] = \sum_{i=1}^m \Pr[i \in S] = \sum_{i=1}^m \left(1 - (1-p_i)^\tau\right) \le \sum_{i=1}^m \left(1 - (1 - \tau p_i)\right) = \tau \sum_{i=1}^m p_i,$$

where $\tau = \frac{6 \ln n}{\epsilon^2}$ as defined in Algorithm 6. Note that

$$\sum_{i=1}^m p_i = \sum_{i=1}^m \|v_i\|_2^2 = \sum_{i=1}^m v_i^T v_i = \sum_{i=1}^m \text{Tr}(v_i^T v_i) = \sum_{i=1}^m \text{Tr}(v_i v_i^T) = \text{Tr}\left(\sum_{i=1}^m v_i v_i^T\right) = \text{Tr}(I_n) = n,$$

where we used $\text{Tr}(AB) = \text{Tr}(BA)$ in Fact 2.34. Therefore, $\mathbb{E}\left[|S|\right] \le \tau \sum_{i=1}^m p_i = \tau n = 6n \ln n/\epsilon^2$, and the result follows from Markov's inequality that $\Pr\left[|S| \ge 10\mathbb{E}\left[|S|\right]\right] \le 1/10$. □

Matrix Chernoff Bound

There is an elegant generalization of the Chernoff-Hoeffding bound to the matrix setting. The proof uses the Golden-Thompson inequality in Fact 2.36.

Theorem 9.13 (Matrix Chernoff Bound [Tro12]). *Let $X_1, \ldots, X_k$ be independent, $n \times n$ real symmetric matrices with $0 \preccurlyeq X \preccurlyeq rI$. Suppose $\mu_{\min} \cdot I_n \preccurlyeq \sum_{i=1}^{k} \mathbb{E}[X_i] \preccurlyeq \mu_{\max} \cdot I_n$. Then, for any $0 \leq \epsilon \leq 1$,*

$$\Pr\left[\lambda_{\max}\left(\sum_{i=1}^{k} X_i\right) \geq (1+\epsilon)\mu_{\max}\right] \leq n e^{-\frac{\epsilon^2 \mu_{\max}}{3r}} \text{ and } \Pr\left[\lambda_{\min}\left(\sum_{i=1}^{k} X_i\right) \leq (1-\epsilon)\mu_{\min}\right] \leq n e^{-\frac{\epsilon^2 \mu_{\max}}{2r}}.$$

Note that it is almost an exact analog of the Chernoff-Hoeffding bound in the scalar setting, by using the maximum eigenvalue and the minimum eigenvalue to measure the "size" of a matrix. Informally, it says that if we consider the sum of independent random matrices, when each matrix is not too "big/influential", the sum is concentrated around the expectation in terms of the eigenvalues.

Concentration

The algorithm was designed in a way such that the proof that the solution is a $(1 \pm \epsilon)$-spectral sparsifier is a direct application of the matrix Chernoff bound. The reweighting by the sampling probability is set to ensure that no random variable is too influential.

Lemma 9.14 (Success Probability of Spectral Approximation). *The output of Algorithm 6 satisfies $(1-\epsilon)I_n \preccurlyeq \sum_{i=1}^{m} s_i v_i v_i^T \preccurlyeq (1+\epsilon)I_n$ with probability at least $1 - \frac{2}{n}$.*

Proof. The random variables are

$$X_{i,t} = \begin{cases} \frac{v_i v_i^T}{\tau p_i} & \text{with probability } p_i \\ 0 & \text{otherwise} \end{cases},$$

for vector i in iteration t. The output of the algorithm is $Y := \sum_{i=1}^{m} \sum_{t=1}^{\tau} X_{i,t}$. The expected output is

$$\mathbb{E}[Y] = \sum_{i=1}^{m} \sum_{t=1}^{\tau} \mathbb{E}[X_{i,t}] = \sum_{i=1}^{m} \sum_{t=1}^{\tau} p_i \cdot \frac{v_i v_i^T}{\tau p_i} = \sum_{i=1}^{m} \sum_{t=1}^{\tau} \frac{v_i v_i^T}{\tau} = \sum_{i=1}^{m} v_i v_i^T = I_n.$$

So, the expected output is exactly the identity matrix, with $\mu_{\max} = \mu_{\min} = 1$. To apply the matrix Chernoff bound in Theorem 9.13, it remains to find a bound r so that $X_{i,t} \preccurlyeq rI$. Note that

$$X_{i,t} = \frac{v_i v_i^T}{\tau p_i} = \frac{v_i v_i^T}{\tau \|v_i\|^2} = \frac{1}{\tau}\left(\frac{v_i}{\|v_i\|}\right)\left(\frac{v_i}{\|v_i\|}\right)^T,$$

which is a rank-one matrix of a unit vector, and so the maximum eigenvalue is just $1/\tau$ and thus we can set $r = 1/\tau$. By Theorem 9.13, as $\tau = \frac{6 \ln n}{\epsilon^2}$,

$$\Pr[\lambda_{\max}(Y) \geq 1+\epsilon] \leq n e^{-\frac{\epsilon^2 \tau}{3}} = n e^{-2 \ln n} = \frac{1}{n}.$$

The lower tail follows similarly. So, with probability at least $1 - \frac{2}{n}$, we have $\lambda_{\max}(Y) \leq 1+\epsilon$ and $\lambda_{\min}(Y) \geq 1-\epsilon$, and thus $(1-\epsilon)I_n \preccurlyeq Y \preccurlyeq (1+\epsilon)I$, proving that the solution is a $(1 \pm \epsilon)$-spectral approximator of the identity matrix. □

Combining Lemma 9.12 and Lemma 9.14 by a union bound, we know that a $(1\pm\epsilon)$-spectral approximator of the identity matrix with $O(n\log n/\epsilon^2)$ vectors exists, and indeed the random sampling algorithm will succeed with constant probability. This proves Theorem 9.9, and the reduction in Lemma 9.11 proves the theorem by Spielman and Srivastava that every graph has a $(1\pm\epsilon)$-spectral approximator with $O(n\log n/\epsilon^2)$ edges.

Discussions

By considering spectral sparsification, there is an elegant and arguably simpler proof of Theorem 9.5 for cut sparsification by Benczur and Karger. In the cut sparsification problem, it was not very clear that what is the right sampling probability, and the conjecture that edge-connectivity can be used for sampling was only answered much later [FHHP19]. In the more general spectral sparsification problem, however, there seems to be only one natural choice for the sampling probability and the analysis follows directly from the matrix Chernoff bound. This is a great example that a more general problem can be easier to solve than a special case, where in the special case there seem to be multiple reasonable approaches while the generalization points to the right approach.

Sampling Probability: For spectral sparsification of graphs, the sampling probability of an edge $e=uv$ is proportional to

$$\|v_e\|_2^2 = \|U^T L_G^{\dagger/2} b_e\|_2^2 = b_e^T L_G^{\dagger/2} U U^T L_G^{\dagger/2} b_e = b_e^T L_G^{\dagger} b_e = \mathrm{Reff}_G(u,v),$$

where $\mathrm{Reff}_G(u,v)$ is the effective resistance of the two endpoints u and v in the graph G, when we view the graph as a resistor network with each edge being a resistor of resistance one. An equivalent characterization of effective resistance is

$$\mathrm{Reff}_G(u,v) = \min_{f:E\to R_{\geq 0}} \left\{ \sum_{e\in E} f(e)^2 \;\middle|\; f \text{ is a unit flow from } u \text{ to } v \right\}.$$

This quantity can be thought of as an interpolation between the shortest path distance and the maximum flow value of a graph. Effective resistance is known to be closely related to some quantities in random walks such as the commute time and cover time. Recently, this concept has various applications in designing fast graph algorithms where spectral sparsification is an excellent example.

Fast Algorithm: Spielman and Srivastava also gave a nearly linear time algorithm to estimate the effective resistances of all edges. The main tools are a nearly linear time algorithm to solve a Laplacian system of equations (which is a breakthrough result by Spielman and Teng), and also the dimension reduction result by Johnson and Lindenstrauss in Theorem 8.8. As a consequence, there is a nearly linear time randomized algorithm for constructing a spectral sparsifier of a graph, which is important for designing fast algorithms for other graph problems.

Tight Example: The analysis of the random sampling algorithm is tight. In a complete graph, the effective resistance of every edge is the same. So, the random sampling algorithm on a complete graph is just the uniform sampling algorithm. A "coupon collector" argument can be used to prove that random sampling won't work to find a cut sparsifier with $o(n\log n)$ edges. It is a good exercise to work out the details.

Randomized Linear Algebra: Random sampling and dimension reduction are very useful in designing fast algorithms for numerical linear algebra problems. We illustrate these ideas in a basic problem, the least square problem. In the least square problem, we are given an $n \times d$ matrix A and $b \in \mathbb{R}^n$, and the objective is to find an $x \in \mathbb{R}^d$ that minimizes $\|Ax - b\|_2$. We are usually interested in the case when $n \gg d$, so the problem is over-constrained. Exact algorithms require $\Omega(n \operatorname{poly}(d))$ time, which is too slow when n is large.

We would like to find an approximation algorithm with $\|Ax' - b\| \leq (1+\epsilon) \min_x \|Ax - b\|_2$ in $\tilde{O}(nd + \operatorname{poly}(\frac{d}{\epsilon}))$ time, which is near linear when $n \gg d$. The idea is to use a nearly linear time algorithm to compress the matrix A into a $k \times d$ matrix $B = SA$ with $k = \operatorname{poly}(\frac{d}{\epsilon})$, and then solve the least square problem on $\min_x \|S(Ax-b)\|_2$ exactly as our approximate solution. The techniques in spectral sparsification can be used for the compression.

Given $A \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^n$, we first reduce the problem to the case when the columns of A are orthonormal. This is reminiscent to the reduction to the identity matrix in Lemma 9.11, so that $A^T A = I_d$ or equivalently $\sum_{i=1}^n a_i a_i^T = I_d$ where a_i is the i-th row of A. Then, we construct a matrix B by sampling and rescaling each row proportional to its squared length, so that $B = \sum_{i=1}^n s_i a_i a_i^T \approx I_d$ with only $O(d \log d/\epsilon^2)$ nonzero scalars. Therefore, B has $O(d \log d/\epsilon^2)$ rows, where each row of B is $\sqrt{s_i} a_i$ so that $(1-\epsilon)A^T A \preccurlyeq B^T B \preccurlyeq (1+\epsilon)A^T A$. This is a good "subspace embedding" as $\|Ax\|_2^2 \approx \|Bx\|_2^2$ because $x^T A^T A x \approx x^T B^T B x$. All the technique details are similar to those in spectral sparsification, e.g. using matrix Chernoff bound. The sampling probability is called the leverage score of a row, a generalization of effective resistance.

9.3 References

[BK15] András A. Benczúr and David R. Karger. Randomized approximation schemes for cuts and flows in capacitated graphs. *SIAM J. Comput.*, 44(2):290–319, 2015. 87, 88, 89, 90

[FHHP19] Wai Shing Fung, Ramesh Hariharan, Nicholas J. A. Harvey, and Debmalya Panigrahi. A general framework for graph sparsification. *SIAM J. Comput.*, 48(4):1196–1223, 2019. 89, 94

[Kar99] David R. Karger. Random sampling in cut, flow, and network design problems. *Math. Oper. Res.*, 24(2):383–413, 1999. 87, 88

[Kar00] David R. Karger. Minimum cuts in near-linear time. *J. ACM*, 47(1):46–76, 2000. 89

[KT19] Ken-ichi Kawarabayashi and Mikkel Thorup. Deterministic edge connectivity in near-linear time. *J. ACM*, 66(1):4:1–4:50, 2019. 89

[SS11] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011. 87, 90

[ST11] Daniel A. Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM J. Comput.*, 40(4):981–1025, 2011. 87, 89, 90

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012. 93, 166

Chapter 10

Barrier Method

In Chapter 9, we have seen that spectral sparsification is a stronger notion than cut sparsification, but this provides a linear algebraic formulation that connects the problem to more general mathematical phenomenon, which leads to an elegant solution that matches the best known result in cut sparsification. In this chapter, we will see that this stronger notion even leads to a surprisingly strong solution that goes beyond what was known (or perhaps thought possible) for cut sparsification. The main theorem that we will study is by Batson, Spielman and Srivastava.

Theorem 10.1 (Linear-Sized Spectral Approximator of Identity Matrix [BSS14]). *For any m vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ satisfying $\sum_{i=1}^m v_i v_i^T = I_n$, there always exist scalars $s_1, \ldots, s_m$ with at most dn nonzeros such that*

$$\left(1 - \frac{1}{\sqrt{d}}\right)^2 \cdot I_n \preccurlyeq \sum_{i=1}^m s_i v_i v_i^T \preccurlyeq \left(1 + \frac{1}{\sqrt{d}}\right)^2 \cdot I_n.$$

It follows from the reduction in Lemma 9.11 that every graph has a linear-sized spectral sparsifier.

Theorem 10.2 (Linear-Sized Spectral Sparsifier [BSS14]). *For any edge-weighted undirected graph $G = ([n], E)$ and any $0 < \epsilon \leq 1$, there is a reweighted subgraph $H = ([n], F)$ on the same vertex set with at most $O(n/\epsilon^2)$ edges such that H is a $(1 \pm \epsilon)$-spectral approximator of G.*

One corollary is that every graph has a $(1 \pm \epsilon)$-cut sparsifier with at most $O(n/\epsilon^2)$ edges, which improves upon Theorem 9.5 by Benczur and Karger. It is quite remarkable that a harder problem leads to a stronger solution in a well-studied special case. Up until now, there is no known alternative way to obtain linear-sized cut sparsifiers without going through the concept of spectral sparsification. It is an interesting challenge especially to those who prefer to see combinatorial algorithms to solve combinatorial problems.

10.1 Deterministic Algorithm and Polynomial Perspective

The approach taken to prove Theorem 10.1 is different from the random sampling approach used in previous algorithms in Theorem 9.2, Theorem 9.5, and Theorem 9.9. It is a deterministic "greedy" approach that uses a potential function to guide the algorithm to add one vector at a time.

Intuition from Characteristic Polynomials

As discussed in [BSS14], the intuition of their approach is from a polynomial perspective to the problem. Let $A \in \mathbb{R}^{n\times n}$ be the current partial solution, where $A = 0$ initially. They considered the characteristic polynomial $p_A(x) = \det(xI - A) = \prod_{j=1}^{n}(x - \lambda_j)$ whose roots are the eigenvalues of A, and studied how it changes after adding one vector. By the matrix determinant formula in Fact 2.29,

$$p_{A+vv^T}(x) = \det(xI - A - vv^T) = \det(xI - A) \cdot \left(1 - v^T(xI - A)^{-1}v\right) = p_A(x) \cdot \left(1 - \sum_{j=1}^{n} \frac{\langle v, u_j\rangle^2}{x - \lambda_j}\right).$$

where λ_j are the eigenvalues of A and u_j are the corresponding orthonormal eigenvectors. Suppose we add a uniformly random vector v from $v_1, \ldots, v_m$ to A. Then, by the isotropy assumption,

$$\mathbb{E}\left[\langle v, u_j\rangle^2\right] = \frac{1}{m}\sum_{i=1}^{m}\langle v_i, u_j\rangle^2 = \frac{1}{m} \cdot u_j^T\left(\sum_{i=1}^{m} v_i v_i^T\right)u_j = \frac{\|u_j\|^2}{m} = \frac{1}{m}.$$

This implies that the expected characteristic polynomial is

$$\mathbb{E}\left[p_{A+vv^T}(x)\right] = p_A(x)\left(1 - \frac{1}{m}\sum_{j=1}^{n}\frac{1}{x - \lambda_j}\right) = p_A(x) - \frac{1}{m}\partial_x p_A(x),$$

as $\partial_x p_A(x)/p_A(x) = \sum_{j=1}^{n} 1/(x - \lambda_j)$. Since we start from $A = 0$, the initial polynomial is $p_A(x) = x^n$. After t iterations, the expected characteristic polynomial becomes

$$p_t(x) = \left(1 - \frac{1}{m}\partial_x\right)^t x^n.$$

This generates a standard family of orthogonal polynomials, called the associated Laguerre polynomials, whose roots are known. After $t = dn$ iterations, the ratio of the largest root to the smallest root of $p_{dn}(x)$ is known to be

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

and this is the ratio of the maximum eigenvalue and the minimum eigenvalue in Theorem 10.1.

This is only a heuristic argument, as there may not be any vector v with its characteristic polynomial $p_{A+vv^T}(x)$ equal to the expected characteristic polynomial. The proof of Theorem 10.1 in [BSS14] is also not based on this approach, but this foreshadows the polynomial approach that we will study in the second part of this course.

Algorithm Structure

As discussed in Chapter 9, one advantage of the algebraic formulation for spectral sparsification in Theorem 9.9 is that we "only" need to keep track of the maximum eigenvalue and the minimum eigenvalue of the current partial solution, instead of the exponentially many cut values as was done in the cut sparsification problem. So the general idea is to maintain an upper bound of the maximum eigenvalue and a lower bound on the minimum eigenvalue of the current partial solution, and to control how they evolve over time. This will be done using two potential functions Φ^u and Φ_l that we will define and study in the next section.

Assuming the existence of the two potential functions, we first describe the structure of the deterministic "greedy" algorithm. Initially, we start from the empty solution $A_0 = 0$, some upper bound u_0 of the maximum eigenvalue of A_0, some lower bound l_0 of the minimum eigenvalue, so that the potential values $\Phi^{u_0}(A_0) \leq \phi_u$ and $\Phi_{l_0}(A_0) \leq \phi_l$ for some values ϕ_u and ϕ_l that will be fixed throughout the algorithm. In each iteration t, we find a vector v_i and a scalar s and add $s \cdot v_i v_i^T$ to the current solution so that $A_{t+1} \leftarrow A_t + s v_i v_i^T$, and shift the upper bound $u_{t+1} \leftarrow u_t + \delta_u$ and the lower bound $l_{t+1} \leftarrow l_t + \delta_l$ by some fixed amount δ_u and δ_l to maintain the invariants that $\Phi^{u_{t+1}}(A_{t+1}) \leq \phi_u$ and $\Phi^{l_{t+1}}(A_{t+1}) \leq \phi_l$ and also u_{t+1} and l_{t+1} are upper and lower bounds of the maximum eigenvalue and the minimum eigenvalue of A_{t+1} respectively.

Algorithm 7 Deterministic Greedy Algorithm for Spectral Sparsification

Require: Vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ satisfying $\sum_{i=1}^m v_i v_i^T = I_n$.

1: Initialization: $A_0 = 0$ and $\tau = dn$.
2: Choose u_0, l_0, ϕ_u, ϕ_l so that $\Phi^{u_0}(A_0) \leq \phi_u$ and $\Phi_{l_0}(A_0) \leq \phi_l$.
3: Choose two parameters δ_u and δ_l and set $u_t = u_0 + t\delta_u$ and $l_t = l_0 + t\delta_l$ for any $t \geq 1$.
4: **for** $1 \leq t \leq \tau$ **do**
5: Find vector $v \in \{v_1, \ldots, v_m\}$ and scalar s and set $A_t = A_{t-1} + s \cdot vv^T$ to maintain the invariants that $\Phi^{u_t}(A_t) \leq \phi_u$ and $\Phi_{l_t}(A_t) \leq \phi_l$ and $\lambda_{\max}(A_t) \leq u_t$ and $\lambda_{\min}(A_t) \geq l_t$.
6: **end for**
7: **return** A_τ.

There are many parameters $u_0, l_0, \phi_u, \phi_l, \delta_u, \delta_l$ to be chosen, and we will only do so in the end.

10.2 Potential Functions

The magical element in this algorithm is the definition of the potential functions, that will make everything works beautifully. Before we state the potential functions used in [BSS14], let us discuss some natural attempts and see what we need.

Norm of Eigenvalues

A natural first attempt is to simply use the maximum eigenvalue and the minimum eigenvalue as the potential functions (i.e. $\Phi^u(A_t) = \lambda_{\max}(A_t)$ and $\Phi_l(A_t) = \lambda_{\min}(A_t)$), and then inductively prove that $\lambda_{\max}(A_t) \leq \lambda_{\max}(A_{t-1}) + \delta_u$ and $\lambda_{\min}(A_t) \leq \lambda_{\min}(A_{t-1}) + \delta_l$. This way of measuring progress does not work well for this problem, as the matrix A_t is n-dimensional, and just focusing on the maximum direction cannot distinguish between the case where every direction is large or where one direction is large and all other orthogonal directions are small. Ideally, we hope to say something such as after n iterations, every direction is increased by one unit. To prove it inductively, we would need a potential function to let us argue that the maximum direction is increased by $1/n$ unit per edge *on average*, but the maximum eigenvalue is not such a smooth/robust quantity for this.

By the above discussion, we would like to have a more global quantity that will take into consideration of all directions. One possible parameter of this kind is $\frac{1}{n}\operatorname{Tr}(A)$, which is the average eigenvalue of the current solution. For this, we can easily argue that the average eigenvalue increases smoothly, but the problem is that we cannot conclude that the maximum eigenvalue is small by using that the average eigenvalue is small.

So, we would like to have a more global quantity that is smooth enough to measure the progress made in each iteration, and also that the maximum eigenvalue is small when this quantity is small. Let $\vec{\lambda} = \big(\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A)\big)$ be the spectrum of the current solution. Note that $\lambda_{\max}(A) := \|\vec{\lambda}\|_\infty$ is the infinity norm of the spectrum, while $\mathrm{Tr}(A) := \|\vec{\lambda}\|_1$ is the 1-norm of the spectrum. Interpolating between these two extremes, we may consider the quantity $\left(\frac{1}{n}\sum_{i=1}^n \lambda_i^p\right)^{1/p} = n^{-1/p} \cdot \|\lambda\|_p$. We know that setting $p \approx \log n$ would approximate $\|\vec{\lambda}\|_\infty$ well, but the p-norm may not be so convenient for calculations. In convex optimization, there is a softmax function that is defined as $\log \sum_{i=1}^n \exp(\lambda_i)$, which is known to be convex and differentiable and approximates the maximum well. In our setting, the softmax function can be nicely written as $\log \mathrm{Tr}(e^A)$, where $e^A := \sum_{k=0}^\infty \frac{1}{k!} X^k$ is the matrix exponential of A. So this seems to be a good potential function to be used for spectral sparsification, and indeed this function is used in the proof of the matrix Chernoff bound. I think this function can be used in Algorithm 7 to give a deterministic algorithm with the same guarantee as the random sampling algorithm by Spielman and Srivastava in Theorem 9.9 (please see [dCSHS16]), but it is not enough for linear-sized spectral sparsification.

Barrier Functions

Batson, Spielman, and Srivastava mentioned in [BSS14] that the definition of their potential functions is inspired by the calculation of the expected characteristic polynomial in Section 10.1.

Definition 10.3 (Barrier Functions). *Given $u, l \in \mathbb{R}$ and a real symmetric matrix $A \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$, the upper barrier function and the lower barrier function are defined as*

$$\Phi^A(u) := \Phi^u(A) := \mathrm{Tr}(uI_n - A)^{-1} = \sum_{i=1}^n \frac{1}{u - \lambda_i} \quad \text{and} \quad \Phi_A(l) := \Phi_l(A) := \mathrm{Tr}(A - lI_n)^{-1} = \sum_{i=1}^n \frac{1}{\lambda_i - l}.$$

We will use the notations $\Phi^u(A)$ and $\Phi_l(A)$ when we fix u and l and see the barrier functions as a function of the matrix A, and we will use the notations $\Phi^A(u)$ and $\Phi_A(l)$ when we fix A and see the barrier functions as a function of u or l.

When $u > \lambda_{\max}(A)$ and $l < \lambda_{\min}(A)$, these functions measure how far the eigenvalues of A are from the barriers u and l, and they blow up as any eigenvalue approaches a barrier. Suppose we could maintain the invariant that say $\Phi^{u_t}(A) \le 1$ for all t. This will ensure that u_t is a "comfortable" upper bound of the maximum eigenvalue, as there could be at most one eigenvalue with value at least $u_t - 1$, at most two eigenvalues with value at least $u_t - 2$, and so on. This is a more global quantity that takes all the eigenvalues into consideration, and has the property that it changes smoothly to measure the progress made in each iteration.

The following properties of the barrier functions are simple but useful. We will see a generalization in the multivariate setting in the second part of the course.

Exercise 10.4 (Monotonicity and Convexity). *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. For any $u > \lambda_{\max}(A)$ and any $\delta > 0$, the upper barrier function satisfies*

$$\Phi^A(u) \ge \Phi^A(u+\delta) \quad \text{and} \quad \Phi^A(u) + \delta \cdot \big(\Phi^A(u+\delta)\big)' \ge \Phi^A(u+\delta).$$

For any $l + \delta < \lambda_{\min}(A)$ and any $\delta > 0$, the lower barrier function satisfies

$$\Phi_A(l) \le \Phi_A(l+\delta) \quad \text{and} \quad \Phi_A(l) + \delta \cdot \big(\Phi_A(l+\delta)\big)' \ge \Phi_A(l+\delta).$$

The strategy in Algorithm 7 is to ensure that u_t is increased slowly while maintaining the invariant that the potential value $\Phi^{u_t}(A_t)$ is small. More explicitly, we can define a family of "soft" bounds on the max/min eigenvalue, parameterized by the value of the potential functions.

Definition 10.5 (ϕ-Soft-Max and ϕ-Soft-Min)**.** *Given a real symmetric matrix $A \in \mathbb{R}^{n\times n}$ and a parameter $\phi > 0$, the ϕ-max of A and the ϕ-min of A are defined as*

$$\phi\text{-}\max(A) := \max\{u \mid \Phi^A(u) = \phi\} \quad \textit{and} \quad \phi\text{-}\min(A) := \min\{l \mid \Phi_A(l) = \phi\}.$$

They can be understood as the inverse of the upper and lower barrier functions.

The parameter ϕ can be thought of as a sensitivity parameter, which controls the tradeoff between how accurate the bound is and how smoothly it varies. The strategy in Algorithm 7 is to fix an ϕ and then bound $\phi\text{-}\max(A_t)$ and prove that $\phi\text{-}\max(A_t) \le \phi\text{-}\max(A_{t-1}) + \delta_u$ for all $t \ge 1$.

The barrier functions are similar to the log-barrier functions used in the interior point method for convex optimization.

Remark 10.6 (Log-Barrier Functions)**.** *Let $p_A(x) = \det(xI - A)$ be the characteristic polynomial of A. Note that*

$$\Phi^x(A) = \frac{\partial_x p_A(x)}{p_A(x)} = \partial_x \log\big(p_A(x)\big) \quad \textit{and} \quad \Phi_x(A) = -\partial_x \log\big(p_A(x)\big).$$

These functions blow up when x is getting close to a root.

10.3 Changes of Potential Values

There are nice formulas to analyze the change of the barrier functions when we add a vector and do a rank-one update.

Upper Barrier Function

For the upper barrier function $\Phi^u(A)$, adding a vector $s\cdot vv^T$ would increase the potential value, but increasing the upper bound u would compensate for it to maintain the invariant $\Phi^{u+\delta_u}(A+s\cdot vv^T) \le \Phi^u(A)$.

Lemma 10.7 (Upper Barrier Change)**.** *Suppose $u > \lambda_{\max}(A)$. For any vector v, if*

$$\frac{1}{s} \ge \frac{v^T\big((u+\delta_u)I - A\big)^{-2}v}{\Phi^u(A) - \Phi^{u+\delta_u}(A)} + v^T\big((u+\delta_u)I - A\big)^{-1}v =: U_A(v),$$

then

$$\Phi^{u+\delta_u}\big(A + s\cdot vv^T\big) \le \Phi^u(A) \quad \textit{and} \quad \lambda_{\max}\big(A + s\cdot vv^T\big) < u + \delta_u.$$

Proof. Let $u' := u + \delta_u$. By the Sherman-Morrison rank-one update formula in Fact 2.20,

$$\begin{aligned}
\Phi^{u+\delta_u}(A + s \cdot vv^T) &= \operatorname{Tr}\left(\left(u'I - A - s \cdot vv^T\right)^{-1}\right) \\
&= \operatorname{Tr}\left(\left(u'I - A\right)^{-1} + \frac{s(u'I - A)^{-1}vv^T(u'I - A)^{-1}}{1 - s \cdot v^T(u'I - A)^{-1}v}\right) \\
&= \Phi^{u+\delta_u}(A) + \frac{s \cdot v^T(u'I - A)^{-2}v}{1 - s \cdot v^T(u'I - A)^{-1}v} \\
&= \Phi^{u}(A) - \underbrace{\left(\Phi^{u}(A) - \Phi^{u+\delta_u}(A)\right)}_{\text{gain}} + \underbrace{\frac{v^T(u'I - A)^{-2}v}{1/s - v^T(u'I - A)^{-1}v}}_{\text{loss}}
\end{aligned}$$

Rearranging shows that $\Phi^{u+\delta_u}\left(A + s \cdot vv^T\right) \leq \Phi^u(A)$ when $1/s \geq U_A(v)$. This also implies that $\lambda_{\max}(A + s \cdot vv^T) \leq u + \delta_u$, as otherwise $\lambda_{\max}(A + s' \cdot vv^T) = u + \delta_u$ for some $s' \leq s$ and thus $\Phi^{u+\delta_u}\left(A + s' \cdot vv^T\right) = \infty$, but this contradicts that $\Phi^{u+\delta_u}\left(A + s' \cdot vv^T\right) \leq \Phi^u(A)$ is bounded. □

Lower Barrier Function

For the lower barrier function $\Phi_l(A)$, adding a vector $s \cdot vv^T$ would decrease the potential value, but increasing the lower bound l would increase the potential value. Note that there is an additional condition about the barrier value to ensure that we still have a lower bound on the minimum eigenvalue.

Lemma 10.8 (Lower Barrier Change)**.** *Suppose $\lambda_{\min}(A) > l$ and $\Phi_l(A) \leq 1/\delta_l$. For any vector v, if*

$$0 < \frac{1}{s} \leq \frac{v^T\left(A - (l + \delta_l)I\right)^{-2}v}{\Phi_{l+\delta_l}(A) - \Phi_l(A)} - v^T\left(A - (l + \delta_l)I\right)^{-1}v =: L_A(v),$$

then

$$\Phi_{l+\delta_l}\left(A + s \cdot vv^T\right) \leq \Phi_l(A) \quad \text{and} \quad \lambda_{\min}\left(A + s \cdot vv^T\right) > l + \delta_l.$$

Proof. Note that $\lambda_{\min}(A) > l$ and $\Phi_l(A) = \sum_{i=1}^n 1/(\lambda_i - l) \leq 1/\delta_l$ imply that $1/(\lambda_{\min} - l) < 1/\delta_l$ and thus $\lambda_{\min} > l + \delta_l$. So, $\lambda_{\min}(A + s \cdot vv^T) \geq \lambda_{\min}(A) > l + \delta_l$. Then, by a similar calculation using the Sherman-Morrison formula as in Lemma 10.7,

$$\Phi_{l+\delta_l}(A + s \cdot vv^T) = \Phi_l(A) + \underbrace{\left(\Phi_{l+\delta_l}(A) - \Phi_l(A)\right)}_{\text{loss}} - \underbrace{\frac{v^T(A - l'I)^{-2}v}{1/s + v^T(A - l'I)^{-1}v}}_{\text{gain}}.$$

Rearranging shows that $\Phi_{l+\delta_l}\left(A + s \cdot vv^T\right) \leq \Phi_l(A)$ when $1/s \leq L_A(v)$. □

10.4 Averaging Argument

We need to prove that there exists a vector v and a scalar s such that both the assumptions in Lemma 10.7 and Lemma 10.8 hold, so that we can conclude that the invariants $\Phi^{u+\delta_u}\left(A + s \cdot vv^T\right) \leq \Phi^u(A)$ and $\lambda_{\max}\left(A + s \cdot vv^T\right) < u + \delta_u$ and $\Phi_{l+\delta_l}\left(A + s \cdot vv^T\right) \leq \Phi_l(A)$ and $\lambda_{\min}\left(A + s \cdot vv^T\right) > l + \delta_l$ in Algorithm 7 hold simultaneously.

The idea in [BSS14] is to prove that $\sum_{i=1}^m L_A(v_i) \geq \sum_{i=1}^m U_A(v_i)$, and so there exists a vector v_i such that $L_A(v_i) \geq U_A(v_i)$. Therefore, by setting s to be a scalar such that $L_A(v_i) \geq 1/s \geq U_A(v_i)$, then both the assumptions in Lemma 10.7 and Lemma 10.8 are satisfied and thus all the invariants hold simultaneously for $A + s \cdot v_i v_i^T$.

The calculations work out quite nicely using the isotropy condition $\sum_{i=1}^m v_i v_i^T = I_n$.

Upper Barrier Function

Lemma 10.9 (Total Upper Barrier Shift). *Given $v_1, \ldots, v_m \in \mathbb{R}^n$ such that $\sum_{i=1}^m v_i v_i^T = I_n$,*

$$\sum_{i=1}^m U_A(v_i) \leq \frac{1}{\delta_u} + \Phi^u(A).$$

Proof. Using the isotropy assumption $\sum_{i=1}^m v_i v_i^T = I_n$, it follows that

$$\sum_{i=1}^m v_i^T \big((u+\delta_u)I - A\big)^{-2} v_i = \sum_{i=1}^m \mathrm{Tr}\left(\big((u+\delta_u)I - A\big)^{-2} v_i v_i^T\right) = \mathrm{Tr}\left(\big((u+\delta_u)I - A\big)^{-2}\right),$$

and similarly

$$\sum_{i=1}^m v_i^T \big((u+\delta_u)I - A\big)^{-1} v_i = \mathrm{Tr}\left(\big((u+\delta_u)I - A\big)^{-1}\right) = \Phi^{u+\delta_u}(A).$$

By the convexity of the barrier function $\Phi^u(A) = \Phi^A(u)$ in terms of u in Exercise 10.4, the "gain" is

$$\Phi^u(A) - \Phi^{u+\delta_u}(A) = \Phi^A(u) - \Phi^A(u+\delta_u) \geq -\delta_u \cdot \Big(\Phi^A(u+\delta_u)\Big)' = \delta_u \cdot \mathrm{Tr}\left(\big((u+\delta_u)I - A\big)^{-2}\right).$$

Therefore,

$$\begin{aligned}
\sum_{i=1}^m U_A(v_i) &:= \sum_{i=1}^m \left(\frac{v_i^T \big((u+\delta_u)I - A\big)^{-2} v_i}{\Phi^u(A) - \Phi^{u+\delta_u}(A)} + v_i^T \big((u+\delta_u)I - A\big)^{-1} v_i\right) \\
&= \frac{\mathrm{Tr}\left(\big((u+\delta_u)I - A\big)^{-2}\right)}{\Phi^u(A) - \Phi^{u+\delta_u}(A)} + \Phi^{u+\delta_u}(A) \\
&\leq \frac{1}{\delta_u} + \Phi^u(A).
\end{aligned}$$

□

Lower Barrier Function

The calculations for the total lower barrier shift is similar, but is a bit trickier. Note that the following lemma also requires the assumption that $\Phi_l(A) \leq 1/\delta_l$ as in Lemma 10.8.

Lemma 10.10 (Total Lower Barrier Shift). *Given $v_1, \ldots, v_m \in \mathbb{R}^n$ such that $\sum_{i=1}^m v_i v_i^T = I_n$, if $\Phi_l(A) \leq 1/\delta_l$, then*

$$\sum_{i=1}^m L_A(v_i) \geq \frac{1}{\delta_l} - \Phi_l(A).$$

Proof. As in the proof of Lemma 10.9, using the isotropy assumption $\sum_{i=1}^m v_i v_i^T = I_n$,

$$\sum_{i=1}^m v_i^T \big(A - (l+\delta_l)I\big)^{-2} v_i = \mathrm{Tr}\left(\big(A - (l+\delta_l)I\big)^{-2}\right) \quad \text{and} \quad \sum_{i=1}^m v_i^T \big(A - (l+\delta_l)I\big)^{-1} v_i = \Phi_{l+\delta_l}(A).$$

Therefore,

$$\begin{aligned}\sum_{i=1}^m L_A(v_i) &:= \sum_{i=1}^m \left(\frac{v_i^T \big(A - (l+\delta_l)I\big)^{-2} v_i}{\Phi_{l+\delta_l}(A) - \Phi_l(A)} - v_i^T \big(A - (l+\delta_l)I\big)^{-1} v_i\right) \\ &= \frac{\mathrm{Tr}\big(A - (l+\delta_l)I\big)^{-2}}{\Phi_{l+\delta_l}(A) - \Phi_l(A)} - \Phi_{l+\delta_l}(A).\end{aligned}$$

Using convexity in Exercise 10.4 as in the proof of Lemma 10.9,

$$\Phi_{l+\delta_l}(A) - \Phi_l(A) = \Phi_A(l+\delta_l) - \Phi_A(l) \le \delta_l \cdot \big(\Phi_A(l+\delta_l)\big)' = \delta_l \cdot \mathrm{Tr}\big(A - (l+\delta_l)I\big)^{-2}.$$

This gives $\sum_{i=1}^m L_A(v_i) \ge \frac{1}{\delta_l} - \Phi_{l+\delta_l}(A)$, which is slightly weaker than the statement and is not enough for the invariants to hold throughout the algorithm. To prove the statement, we need to work harder and show that

$$\frac{\mathrm{Tr}\big(A - (l+\delta_l)I\big)^{-2}}{\Phi_{l+\delta_l}(A) - \Phi_l(A)} - \Phi_{l+\delta_l}(A) \ge \frac{1}{\delta_l} - \Phi_l(A),$$

which is equivalent to the following claim by rearranging.

Claim 10.11 (Lemma 4.3 of [MSS21]). *If $\Phi_l(A) \le 1/\delta_l$, then*

$$\big(\Phi_{l+\delta_l}(A) - \Phi_l(A)\big)^2 \le \mathrm{Tr}\big(A - (l+\delta_l)I\big)^{-2} - \frac{1}{\delta_l}\big(\Phi_{l+\delta_l}(A) - \Phi_l(A)\big).$$

Proof. By definition of the lower barrier function in Definition 10.3

$$\big(\Phi_{l+\delta_l}(A) - \Phi_l(A)\big)^2 = \left(\sum_{i=1}^n \frac{1}{\lambda_i - (l+\delta_l)} - \frac{1}{\lambda_i - l}\right)^2 = \left(\sum_{i=1}^n \frac{\delta_l}{(\lambda_i - (l+\delta_l)) \cdot (\lambda_i - l)}\right)^2$$

Using Cauchy-Schwarz inequality and then the assumption $\delta_l \cdot \Phi_l(A) \le 1$, the RHS is

$$\le \left(\sum_{i=1}^n \frac{\delta_l}{(\lambda_i - l)}\right)\left(\sum_{i=1}^n \frac{\delta_l}{\big(\lambda_i - (l+\delta_l)\big)^2 \cdot (\lambda_i - l)}\right) \le \left(\sum_{i=1}^n \frac{\delta_l}{\big(\lambda_i - (l+\delta_l)\big)^2 \cdot (\lambda_i - l)}\right).$$

Check that this is equal to the RHS of the statement of this claim. □

The claim completes the proof of this lemma. □

Both Barrier Functions

Combining Lemma 10.9 and Lemma 10.10 with the averaging argument in the beginning of this section, we arrive at the following conditions for the invariants in Algorithm 7 to hold throughout.

Lemma 10.12 (Invariants). *Let $A_0 = 0$. If we choose $u_0 > 0, l_0 < 0, \phi_u, \phi_l, \delta_u, \delta_l$ so that*

$$\Phi^{u_0}(A_0) \le \phi_u \quad \text{and} \quad \Phi_{l_0}(A_0) \le \phi_l \quad \text{and} \quad \phi_l \le \frac{1}{\delta_l} \quad \text{and} \quad \frac{1}{\delta_l} - \phi_l \ge \frac{1}{\delta_u} + \phi_u,$$

then Algorithm 7 can always find a vector v and a scalar s in each iteration t to maintain the invariants that $\Phi^{u_t}(A_t) \le \phi_u$ and $\Phi_{l_t}(A_t) \le \phi_l$ and $\lambda_{\max}(A_t) \le u_t$ and $\lambda_{\min}(A_t) \ge l_t$, where $u_t = u_0 + t\delta_u$ and $l_t = l_0 + t\delta_l$ as defined in Algorithm 7.

Proof. The proof is by a simple induction. The induction hypothesis is that $\Phi^{u_t}(A_t) \le \phi_u$ and $\Phi_{l_t}(A_t) \le \phi_l$ and $\lambda_{\max}(A_t) \le u_t$ and $\lambda_{\min}(A_t) \ge l_t$. This holds at $t = 0$ by our assumptions. For the induction step, by Lemma 10.9 and Lemma 10.10 and our assumption,

$$\sum_{i=1}^{m} U_{A_t}(v_i) \le \frac{1}{\delta_u} + \Phi^{u_t}(A_t) \le \frac{1}{\delta_u} + \phi_u \le \frac{1}{\delta_l} - \phi_l \le \frac{1}{\delta_l} - \Phi_{l_t}(A_t) \le \sum_{i=1}^{m} L_{A_t}(v_i).$$

So there exists some $v \in \{v_1, \dots, v_m\}$ such that $U_{A_t}(v) \le L_{A_t}(v)$. Let s be a scalar such that $U_{A_t}(v) \le 1/s \le L_{A_t}(v)$. Then, it follows from Lemma 10.7 and Lemma 10.8 that the invariants hold for $t+1$ with $A_{t+1} = A_t + s \cdot vv^T$ and $u_{t+1} = u_t + \delta_u$ and $l_{t+1} = l_t + \delta_l$. □

Wrapping Up

With Lemma 10.12, it remains to choose $u_0, l_0, \phi_u, \phi_l, \delta_u, \delta_l$ to prove Theorem 10.1. Batson, Spielman and Srivastava set

$$l_0 := -\sqrt{d}n, \quad u_0 := \Big(\frac{d+\sqrt{d}}{\sqrt{d}-1}\Big)n, \quad \phi_l := \Phi_{l_0}(A_0) = -\frac{n}{l_0} = \frac{1}{\sqrt{d}}, \quad \phi_u := \Phi^{u_0}(A_0) = \frac{n}{u_0} = \frac{\sqrt{d}-1}{d+\sqrt{d}},$$

so that the first three conditions in Lemma 10.12 are satisfied. Then, they set

$$\delta_l := 1 \quad \text{and} \quad \delta_u := \frac{\sqrt{d}+1}{\sqrt{d}-1} \quad \Longrightarrow \quad \frac{1}{\delta_l} - \phi_l = \frac{1}{\delta_u} + \phi_u,$$

and so the last condition in Lemma 10.12 is also satisfied. Therefore, after dn iterations of Algorithm 7,

$$\frac{\lambda_{\max}(A_{dn})}{\lambda_{\min}(A_{dn})} \le \frac{u_{dn}}{l_{dn}} = \frac{u_0 + dn \cdot \delta_u}{l_0 + dn \cdot \delta_l} = \frac{\frac{d+\sqrt{d}}{\sqrt{d}-1} + d \cdot \frac{\sqrt{d}+1}{\sqrt{d}-1}}{-\sqrt{d}+d} = \left(\frac{\sqrt{d}+1}{\sqrt{d}-1}\right)^2,$$

completing the proof of Theorem 10.1.

10.5 Discussions

There are many subsequent work on spectral sparsification and we discuss some of them here.

- Allen-Zhu, Liao, and Orecchia [ALO15] constructed linear-sized spectral sparsifier using the regret minimization framework in convex optimization. This provides a more systematic way to derive the result and a different interpretation of Batson, Spielman, Srivastava's result as using a different regularizer in the regret minimization framework. Their tools developed are also more convenient in some applications as we will discuss more in the next chapter.

- Lee and Sun gave an almost linear time algorithm [LS18] and then a nearly linear time algorithm [LS17] to construct linear-sized spectral sparsifiers. Their first algorithm in [LS18] is an adaptive sampling algorithm that is interesting and easy to describe. In the first iteration, the algorithm samples say $n^{0.99}$ vectors using effective resistance as in Theorem 9.9. Then, it will update the sampling probability using barrier functions, and repeat this process for $n^{0.01}$ iterations. Their intuition is from balls-and-bins that the maximum load is only a constant after throwing $n^{0.99}$ balls to n bins.

- One may wonder whether there are other sampling algorithms for constructing spectral sparsifiers. An interesting result by Kyng and Song [KS18] is that the union of $\log n/\epsilon^2$ random spanning trees is an $(1 \pm \epsilon)$-spectral approximator, and this is tight.

- de Carli Silva, Harvey and Sato [dCSHS16] generalized the result in Theorem 10.1 to sparsifying sums of positive semidefinite matrices that have arbitrary rank, with applications in hypergraph sparsification.

A main open question is if every vector is of the same length (or more generally when every vector is short), then is there an efficient algorithm to construct an *unweighted* sparsifier? This is closely related to the Kadison-Singer problem that we will study in the second part of the course.

10.6 References

[ALO15] Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 237–245. ACM, 2015. 105, 108, 113, 114

[BSS14] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Rev.*, 56(2):315–334, 2014. 97, 98, 99, 100, 103, 108, 110, 117

[dCSHS16] Marcel Kenji de Carli Silva, Nicholas J. A. Harvey, and Cristiane M. Sato. Sparse sums of positive semidefinite matrices. *ACM Trans. Algorithms*, 12(1):9:1–9:17, 2016. 100, 106

[KS18] Rasmus Kyng and Zhao Song. A matrix chernoff bound for strongly rayleigh distributions and spectral sparsifiers from a few random spanning trees. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 373–384. IEEE Computer Society, 2018. 106

[LS17] Yin Tat Lee and He Sun. An sdp-based algorithm for linear-sized spectral sparsification. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 678–687. ACM, 2017. 106

[LS18] Yin Tat Lee and He Sun. Constructing linear-sized spectral sparsification in almost-linear time. *SIAM J. Comput.*, 47(6):2315–2336, 2018. 106

[MSS21] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families iii: Sharper restricted invertibility estimates. *Israel Journal of Mathematics*, pages 1–28, 2021. 104, 111, 119, 122, 123, 135

Chapter 11

Spectral Rounding

In the spectral sparsification problem in Chapter 9 and Chapter 10, we are given $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$, and the goal is to find a "reweighting" $s_1, \ldots, s_m$ with few nonzeros so that $\sum_{i=1}^m s_i v_i v_i^T \approx \sum_{i=1}^m v_i v_i^T$.

In this chapter, we consider the following spectral rounding problem where the goal is to find an "integral reweighting" that approximates the input.

Definition 11.1 (Spectral Rounding)**.** *Given $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$ and scalars $x_1, \ldots, x_m \in \mathbb{R}_{\geq 0}$, find integer scalars $z_1, \ldots, z_m \in \mathbb{Z}_{\geq 0}$ such that*

$$\sum_{i=1}^m x_i v_i v_i^T \approx \sum_{i=1}^m z_i v_i v_i^T.$$

More generally, we are also given k linear constraints in a matrix $A \in \mathbb{R}_{\geq 0}^{k \times m}$ and are required to find integer scalars $z_1, \ldots, z_m \in \mathbb{Z}_{\geq 0}$ that also satisfies

$$A\vec{x} \approx A\vec{z}.$$

Note that there is no requirement on the number of nonzeros in $\vec{z}$ as in the spectral sparsification problem, rather the requirement is on the integrality of $\vec{z}$.

The motivation of this problem is from designing approximation algorithms for some discrete optimization problems, where we should think of $\vec{x}$ as an optimal fractional solution to some convex relaxation of a combinatorial problem, and our goal is to find an integer solution $\vec{z}$ that is almost as good as $\vec{x}$. The additional linear constraints can be used to incorporate the objective value of the solutions, and/or some other constraints such as upper and lower bound on the size of the solutions. We will see two concrete applications in the next section.

This problem in its strongest form is as general as the Kadison-Singer problem that we will study in the second part of the course. In this chapter, we consider a simpler setting where the two-sided approximation requirement is replaced by a one-sided approximation requirement. The main result that we will study is by Allen-Zhu, Li, Singh and Wang [ALSW17], who formulated the following minimum eigenvalue problem and used it to design approximation algorithms for experimental design problems.

Theorem 11.2 (Minimum Eigenvalue Problem [ALSW17])**.** *Given $v_1, \ldots, v_m \in \mathbb{R}^n$ and scalars $x_1, \ldots, x_m \in \mathbb{R}_{\geq 0}$ satisfying*

$$\sum_{i=1}^m x_i v_i v_i^T = I_n \quad \text{and} \quad \sum_{i=1}^m x_i = k,$$

there is a polynomial time algorithm to find integer scalars $z_1, \ldots, z_m \in \mathbb{Z}_{\geq 0}$ *satisfying*

$$\sum_{i=1}^{m} z_i v_i v_i^T \succcurlyeq \left(1 - \sqrt{\frac{n-1}{k}}\right)^2 \cdot I_n \quad \textit{and} \quad \sum_{i=1}^{m} z_i = k.$$

To see its connection to the spectral rounding problem in Definition 11.1, first we apply the same reduction as in Lemma 9.11 to reduce to the case when $\sum_{i=1}^m x_i v_i v_i^T = I_n$. Then the two-sided approximation requirement in Definition 11.1 becomes $(1-\epsilon)I_n \preccurlyeq \sum_{i=1}^m z_i v_i v_i^T \preccurlyeq (1+\epsilon)I_n$ for an ϵ as small as possible. In Theorem 11.2, the two-sided requirement is replaced by the one-sided requirement $\sum_{i=1}^m z_i v_i v_i^T \succcurlyeq (1-\epsilon)I_n$. And there is one linear "cardinality/budget" constraint $\sum_{i=1}^m x_i \approx \sum_{i=1}^m z_i$ to satisfy, without which the problem is trivial.

The proof of Theorem 11.2 in [ALSW17] is based on the regret minimization framework developed for spectral sparsification by Allen-Zhu, Liao, and Orecchia [ALO15]. It will take quite some time to introduce this framework properly and we will not do so, but we will briefly describe their framework at the end of this chapter.

Instead, we will present a new proof of Theorem 11.2, following the (informal) polynomial perspective for spectral sparsification from [BSS14] that we described in the beginning of Chapter 10. I hope this proof serves better as a bridge to connect to the second part of the course, starting next chapter.

11.1 Applications

Before we see the proof, let's first see some applications of Theorem 11.2, which is useful in designing approximation algorithms for choosing a good subset of points/vectors/edges.

Experimental Design

In experimental design problems, we are given vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ and a parameter $k \geq n$, and the goal is to choose a (multi-)subset S of k vectors so that $\sum_{i \in S} v_i v_i^T$ optimizes some objective function. The most popular and well-studied objective functions are:

- D-design: Maximizing $\left(\det\left(\sum_{i\in S} v_i v_i^T\right)\right)^{\frac{1}{n}}$.
- A-design: Minimizing $\mathrm{Tr}\left(\left(\sum_{i\in S} v_i v_i^T\right)^{-1}\right)$.
- E-design: Maximizing $\lambda_{\min}\left(\sum_{i\in S} v_i v_i^T\right)$.

These problems of choosing a representative subset of vectors have a wide range of applications (see [ALSW21, LZ21]), but these are all NP-hard. To design approximation algorithms, we consider the following natural convex programming relaxations for D/A/E-design.

$$\begin{aligned} \max \quad & \left(\det\left(\sum_{i=1}^n x_i \cdot v_i v_i^T\right)\right)^{\frac{1}{n}} \Big/ \min \ \mathrm{Tr}\left(\sum_{i=1}^n x_i \cdot v_i v_i^T\right)^{-1} \Big/ \max \ \lambda_{\min}\left(\sum_{i=1}^n x_i \cdot v_i v_i^T\right) \\ \text{s.t.} \quad & \sum_{i=1}^m x_i \leq k. \\ & x_i \geq 0, \qquad \text{for } 1 \leq i \leq n. \end{aligned}$$

After we computed an approximately optimal solution x in polynomial time, we can apply the transformation as in Lemma 9.11 to reduce to the case where $\sum_{i=1}^m x_i v_i v_i^T = I_n$. Then we can apply Theorem 11.2 to obtain an integral solution z, and then apply the reverse transformation in Lemma 9.11 to see that z has the following performance guarantee.

Problem 11.3 (Experimental Design). *Prove that Theorem 11.2 can be used to obtain a $(1 \pm \epsilon)$-approximation algorithm for D/A/E-design when $k \gtrsim n/\epsilon^2$.*

This approach is used in [ALSW21, LZ21] to provide a unifying algorithmic framework for designing the best known approximation algorithms for a large class of experimental design problems. We will discuss some ideas of these work in the end of this chapter.

Network Design

The general setting of network design is to find a minimum cost subgraph satisfying certain requirements. The most well-studied problem is the survivable network design problem, where the requirement is to have at least a specified number $f_{u,v}$ of edge-disjoint paths between every pair of vertices u, v. Linear programming is the default approach in designing approximation algorithms for network design problems. It is observed in [LZ20] that spectral techniques can also be used for survivable network design problems, as well as to incorporate additional spectral constraints. For example, consider the following convex relaxation:

$$\begin{aligned}
\min_x \quad & \sum_{e\in E} c_e x_e & & \\
& \sum_{e\in\delta(S)} x_e \geq \max_{u\in S, v\notin S} \{f_{u,v}\} & \forall S \subseteq V \quad & \text{(connectivity constraints)} \\
& \lambda_2(L_x) \geq \lambda & & \text{(algebraic connectivity constraint)} \\
& 0 \leq x_e \leq 1 & \forall e \in E \quad & \text{(capacity constraints)}
\end{aligned}$$

where c_e is the given cost of an edge $e \in E$, and L_x is the Laplacian matrix where each edge e has weight x_e. The algebraic connectivity constraint can be used to lower bound the edge expansion of the solution.

Exercise 11.4 (Second Laplacian Eigenvalue and Edge Expansion). *Let $G = (V, E)$ be an undirected graph. Prove that*

$$\lambda_2(L_G) \leq 2 \min_{0\leq|S|\leq|V|/2} \frac{|\delta(S)|}{|S|}.$$

Without the algebraic connectivity constraint, the above is a linear program and there is an elegant iterative rounding 2-approximation algorithm by Jain to solve the problem. With the algebraic connectivity constraint, the above becomes a convex program and it was not known how to handle both connectivity constraints and the algebraic connectivity constraint simultaneously. The observation in [LZ20] is that the one-sided spectral rounding result in Theorem 11.2 can be adapted to design an approximation algorithm for this problem.

Exercise 11.5 (Spectral Rounding for Network Design). *Let $x \in [0,1]^m$ be a fractional solution to the above convex program and L_x be its Laplacian matrix. Show that if $z \in \{0,1\}^m$ is an integral solution satisfying $L_z \succcurlyeq L_x$, then z satisfies the connectivity constraints and the algebraic connectivity constraint simultaneously.*

In [LZ20], Theorem 11.2 is extended to find $z \in \{0,1\}^m$ satisfying $L_z \succcurlyeq L_x$ with

$$\sum_{e\in E} c_e z_e \leq (1+O(\epsilon)) \cdot \sum_{e\in E} c_e x_e + O\left(\frac{n \cdot c_{\max}}{\epsilon}\right)$$

for any $0 < \epsilon < 1/4$, where $c_{\max} := \max_{e\in E}\{c_e\}$ is the maximum cost of an edge. This spectral rounding approach enlarges the set of constraints that one could handle in designing approximation algorithms for network design problems. We will discuss some technical ideas in the end of this chapter.

11.2 Barrier Method with Polynomials

The goal of this section is to present a proof of Theorem 11.2 using the polynomial perspective from [BSS14]. First, we will rephrase the barrier functions in Definition 10.3 in terms of polynomials. Then, we will present the plan following the intuition in the beginning of Chapter 10. Finally, we will proceed with the analysis and introduce some ideas about interlacing polynomials.

Soft-Max and Soft-Min of Polynomials

Recall the ϕ-soft-max and ϕ-soft-min in Definition 10.5 using the barrier functions in Definition 10.3. The strategy in the deterministic greedy Algorithm 7 in Chapter 10 is to fix ϕ_u, ϕ_l and then prove that $\phi_u\text{-}\max(A_t) \leq \phi_u\text{-}\max(A_{t-1}) + \delta_u$ and $\phi_l\text{-}\min(A_t) \geq \phi_l\text{-}\min(A_{t-1}) + \delta_l$ for all $t \geq 1$.

There are natural interpretations of the barrier functions from the polynomial perspective.

Remark 11.6 (Soft-Max and Soft-Min of Polynomials)**.** *Let $p_A(y) = \det(yI - A)$ be the characteristic polynomial of A. Note that*

$$\phi\text{-}\max(p_A) \;:=\; \phi\text{-}\max(A) \;=\; \max\left\{u \;\middle|\; \Phi^A(u) = \frac{p'_A(u)}{p_A(u)} = \phi\right\} \;=\; \lambda_{\max}\Big(p_A - \frac{1}{\phi}p'_A\Big)$$

and

$$\phi\text{-}\min(p_A) \;:=\; \phi\text{-}\min(A) \;=\; \min\left\{l \;\middle|\; \Phi_A(l) = -\frac{p'_A(l)}{p_A(l)} = \phi\right\} \;=\; \lambda_{\min}\Big(p_A + \frac{1}{\phi}p'_A\Big).$$

So, using the ϕ-soft-min to lower bound $\lambda_{\min}(p)$ can be understood as using the minimum root a related polynomial $p + \frac{1}{\phi}p'$ to lower bound $\lambda_{\min}(p)$. Actually, slightly more can be said.

Exercise 11.7 (Soft-Min)**.** *Let A be a real symmetric matrix. Show that $\lambda_{\min}(p_A) \geq \phi\text{-}\min(p_A) + \frac{1}{\phi}$ for any $\phi > 0$.*

Proof Plan

Given $v_1, \ldots, v_m \in \mathbb{R}^n$ and scalars $x_1, \ldots, x_m \in \mathbb{R}_{\geq 0}$ such that $\sum_{i=1}^m x_i v_i v_i^T = I_n$ and $\sum_{i=1}^m x_i = k$, our goal is to find $z_1, \ldots, z_m \in \mathbb{Z}_{\geq 0}$ with $\sum_{i=1}^m z_i \leq k$ and $\lambda_{\min}\big(\sum_{i=1}^m z_i v_i v_i^T\big)$ as large as possible.

Initially, we start with A_0 being the $n \times n$ zero matrix. In each iteration $1 \leq t \leq k$, we would like to find a vector $v \in \{v_1, \ldots, v_m\}$ and set $A_t = A_{t-1} + vv^T$. This will ensure that $A_t = \sum_{i=1}^m z_i v_i v_i^T$ for integers $z_1, \ldots, z_m$ for any $t \geq 0$.

As in Section 10.1, we consider the expected characteristic polynomial when we add a random vector with probability proportional to x_i. The following lemma is by the same calculation as in Section 10.1.

Exercise 11.8 (Expected Rank-One Update)**.** *The expected characteristic polynomial after we add a vector v_i with probability x_i/k is*

$$\mathbb{E}\left[p_{A+vv^T}\right] := \sum_{i=1}^{m} \frac{x_i}{k} \cdot p_{A+v_i v_i^T} = p_A - \frac{1}{k} p'_A.$$

Now, instead of considering the roots of $(1 - \frac{1}{k}\partial)^k x^n$ after k iterations as in Section 10.1, we would like to use ϕ-min to show that (i) the "expected progress" after one iteration is good and (ii) there is a vector v which achieves this expected progress. Concretely, the plan is to prove that there exists a vector v with

$$\phi\text{-}\min(p_{A+vv^T}) \geq \phi\text{-}\min\left(\mathbb{E}_x[p_{A+vv^T}]\right) = \phi\text{-}\min\left(p_A - \frac{1}{k} \cdot p'_A\right) \geq \phi\text{-}\min(p_A) + \frac{1}{k+\phi}. \tag{11.1}$$

The equality is from Exercise 11.8. We will prove the last inequality in the next subsection, and then the first inequality in the subsection after. Assume the two inequalities in Equation 11.1 always hold. Then, by induction, after k iterations,

$$\lambda_{\min}(p_{A_k}) \geq \frac{1}{\phi} + \phi\text{-}\min(p_{A_k}) \geq \frac{1}{\phi} + \phi\text{-}\min(p_{A_0}) + \frac{k}{k+\phi} = -\frac{n-1}{\phi} + \frac{k}{k+\phi},$$

where the first inequality is from Exercise 11.7. Some calculations show that choosing

$$\phi = \frac{(n-1)k}{\sqrt{(n-1)k} - (n-1)} \quad \Longrightarrow \quad \lambda_{\min}\left(\sum_{i \in S} v_i v_i^T\right) \geq \left(1 - \sqrt{\frac{n-1}{k}}\right)^2.$$

This proves Theorem 11.2. It remains to prove the two inequalities in Equation 11.1 in the following two subsections.

Shifting Lower Barrier

It turns out that the techniques developed for the barrier functions in Chapter 10 can be used to bound the maximum and minimum root of a real-rooted polynomial as well. The following lemma is from Lemma 4.3 of [MSS21], proving the last inequality in Equation 11.1. The proof is very similar to the proofs in Lemma 10.8 and Lemma 10.10 for the lower barrier function, just rephrased in the language of polynomials.

Lemma 11.9 (Lower Barrier Shift [MSS21])**.** *If p is real-rooted and $s, \phi > 0$, then $p - sp'$ is real rooted and*

$$\phi\text{-}\min\left(p - sp'\right) \geq \phi\text{-}\min(p) + \frac{1}{\frac{1}{s} + \phi}.$$

Proof. It is well known and we will prove it in the next chapter that $p - sp'$ is real-rooted if p is. Let $l = \phi\text{-}\min(p)$ such that l is the minimum value with $\Phi_p(l) = \phi$. Let

$$\delta := \frac{1}{\frac{1}{s} + \phi}.$$

To prove the lemma, we will prove that (i) $l+\delta < \lambda_{\min}(p - sp')$ and (ii) $\Phi_{p-sp'}(l+\delta) \le \phi$, and this would imply that $\phi\text{-}\min(p - sp') \ge l + \delta = \phi\text{-}\min(p) + \delta$.

For (i), we claim that

$$\lambda_{\min}(p - sp') \ge \lambda_{\min}(p) \ge \phi\text{-}\min(p) + \frac{1}{\phi} > l + \delta.$$

The second inequality is from Exercise 11.7 and the third inequality is from the definition of δ. To see the first inequality, note that $p(y)$ and $-s \cdot p'(y)$ with $s > 0$ have the same sign for all $y < \lambda_{\min}(p)$, and thus any $y < \lambda_{\min}(p)$ cannot be a root of $p(y) - s \cdot p'(y)$, which implies that $\lambda_{\min}(p - sp') \ge \lambda_{\min}(p)$.

For (ii), we write $\Phi_{p-sp'}$ in terms of $\Phi_p = -p'/p$ so that

$$\Phi_{p-sp'} = -\frac{(p-sp')'}{p-sp'} = -\frac{((1+s\Phi_p)p)'}{(1+s\Phi_p)p} = -\frac{p'}{p} - \frac{s\Phi_p'}{1+s\Phi_p} = \Phi_p - \frac{\Phi_p'}{\frac{1}{s}+\Phi_p},$$

whenever all the quantities are finite, which happens everywhere except at the roots of p and $p - sp'$. Since $l+\delta$ is below the roots of p and $p - sp'$, it follows that

$$\Phi_{p-sp'}(l+\delta) = \Phi_p(l+\delta) - \frac{\Phi_p'(l+\delta)}{\frac{1}{s}+\Phi_p(l+\delta)} = \Phi_p(l) + \underbrace{(\Phi_p(l+\delta) - \Phi_p(l))}_{\text{loss}} - \underbrace{\frac{\Phi_p'(l+\delta)}{\frac{1}{s}+\Phi_p(l+\delta)}}_{\text{gain}}.$$

Therefore,

$$\Phi_{p-sp'}(l+\delta) \le \Phi_p(l) = \phi \quad \iff \quad \Phi_p(l+\delta) - \Phi_p(l) \le \frac{\Phi_p'(l+\delta)}{\frac{1}{s}+\Phi_p(l+\delta)}. \tag{11.2}$$

As in Lemma 10.10, using convexity of $\Phi_p(l)$ will get us close but not enough; see Remark 11.10. So we need to work a bit harder as in Lemma 10.10. Using $\frac{1}{s} = \frac{1}{\delta} - \phi$ and rearranging, the condition in Equation 11.2 is equivalent to

$$\big(\Phi_p(1+\delta) - \Phi_p(l)\big)^2 \le \Phi_p'(l+\delta) - \frac{1}{\delta}\big(\Phi_p(1+\delta) - \Phi_p(l)\big).$$

This is exactly what Claim 10.11 proved, which completes the proof of (ii) that $\Phi_{p-sp'}(l+\delta) \le \phi$. □

Remark 11.10 (Convexity Not Enough)**.** *Using convexity* $\Phi_p(l+\delta) - \Phi_p(l) \le \delta \cdot \Phi_p'(l+\delta)$ *in Exercise 10.4, the condition in Equation 11.2 holds if*

$$\delta \cdot \Phi_p'(l+\delta) \le \frac{\Phi_p'(l+\delta)}{\frac{1}{s}+\Phi_p(l+\delta)} \quad \iff \quad \delta \le \frac{1}{\frac{1}{s}+\Phi_p(l+\delta)},$$

but we cannot conclude that δ *being* $1/\big(\frac{1}{s}+\Phi_p(l)\big) = 1/\big(\frac{1}{s}+\phi\big)$ *suffices to maintain the nonincreasing potential. This is the same situation as in Lemma 10.10.*

Common Interlacing

Now we would like to prove the first inequality in Equation 11.1. In general, given real-rooted polynomials $p_1, \ldots, p_m$ and a convex combination $q := \sum_{i=1}^m \mu_i p_i$ of them, there may not be any

useful relations between the roots of $p_1, \ldots, p_m$ and q. For example, q may not have real roots even if $p_1, \ldots, p_m$ are all real-rooted, so we cannot hope to prove inequalities such as $\max_{1 \le i \le m}\{\lambda_{\min}(p_i)\} \ge \lambda_{\min}(q)$. We will discuss more in the next chapter.

An important observation of Marcus, Spielman, and Srivastava is that if $p_1, \ldots, p_m$ have a "common interlacing", then we can relate the roots of $p_1, \ldots, p_m$ and the roots of q and prove the inequality $\max_{1 \le i \le m}\{\lambda_{\min}(p_i)\} \ge \lambda_{\min}(q)$ and more. We will introduce interlacing polynomials properly in the next chapter. After reading the next chapter, it will be a good exercise to prove that the polynomials

$$p_1 := p_{A+v_1v_1^T} + \frac{1}{\phi}p'_{A+v_1v_1^T}, \quad \ldots, \quad p_m := p_{A+v_mv_m^T} + \frac{1}{\phi}p'_{A+v_mv_m^T}$$

have a common interlacing, to establish the first inequality in Equation 11.1.

Problem 11.11 (Interlacing Property of ϕ-min)**.** *If $p_1, \ldots, p_m$ are real-rooted polynomials that have a common interlacing, then for any expected polynomial $q = \sum_{i=1}^m \mu_i p_i$ with $\sum_{i=1}^m \mu_i = 1$ and $\mu_i \ge 0$ for $1 \le i \le m$, there exists $i \in [m]$ with*

$$\phi\text{-}\min(p_i) \ge \phi\text{-}\min(q).$$

Assuming Problem 11.11, we have completed the proof of Theorem 11.2 using polynomials.

11.3 Regret Minimization

The original proof of Theorem 11.2 is based on the regret minimization framework developed in [ALO15]. The general idea in regret minimization is to find some distributions of experts that are almost as good as the best experts. I won't be able to introduce regret minimization properly, but let me try to give an informal high level idea of the regret minimization framework in the specific setting of spectral rounding.

In the minimum eigenvalue problem in Theorem 11.2, the objective is to find a (multi-)subset S with large $\lambda_{\min}\big(\sum_{i\in S} v_iv_i^T\big)$, or equivalently a (multi-)subset S such that $x^T\big(\sum_{i\in S} v_iv_i^T\big)x$ is large for all vectors $x \in \mathbb{R}^n$ on the unit sphere. In this problem, we think of each direction x on the unit sphere as an expert. In each iteration t, given the current solution A_t, the regret minimization framework would maintain a "smart" probability distribution μ_t on the unit sphere, which puts higher probability on x if x^TA_tx is small and a lower probability on x if x^TA_tx is large. In words, the probability distribution puts more focus on the directions that the current solution A_t has not covered well. The distribution is summarized succinctly by a density matrix $P_t = \int xx^T d\mu_t$. This density matrix guides us naturally to add a vector v_t that maximizes the inner product $\langle v_iv_i^T, P_t\rangle$ to A_t, to cover the directions that are not covered well. The analysis in the regret minimization framework proves that if the probability distributions μ_t are smart, then the "regret"

$$\sum_{t\ge1}\langle v_tv_t^T, P_t\rangle - \min_{x\in\mathbb{R}^n:\|x\|=1}\sum_{t\ge1}\langle v_tv_t^T, xx^T\rangle$$

of using P_t over time instead of focusing on the worst directions (or best experts) is small. So, if we could always find a vector v_t in each iteration with a large inner product with P_t, which we can because of the isotropy condition, then we can conclude that $\sum_t\langle v_tv_t^T, P_t\rangle$ is large and hence $\min_x \sum_{t\ge1}\langle v_tv_t^T, xx^T\rangle = \min_x x^T\big(\sum_{t\ge1} v_tv_t^T\big)x = \lambda_{\min}\big(\sum_{t\ge1} v_tv_t^T\big)$ is also large.

A versatile and commonly used approach to maintain the distributions is by the multiplicative weight update method. If we use it for spectral sparsification, then we can recover the $O(n \log n/\epsilon^2)$ result by Spielman and Srivastava in Theorem 9.9.

The insight in [ALO15] is that the barrier functions used by Batson, Spielman and Srivastava in Definition 10.3 can be interpreted as a new way of updating the probability distributions, by setting $P_t = (\phi A_t - l_t I)^{-2}$ where A_t is the current solution and l_t is chosen such that $P_t \succcurlyeq 0$ and $\mathrm{Tr}(P_t) = 1$. This can be integrated into the regret minimization framework to give the following regret bound:

$$\lambda_{\min}\left(\sum_{t=1}^{\tau} v_t v_t^T\right) \geq \underbrace{-\frac{2\sqrt{n}}{\phi}}_{\text{initial lower bound}} + \sum_{t=1}^{\tau} \underbrace{\frac{\langle v_t v_t^T, P_t\rangle}{1+\phi\langle v_t v_t^T, P_t^{1/2}\rangle}}_{\text{increase of } \phi\text{-soft-min}} \tag{11.3}$$

where we should interpret ϕ as the same parameter in soft-min, the negative term is the initial lower bound in the barrier method, and each term in the summation as the increase of the ϕ-soft-min. Using the isotropy condition $\sum_{i=1}^m v_i v_i^T = I_n$, it is possible to show that there always exists a vector with a large increase of ϕ-soft-min to prove Theorem 11.2.

Local Search

In applications of spectral rounding or the minimum eigenvalue problem, often we are given $x \in [0,1]^m$ and we would like to find $z \in \{0,1\}^m$ instead of just $z \in \mathbb{Z}^m$. This is called the "without repetition" setting, where each vector can be chosen at most once, which is a more general setting than the "with repetition" setting, where each vector can be chosen more than once. In [ALSW17], it was shown that the same greedy algorithm can only achieve a constant factor approximation algorithm, not the $(1\pm\epsilon)$-approximation algorithm in Problem 11.3 when $k \geq n/\epsilon^2$.

An interesting new idea in [ALSW21] is to analyze a local search algorithm, where we start from an arbitrary subset S_0 with k vectors, and in each iteration $t \geq 1$ we find a pair $i_{t-1} \in S_{t-1}$ and $j_{t-1} \notin S_{t-1}$ and set $S_t := S_{t-1} - i_{t-1} + j_{t-1}$. This guarantees that each vector is chosen at most once. They developed the following rank-two update formula for regret minimization:

$$\lambda_{\min}\left(\sum_{l\in S_t} v_l v_l^T\right) \geq -\frac{2\sqrt{n}}{\phi} + \sum_{t=1}^{\tau}\left(\frac{\langle v_{j_t} v_{j_t}^T, P_t\rangle}{1+2\phi\langle v_{j_t} v_{j_t}^T, P_t^{1/2}\rangle} - \frac{\langle v_{i_t} v_{i_t}^T, P_t\rangle}{1-2\phi\langle v_{i_t} v_{i_t}^T, P_t^{1/2}\rangle}\right), \tag{11.4}$$

and used it to get the same result as in Theorem 11.2 in the more challenging without repetition setting. It would be interesting to recover this result using the polynomial approach in Section 11.2. See Problem 11.14 for a possible starting point.

Randomized Local Search

The local search approach in [ALSW21] is extended in [LZ20] to handle linear constraints as described in Definition 11.1. The idea is to randomly choose a vector v_{i_t} to remove from S_{t-1} with probability proportional to $(1-x_i)\cdot\left(1+2\alpha\langle v_{j_t} v_{j_t}^T, P_t^{1/2}\rangle\right)$, where x_i is the fractional value of the i-th vector and the other term is in the denominator in Equation 11.4, and similarly choose a vector v_{j_t} to add to S_{t-1} with probability proportional $x_j \cdot \left(1-2\phi\langle v_{i_t} v_{i_t}^T, P_t^{1/2}\rangle\right)$. Informally, the terms $(1-x_i)$ and x_j are to ensure that the linear constraints are approximately preserved, and the terms from the denominators in Equation 11.4 are to ensure that the minimum eigenvalue is improving. The proof is by showing that these sums are concentrated around their expected values.

In [LZ21], it was shown that this randomized local search approach gives the best known algorithms for experimental design problems, where for D/A-design the randomized local search algorithm achieves a $(1\pm\epsilon)$-approximation when $k \gtrsim n/\epsilon$, better than the requirement $k \gtrsim n/\epsilon^2$ for E-design.

Again, it would be interesting to recover these results using the polynomial approach in Section 11.2.

Two-Sided Spectral Rounding

We will study the two-sided spectral rounding problem in Definition 11.1 in the second part of the course.

11.4 Problems

Problem 11.12 (Total Effective Resistance). *Let L_G be the Laplcian matrix of a graph $G=(V,E)$. Recall that $\mathrm{Reff}_G(u,v) = (\chi_u-\chi_v)^T L_G^\dagger(\chi_u-\chi_v)$ is the effective resistance between vertices $u,v\in V$. Show that $|V|\cdot \mathrm{Tr}(L_G^\dagger) = \frac{1}{2}\sum_{u,v\in V}\mathrm{Reff}_G(u,v)$. Use this fact with Theorem 11.2 to obtain an approximation algorithm for minimizing the total effective resistance subject to the constraint that the (multi-)subgraph has at most k edges.*

Problem 11.13 (Upper Barrier Shift). *Prove the following analog of Lemma 11.9 for the upper barrier function. If p has real roots and $s,\phi>0$, then $p-sp'$ is real-rooted and*

$$\phi\text{-}\max\left(p-sp'\right) \le \phi\text{-}\max(p) + \frac{1}{\frac{1}{s}-\phi}.$$

Problem 11.14 (Expected Polynomial After Removal). *This problem might be helpful in obtaining the bound in Theorem 11.2 in the more challenging without repetition setting using the polynomial approach. Suppose the current solution A has k vectors say $v_1,\ldots,v_k\in\mathbb{R}^n$. Show that the expected characteristic polynomial after removing a uniformly random vector is*

$$\mathbb{E}\left[p_{A-vv^T}(y)\right] := \frac{1}{k}\sum_{i=1}^k p_{A-v_iv_i^T}(y) = \left(1-\frac{n}{k}\right)\cdot p_A(y) + \frac{y}{k}\cdot p_A'(y).$$

Question 11.15 (Improved Approximation Ratio when $k=n$). *Is it possible to improve the $\Theta(1/n^2)$ minimum eigenvalue bound in Theorem 11.2 when $k=n$?*

11.5 References

[ALO15] Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 237–245. ACM, 2015. 105, 108, 113, 114

[ALSW17] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal design of experiments via regret minimization. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 126–135. PMLR, 2017. 107, 108, 114

[ALSW21] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal discrete optimization for experimental design: a regret minimization approach. *Math. Program.*, 186(1):439–478, 2021. 108, 109, 114

[BSS14] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Rev.*, 56(2):315–334, 2014. 97, 98, 99, 100, 103, 108, 110, 117

[LZ20] Lap Chi Lau and Hong Zhou. A spectral approach to network design. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 826–839. ACM, 2020. 109, 110, 114, 167

[LZ21] Lap Chi Lau and Hong Zhou. A local search framework for experimental design. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1039–1058. SIAM, 2021. 108, 109, 115

[MSS21] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families iii: Sharper restricted invertibility estimates. *Israel Journal of Mathematics*, pages 1–28, 2021. 104, 111, 119, 122, 123, 135

Chapter 12

Method of Interlacing Polynomials

Marcus, Spielman, and Srivastava [MSS14] turned the heuristic argument from [BSS14] about expected characteristic polynomial described in Section 10.1 into a powerful probabilistic method. We have already previewed this method in a simple form in Section 11.2 without seeing the details. In this chapter, we will go through the relevant concepts and describe the method in its general form. Then we will see an interesting and relatively simple application to the restricted invertibility problem, in two different ways.

12.1 New Probabilistic Method

In standard probabilistic method, we compute the expectation of a random variable $\mathbb{E}[X]$, and then conclude that there is an outcome in the sample space with value at least or at most $\mathbb{E}[X]$. Consider the minimum eigenvalue problem in Theorem 11.2, in which the quantity of interest is $\lambda_{\min}(\sum_{i\in S} v_i v_i^T)$ for some multi-subset S with $|S| = k$. To prove that there is a multi-subset with large minimum eigenvalue, the standard way is to compute $\mu := \mathbb{E}_{S:|S|=k}\big[\lambda_{\min}\big(\sum_{i\in S} v_i v_i^T\big)\big]$ and then conclude that there is a mult-subset S with $|S| = k$ and $\lambda_{\min}(\sum_{i\in S} v_i v_i^T) \geq \mu$.

Marcus, Spielman, and Srivastava took an unusual route to solve this kind of problems. First, instead of working with the random matrix $A = \sum_{i\in S} v_i v_i^T$ directly, they consider the characteristic polynomial $p_A(x) = \det(xI - A)$ of the random matrix. Note that $\lambda_{\min}(A)$ is simply the minimum root of the characteristic polynomial $\lambda_{\min}(p_A)$. Then, quite surprisingly, instead of computing the expected minimum eigenvalue of a random characteristic polynomial $\mathbb{E}_A[\lambda_{\min}(p_A)]$, they compute the minimum eigenvalue of the expected polynomial $\lambda_{\min}\big(\mathbb{E}_A[p_A]\big)$. The following is an instantiation of their new probabilistic method for the minimum eigenvalue problem, when each vector is chosen independently and uniformly randomly.

Proposition 12.1 (Probabilistic Method for Minimum Eigenvalue)**.** *Suppose $v_1, \ldots, v_m \in \mathbb{R}^n$ are vectors with $\sum_{i=1}^m v_i v_i^T = I_n$. For any $k \geq n$, suppose $r_1, \ldots, r_k$ are independent uniformly random vectors in $\{v_1, \ldots, v_m\}$ and let $A := \sum_{i=1}^k r_i r_i^T$ be a random matrix. Then, with positive probability,*

$$\lambda_{\min}(p_A) \geq \lambda_{\min}\big(\mathbb{E}[p_A]\big).$$

In general, $\mathbb{E}[\lambda_{\min}(p_A)] \neq \lambda_{\min}(\mathbb{E}[p_A])$, and in fact the latter term could be bigger than the former term, and so this proposition is not trivial at all.

Characteristic polynomials have not played an important role in much of spectral graph theory. One disadvantage for instance is that the information about the eigenvectors is lost. Very interestingly,

the method by Marcus, Spielman, and Srivastava showed that they often satisfy a number of very nice algebraic identities and are amenable to a set of very elegant analytic tools that do not naturally apply to matrices.

12.2 Interlacing Polynomials

Let $p_1, \ldots, p_m$ be real-rooted polynomials and $q = \sum_{i=1}^m \mu_i p_i$ be a convex combination of $p_1, \ldots, p_m$ where $\sum_{i=1}^m \mu_i = 1$ and $\mu_i \geq 0$ for $1 \leq i \leq m$. Under what conditions can we conclude that say $\max_i \lambda_{\min}(p_i) \geq \lambda_{\min}(q)$? In general, it could be far from true. For example, $p_1 = (x-1)(x-2)$ and $p_2 = (x-3)(x-4)$ are both real-rooted, but their average $\frac{1}{2}(p_1 + p_2)$ is not even real-rooted. Even if assume $p_1 + p_2$ is real-rooted, there is in general no simple relationship between the roots of two polynomials and the roots of their average.

The main insight of Marcus, Spielman and Srivastava is that in several problems of interest, the (characteristic) polynomials satisfy some interlacing properties that would allow us to conclude that $\max_i \lambda_{\min}(p_i) \geq \lambda_{\min}(q)$.

Definition 12.2 (Interlacing Polynomials)**.** *Let p be a degree n polynomial with real roots $\alpha_1 \geq \ldots \geq \alpha_n$ and let q be a degree n or $n-1$ polynomial with real roots $\beta_1 \geq \ldots \geq \beta_n$ (ignoring β_n in the degree $n-1$ case). We say that q interlaces p if their roots alternate and the largest root belongs to p such that*

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \ldots \beta_{n-1} \geq \alpha_n \geq \beta_n.$$

Definition 12.3 (Common Interlacing)**.** *A set of degree n real-rooted polynomials $p_1, \ldots, p_m$ is said to have a common interlacing if there is a polynomial q that interlaces each p_i for $1 \leq i \leq m$.*

Equivalently, $p_1, \ldots, p_m$ have a common interlacing if there are inner-disjoint intervals $I_1 \geq I_2 \geq \ldots \geq I_n$ on the real line such that the k-th largest root of each p_i for $1 \leq i \leq m$ is contained in I_k

An important class of interlacing polynomials are characteristic polynomials of matrices under rank-one updates. The following is also called Cauchy's interlacing theorem, and one can prove it in a similar way as in Cauchy's interlacing Theorem 2.13, using Courant-Fischer Theorem 2.12.

Exercise 12.4 (Cauchy's Interlacing Theorem)**.** *Let $A \in \mathbb{R}^{n \times n}$ be a real symmetric matrix and $v \in \mathbb{R}^n$. Then p_A interlaces p_{A+vv^T}.*

Note that Exercise 12.4 implies that if $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix and $v_1, \ldots, v_m \in \mathbb{R}^n$, then $p_{A+v_1v_1^T}, \ldots, p_{A+v_mv_m^T}$ have a common interlacing.

Common Interlacing and Probabilistic Method

If $p_1, \ldots, p_m$ have a common interlacing, then any convex combination q of $p_1, \ldots, p_m$ is also real-rooted and we can compare the roots of $p_1, \ldots, p_m$ with the roots of q. The proof is a simple application of the intermediate value theorem in the interval I_j for the j-th root for each j.

Theorem 12.5 (Probabilistic Method for Common Interlacing Polynomials)**.** *Suppose $p_1, \ldots, p_m$ are real-rooted polynomials of degree n with positive leading coefficients. Let $\lambda_k(p_j)$ be the k-th largest root of p_j. If $p_1, \ldots, p_m$ have a common interlacing, then for any non-negative numbers $\mu_1, \ldots, \mu_m$ with $\sum_{i=1}^m \mu_i = 1$ and for any $1 \leq k \leq n$,*

$$\min_j \lambda_k(p_j) \leq \lambda_k\big(\mathbb{E}_{j \sim \mu}[p_j]\big) \leq \max_j \lambda_k(p_j)$$

Proof. Let $q = \sum_{i=1}^{m} \mu_i p_i$. Let $u = \max_j \lambda_{\max}(p_j)$ and $l = \min_j \lambda_{\max}(p_j)$. We would like to argue that $\lambda_{\max}(q)$ is contained in $[l, u]$.

First, we argue that $\lambda_{\max}(q) \leq u$. As $p_1, \ldots, p_m$ all have positive leading coefficients, all polynomials are positive in the range (u, ∞). As q is a convex combination of $p_1, \ldots, p_m$, so q is also positive in the range (u, ∞). Therefore, q cannot have a root in the range (u, ∞), and thus $\lambda_{\max}(q) \leq u$.

Next, we argue that $\lambda_{\max}(q) \geq l$. If $l = u$, then $p_j(u) = 0$ for all $1 \leq j \leq m$, and thus $q(u) = 0$ and hence $\lambda_{\max}(q) = u \geq l$. Henceforth we assume $l < u$. On one hand, note that $q(u) > 0$ as each $p_j(u) \geq 0$ and there exists i with $p_i(u) > 0$ (e.g. the one with $\lambda_{\max}(p_i) = l < u$). On the other hand, since $p_1, \ldots, p_m$ have a common interlacing, $\lambda_2(p_j) \leq l$ for each j, and since $p_1, \ldots, p_m$ all have positive leading coefficients, each polynomial p_j is non-positive in the range $[\lambda_2(p_j), \lambda_1(p_j)]$ with $\lambda_2(p_j) \leq l$ and $\lambda_1(p_j) \geq l$ for all $1 \leq j \leq m$. Therefore, $p_j(l) \leq 0$ for all $1 \leq j \leq m$, and thus $q(l) \leq 0$. Since $q(u) > 0$ and $q(l) \leq 0$, by the intermediate value theorem, there exists $r \in [l, u)$ such that $q(r) = 0$, and therefore $\lambda_{\max}(q) \geq l$.

A similar argument works for any $1 \leq k \leq n$ and is left to the reader (see Lemma 2.11 of [MSS21]). It may be more convenient for the argument to first reduce to the case when $p_1, \ldots, p_m$ have no common roots. □

So, if we could show that a set of polynomials have a common interlacing, then we can apply Theorem 12.5 to show that there exists a polynomial with large/small k-th largest root by showing that *some* weighted average polynomial has large/small k-th largest root.

Common Interlacing and Real-Rootedness

We are thus interested in some general techniques to prove that a set of polynomials have a common interlacing. Note that Theorem 12.5 proves that if $p_1, \ldots, p_m$ are real-rooted and have a common interlacing, then any convex combination of $p_1, \ldots, p_m$ is also real-rooted. It turns out that the converse is also true. This gives us a characterization when a set of real-rooted polynomials have a common interlacing. We use the following simple fact in the proof.

Exercise 12.6 (Common Interlacing is a Pairwise Property)**.** *A set of polynomials $p_1, \ldots, p_m$ have a common interlacing if and only if each pair of polynomials p_i, p_j have a common interlacing for all $1 \leq i \neq j \leq n$.*

We also use the following well-known result from elementary complex analysis without proof.

Theorem 12.7 (Continuity of Roots)**.** *The roots of a polynomial are continuous functions of its coefficients.*

Theorem 12.8 (Common Interlacing and Real-Rootedness)**.** *If $p_1, \ldots, p_m$ are degree n polynomials and all of their convex combinations $\sum_{i=1}^{m} \mu_i p_i$ are real-rooted, then $p_1, \ldots, p_m$ have a common interlacing.*

Proof. By Exercise 12.6, we only need to prove the lemma for two polynomials. We assume without loss of generality that p_1 and p_2 have no common roots, as otherwise we can just divide both polynomials by their common factors, prove that the resulting polynomials have a common interlacing, and conclude that the original polynomials also have a common interlacing.

Let $q_\mu = (1-\mu)\cdot p_1 + \mu \cdot p_2$ for $\mu \in [0,1]$. If we keep track of the roots of q_μ from $\mu = 0$ and $\mu = 1$ as a continuous function of μ, then each root of q_μ is a continuous curve on the complex plane as μ varies from 0 to 1 by Theorem 12.7. Since each q_μ is real-rooted by assumption, the curve of each root j is an interval J_j on the real line, with one endpoint being a root of p_1 and the other endpoint being a root of p_2.

We would like to argue that these intervals are pairwise inner-disjoint (i.e. they are disjoint except possibly at the endpoints). Suppose to the contrary that this is not the case, that one endpoint of an interval is contained in the interior of some other interval. This implies that some root r of a polynomial, say p_1, is a root of q_μ for some $0 < \mu < 1$, but then

$$0 = q_\mu(r) = (1-\mu)\cdot p_1(r) + \mu \cdot p_2(r) = \mu \cdot p_2(r) \quad \Longrightarrow \quad p_2(r) = 0,$$

contradicting that p_1 and p_2 have no common roots. Therefore, these intervals are pairwise inner-disjoint. This implies that the intervals can be arranged so that $J_1 \geq J_2 \geq \ldots \geq J_n$, and thus p_1 and p_2 have a common interlacing. □

By Theorem 12.8, to prove a set of polynomials have a common interlacing (in order to apply the probabilistic method), it is equivalent to proving that all convex combinations of any two polynomials are real-rooted. In the next chapter, we will study methods to prove that a polynomial is real-rooted.

Differential Operators and Common Interlacing

In Section 10.1 and in Section 11.2, the expected characteristic polynomials are of the form $(1-s\partial)p$ for some scalar s. With the results in the previous subsections, we can show that this differential operator preserves real-rootedness and also common interlacing.

Problem 12.9 (Differential Operators and Common Interlacing)**.** *Prove that if p is real-rooted, then $(1-s\partial)p$ is also real-rooted. Furthermore, prove that if $p_1, \ldots, p_m$ have positive leading coefficients and a common interlacing, then $(1-s\partial)p_1, \ldots, (1-s\partial)p_m$ also have a common interlacing.*

With Problem 12.9, it should be straightforward to solve Problem 11.11, and thus completing the proof of Theorem 11.2 using a simple version of the method of interlacing polynomials.

12.3 Interlacing Family

Recall that in Proposition 12.1, the goal is to prove that there is a positive probability that a random matrix $A = \sum_{i=1}^k r_i r_i^T$ satisfies $\lambda_{\min}(p_A) \geq \lambda_{\min}\big(\mathbb{E}[p_A]\big)$, where there are m^k possibilities of A. To prove this statement by directly applying the probabilistic method in Theorem 12.5, we need to prove that these m^k different characteristic polynomials have a common interlacing. A moment of thought reveals that this is clearly not true in general.

The idea of Marcus, Spielman, and Srivastava is to build a tree structure among these polynomials and show that the children of each internal node have a common interlacing. This idea is similar to the method of conditional expectations used in derandomization.

Definition 12.10 (Interlacing Family)**.** *An interlacing family consists of a finite rooted tree T and a labeling of the nodes $v \in T$ by monic real-rooted polynomials $p_v(x) \in \mathbb{R}[x]$, with two properties:*

1. *Every polynomial $p_v(x)$ corresponding to a non-leaf node v is a convex combination of the polynomials corresponding to the children of v.*
2. *For all nodes $v_1, v_2 \in T$ with a common parent, all convex combinations of $p_{v_1}(x)$ and $p_{v_2}(x)$ are real-rooted.*

We say that a set of polynomials is an interlacing family if they are the labels of the leaves of such a tree.

Note that, by Theorem 12.8 and Exercise 12.6, the second condition implies that all the children have a common interlacing, and it follows from Theorem 12.5 that all convex combinations of all children are real-rooted.

The above definition may look a bit abstract, but in applications the root polynomial will usually simply be the average polynomial of all the leaves, while the internal nodes will usually simply be the average polynomial of the leaves of the corresponding subtrees. Let us see a concrete example that is useful for the minimum eigenvalue problem in Theorem 11.2.

Example 12.11 (Interlacing Family of Multi-Subset of k Vectors)**.** *Let $v_1, \ldots, v_m \in \mathbb{R}^n$. For any $s_1, \ldots, s_k \in [m]$, define*

$$p_{s_1,\ldots,s_k}(x) := \det\left(xI_n - \sum_{i=1}^{k} v_{s_i} v_{s_i}^T\right).$$

The tree T is a complete m-ary tree, with depth k, and thus m^k leaves. Each leaf of the tree is labeled by a sequence $s_1, \ldots, s_k$, representing a path from the root to the leaf, where s_i represents the s_i-th child of the internal node in the $(i-1)$-th level, with the root being in the 0-th level. The polynomials in the internal nodes are defined inductively as

$$p_{s_1,\ldots,s_t}(x) = \frac{1}{m}\sum_{j=1}^{m} p_{s_1,\ldots,s_t,j}(x) = \frac{1}{m^{k-t}} \sum_{s_{t+1},\ldots,s_k} p_{s_1,\ldots,s_k}(x)$$

for any $t < k$ and the root polynomial is

$$p_\emptyset(x) = \frac{1}{m^k} \sum_{s_1,\ldots,s_k \in [m]^k} p_{s_1,\ldots,s_k}(x).$$

We will prove in the next chapter that these polynomials $\mathcal{P} := \left\{p_{s_1,\ldots,s_k}(x)\right\}_{s_1,\ldots,s_k \in [m]^k}$ form an interlacing family.

It may not be easy to establish that a set of polynomials forms an interlacing family, and in some applications the theory of real stable polynomials is needed to prove so, which we will study in the next chapter.

But once we have established that a family is an interlacing family, we can then easily relate the roots of the root-polynomial to the roots of the polynomials in the leaves. The following theorem follows from a simple induction using Theorem 12.5.

Theorem 12.12 (Probabilistic Method for Interlacing Family)**.** *Let $\mathcal{P}$ be an interlacing family of degree n polynomials with root labeled by $p_\emptyset(x)$ and leaves by $\{p_l(x)\}_{l \in L}$ where L is the set of leaves. Then, for any $1 \le j \le n$, there exist leaves $a \in L$ and $b \in L$ such that*

$$\lambda_j(p_a) \le \lambda_j(p_\emptyset) \le \lambda_j(p_b).$$

Proof. The proof is by a simple induction on the depth of the internal node. By Theorem 12.8, the second condition in Definition 12.10 implies that every pair of children of the root node have a common interlacing. By Exercise 12.6, it follows that all the children of the root node have a common interlacing. Then, Theorem 12.5 proves that there is a child a_1 of the root node with $\lambda_j(p_{a_1}) \leq \lambda_j(p_\emptyset)$ and there is a child b_1 of the root node with $\lambda_j(p_{b_1}) \geq \lambda_j(p_\emptyset)$. By induction, there is a leaf node a in the subtree of a_1 with $\lambda_j(p_a) \leq \lambda_j(p_{a_1}) \leq \lambda_j(p_\emptyset)$, and there is a leaf node b in the subtree of b_1 with $\lambda_j(p_b) \geq \lambda_j(p_{b_1}) \geq \lambda_j(p_\emptyset)$. □

12.4 Restricted Invertibility

In this section, we see an interesting application of the techniques developed so far to the restricted invertibility problem. This is not the first application of the method of interlacing family, but it is the simplest as it only involves univariate polynomials, and so we present it first to separate the ideas of the interlacing family method from the theory of real-stable (multivariate) polynomials.

The restricted invertibility problem is a well-studied problem in functional analysis, which says that a matrix of high stable rank has a large column submatrix with large smallest singular value. We consider an equivalent formulation that is very close to the minimum eigenvalue problem in Theorem 11.2.

Definition 12.13 (Restricted Invertibility Problem)**.** *Given $v_1, \ldots, v_m \in \mathbb{R}^n$ and an integer $k < n$, find a subset $S \subseteq [m]$ with $|S| = k$ to maximize $\lambda_k\big(\sum_{i\in S} v_i v_i^T\big)$, where $\lambda_k(A)$ denotes the k-th largest eigenvalue of matrix A.*

To illustrate the method of interlacing family, we only consider the special "isotropy" case when $\sum_{i=1}^m v_i v_i^T = I_n$. We remark that, unlike the minimum eigenvalue problem, it is no longer true that the general case can be reduced to this special case, because of $k < n$. Marcus, Spielman, and Srivastava [MSS21] used the method of interlacing family to derive a sharp result in the isotropy case.

Theorem 12.14 (Restricted Invertibility in the Isotropy Case)**.** *Suppose $v_1, \ldots, v_m \in \mathbb{R}^n$ are vectors with $\sum_{i=1}^m v_i v_i^T = I_n$. Then, for every integer $k \leq n$, there exists a subset $S \subset [m]$ with $|S| = k$ and*

$$\lambda_k\left(\sum_{i\in S} v_i v_i^T\right) \geq \left(1 - \sqrt{\frac{k}{n}}\right)^2 \frac{n}{m}.$$

Although this result is sharp for a large regime of k, we do not know whether it is tight when $k \approx n$. The following question is closely related to Question 11.15

Question 12.15 (Restricted Invertibility when $k \approx d$)**.** *When $m = O(n)$ and $k = n - 1$, the lower bound in Theorem 12.14 is $\Omega(1/n^2)$. Is this tight? To my knowledge, the best lower bound that we can hope for in this regime is $\Omega(1/n)$.*

Ravichandran [Rav18] presented a different way to use the interlacing family method to derive the results in [MSS21], with an additional application of proving a quantitative Gauss-Lucas theorem which we may mention later. We will present both approaches, as this will allow us to see two different interlacing families for the problem.

Interlacing Family of Multi-Subset of Vectors

The proof in [MSS21] uses the interlacing family that we described in Example 12.11. We have not proved that it is indeed an interlacing family yet, but we assume it is in this subsection. Then, to apply the probabilistic method in Theorem 12.12, we just need to compute the polynomial in the root of the tree and bound its k-th eigenvalue. The calculations for the expected characteristic polynomial in Section 10.1 and in Exercise 11.8 can be used to compute the root polynomial from the leaves up.

Exercise 12.16 (Root Polynomial in Example 12.11). *When $\sum_{i=1}^m v_i v_i^T = I_n$, the root polynomial $p_\emptyset$ in Example 12.11 is*

$$p_\emptyset(x) = \left(1 - \frac{1}{m}\partial_x\right)^k x^n$$

Note that the k-th largest root of $p_\emptyset(x)$ is simply the smallest root of the polynomial $x^{-(n-k)}p_\emptyset(x) = x^{-(n-k)}\left(1-\frac{1}{m}\partial_x\right)^k x^n$. Marcus, Spielman, and Srivastava observed that it is a slight transformation of an associated Laguerre polynomial and a known result by Krasikov implies that

$$\lambda_k(p_\emptyset) \geq \left(1 - \sqrt{\frac{k}{n}}\right)^2 \frac{n}{m}.$$

Therefore, we can conclude from Theorem 12.12 that there is a leaf with the k-th largest eigenvalue at least $\lambda_k(p_\emptyset)$, proving Theorem 12.14.

It is quite amazing that the method of interlacing family reduces the restricted invertibility problem to a pure mathematical problem of bounding the smallest root of a well-known polynomial. So, the heuristic argument in Section 10.1 can indeed be made precise, using the method of interlacing family, at least for the restricted invertibility problem.

Question 12.17 (Polynomial Proof for Spectral Sparsification). *Can you prove the spectral sparsification result in Theorem 10.1 by turning the heuristic argument in Section 10.1 into a precise proof (possibly using the method of interlacing family of polynomials)?*

Interlacing Family of Principle Submatrices

Ravichandran's approach is based on the family of characteristic polynomials of principal submatrices of a matrix. Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. For $1 \leq i \leq n$, let $A_{\{i\}} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the principal submatrix of A obtained by deleting the i-th row and i-th column of A. Note that the characteristic polynomials $p_{A_{\{1\}}}, \ldots, p_{A_{\{n\}}}$ of $A_{\{1\}}, \ldots, A_{\{n\}}$ have a common interlacing by Cauchy's interlacing Theorem 2.13. So, by Theorem 12.5,

$$\max_i \left\{\lambda_{\max}\left(A_{\{i\}}\right)\right\} \geq \lambda_{\max}\left(\sum_{i=1}^m p_{A_{\{i\}}}\right) \geq \min_i \left\{\lambda_{\max}\left(A_{\{i\}}\right)\right\}.$$

Ravichandran noted that there is a very nice formula for $\sum_{i=1}^m p_{A_{\{i\}}}$ and observed that it can be used to define an interlacing family of the characteristic polynomials of principal submatrices.

Theorem 12.18 (Thompson's Theorem [Tho66]). *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and let $A_{\{1\}}, \ldots, A_{\{n\}}$ be the $(n-1)\times(n-1)$ principal submatrices of A. Then*

$$\sum_{i=1}^m p_{A_{\{i\}}} = p'_A.$$

Theorem 12.19 (Interlacing Family of Principal Submatrices [Rav18]). *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. For $S \subseteq [n]$, let A_S be the principal submatrix of A obtained by deleting the rows and columns of S. For any $0 \le k \le n$, the set of characteristic polynomials $\{p_{A_S}\}_{|S|=k}$ forms an interlacing family, with the root polynomial being $p_A^{(k)}(x)$ which is the k-th derivative of $p_A(x)$.*

Proof. There are $\binom{n}{k}$ polynomials in this family. We organize them as the leaves of a tree T of depth k, where each node in T at depth i corresponds to a subset $S \subseteq [n]$ of size i and a principle matrix A_S of size $(n-i)\times(n-i)$. The root is at depth 0, and it corresponds to the empty set and the original matrix $A_\emptyset = A$. The i-th node at depth 1 corresponds to the singleton subset $\{i\}$ and the principal submatrix $A_{\{i\}}$. Inductively, given a node of T at depth i which corresponds to a subset $X \subseteq [n]$ of size i, it has $n-i$ children which correspond to the subsets $X \cup \{j\}$ for each $j \in [n] \setminus X$. The tree then has $n \times (n-1) \times \ldots \times (n-k+1) = k! \cdot \binom{n}{k}$ leaves, where each subset of size k is associated with $k!$ leaves of T (one for each permutation).

Next we define the polynomials in the nodes of T. For a leaf node, let S be the corresponding subset of size k, the polynomial is simply p_{A_S} which is the characteristic polynomial of A_S. Inductively, from the leaves to the root, the polynomial of an internal node of T is defined as the sum of the polynomials of its children.

Now we compute the polynomials in the nodes of T. The leaves at depth k are the base cases. For a node at depth $k-1$, it corresponds to a subset $X \subseteq [n]$ of size $k-1$, with the polynomials in its children being $p_{A_{X\cup\{j\}}}$ for $j \in [n] \setminus X$. By Thompson's Theorem 12.18,

$$\sum_{j\in[n]\setminus X} p_{A_{X\cup\{j\}}} = p'_{A_X}.$$

For a node at depth l which corresponds to a subset $Y \subseteq [n]$ of size l, the induction hypothesis is that the polynomials at its children are $p_{A_{Y\cup\{j\}}}^{(k-l-1)}$ for $j \in [n] \setminus Y$. Then, by Thompson's theorem, the polynomial at this node is

$$\sum_{j\in[n]\setminus Y} p_{A_{Y\cup\{j\}}}^{(k-l-1)} = \left(\sum_{j\in[n]\setminus Y} p_{A_{Y\cup\{j\}}}\right)^{(k-l-1)} = \left(p'_{A_Y}\right)^{(k-l-1)} = p_{A_Y}^{(k-l)},$$

proving the induction step. Therefore, for the root node, the polynomial is $p_A^{(k)}$ as stated.

Finally, we check that these polynomials satisfy the conditions in Definition 12.10. Property (1) is satisfied as the polynomial at a non-leaf node is the sum of the polynomials of its children, which is the same as the average polynomial up to a scalar which does not change the locations of the roots. For property (2), first we consider the case that the non-leaf node is at depth $k-1$, then the polynomials at its children have a common interlacing by Cauchy's interlacing Theorem 2.13, and thus the second property is satisfied by Theorem 12.8. Note that common interlacing is preserved by the differential operator ∂_x, using the same proof as in Problem 12.9. Therefore, for a node at depth l which corresponds to a subset $Y \subseteq [n]$ of size l, the polynomials $p_{A_{Y\cup\{j\}}}^{(k-l-1)}$ for $j \in [n]-Y$ at its children have a common interlacing, because $p_{A_{Y\cup\{j\}}}$ for $j \in [n]-Y$ have a common interlacing by Cauchy's interlacing theorem and applying the differential operator ∂_x at each of these polynomials (multiple times) preserves the common interlacing property. We conclude that the polynomials at the leaves form an interlacing family, with the root polynomial being $p^{(k)}(A)$. □

As a consequence, the method of interlacing family in Theorem 12.12 implies the following bound.

Theorem 12.20 (Ravichandran's Theorem [Rav18]). *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. There exists a subset $S \subset [n]$ with $|S| = k$ and*

$$\lambda_{\max}(A_S) \leq \lambda_{\max}\Big(p_A^{(k)}\Big).$$

Ravichandran applies Theorem 12.20 to the restricted invertibility problem in the following way. Given $v_1, \ldots, v_m \in \mathbb{R}^n$ with $\sum_{i=1}^m v_i v_i^T = I_n$, let $V \in \mathbb{R}^{n\times m}$ be the matrix with the i-th column be v_i. Consider the $m \times m$ matrix $B = I_m - V^T V$. For a subset $S \subseteq [m]$ with $|S| = k$, check that

$$\lambda_k\Big(\sum_{i\in S} v_i v_i^T\Big) = 1 - \lambda_{\max}\big(B_{[m]\setminus S}\big).$$

So the restricted invertibility problem is reduced to finding a subset X of size $m-k$ with small maximum eigenvalue $\lambda_{\max}(B_X)$. Using Theorem 12.20,

$$\max_{S:|S|=k} \lambda_k\Big(\sum_{i\in S} v_i v_i^T\Big) = 1 - \min_{X:|X|=m-k} \lambda_{\max}\big(B_X\big) \geq 1 - \lambda_{\max}\Big(p_B^{(m-k)}\Big) = 1 - \lambda_{\max}\Big(\partial_x^{m-k}(x-1)^{m-n}x^n\Big),$$

as the matrix B has eigenvalue 1 with multiplicity $m-n$ and eigenvalue 0 with multiplicity n, because $V^T V$ has the same spectrum as $VV^T = I_n$ by Fact 2.28. Therefore, once again, we have reduced the bound in the restricted invertibility problem to a pure mathematical problem about the maximum root of a well-studied polynomial.

Discussions

We end with two concluding remarks. One is that instead of looking up the known results for the roots of the specific polynomial in Exercise 12.16 and $\partial_x^{m-k}(x-1)^{m-n}x^n$ in Ravichandran's approach, we can use the results in Lemma 11.9 and Problem 11.13 from the barrier method to bound the roots of these polynomials. So, combining the method of interlacing family with the barrier method would give self-contained proofs of Theorem 12.14.

Another is that the proofs are constructive in that they give polynomial time algorithms to find such a subset. We leave this to the reader to check.

12.5 References

[BSS14] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Rev.*, 56(2):315–334, 2014. 97, 98, 99, 100, 103, 108, 110, 117

[MSS14] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Ramanujan graphs and the solution of the Kadison-Singer problem. In *Proceedings of the international Congress of Mathematicians*, volume 3, pages 363–386. 2014. 117, 145

[MSS21] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families iii: Sharper restricted invertibility estimates. *Israel Journal of Mathematics*, pages 1–28, 2021. 104, 111, 119, 122, 123, 135

[Rav18] Mohan Ravichandran. Principal submatrices, restricted invertibility, and a quantitative gauss-lucas theorem. *International Mathematics Research Notices*, 2020(15):4809–4832, 2018. 122, 124, 125, 128

[Tho66] R. C. Thompson. Principal submatrices of normal and Hermitian matrices. *Illinois Journal of Mathematics*, 10(2):296 – 308, 1966. 123

Chapter 13

Real Stable Polynomials

To use the method of interlacing family of polynomials in Chapter 12, we need to check whether a set of polynomials $p_1, \ldots, p_m$ have a common interlacing, and this is reduced to checking whether all convex combinations $q = \sum_{i=1}^m \mu_i p_i$ are real-rooted polynomials by Theorem 12.8. In this chapter, we will see some characterizations of real-rooted polynomials. The main object that we will study is the class of real-stable polynomials, a multivariate generalization of the class of real-rooted polynomials. We will use the techniques in the theory of real-stable polynomials to prove that the family in Example 12.11 is an interlacing family.

13.1 Real-Rooted Polynomials

A polynomial is real-rooted if all of its roots are real numbers. One important example of real-rooted polynomials is the characteristic polynomial of a real symmetric matrix (more generally Hermitian matrix), as the roots are the eigenvalues of the matrix and they are real numbers by Theorem 2.5.

Besides computing all the roots of a polynomial, there is a general characterization for checking whether a given polynomial is real-rooted.

Theorem 13.1 (Hermite-Sylvester). *A polynomial $p(x) = \prod_{l=1}^n (x - \lambda_l)$ is a real-rooted polynomial if and only if the $n \times n$ matrix H with $H_{ij} = \sum_{l=1}^n \lambda_l^{i+j-2}$ is a positive semidefinite matrix.*

Given a polynomial in the coefficient form $p(x) = \sum_{i=0}^n c_i x^i$, note that the entries of H (which are moments of the roots) can be computed from the coefficients efficiently by Newton's identities, and thus Hermite-Sylvester's theorem provides a polynomial time algorithm to check whether a polynomial is real-rooted. We will not use this theorem to check whether a polynomial is real-rooted, and we leave the proof as an interesting problem to solve for the reader.

Another approach to show that a polynomial $p(x)$ is real-rooted is to start with a known real-rooted polynomial $q(x)$ (e.g. the characteristic polynomial of a real symmetric matrix) and show that $p(x)$ can be obtained from $q(x)$ by some real-rootedness preserving operations.

Exercise 13.2 (Real-Rootedness Preserving Operations). *Prove that the following operations are real-rootedness preserving operations:*

1. *(Scaling:) If $p(x)$ is real-rooted, then $p(cx)$ is real-rooted for any $c \in \mathbb{R}$.*

2. *(Inversion:) If* $p(x) = \sum_{i=0}^{n} c_i \cdot x^i$ *is a degree* n *real-rooted polynomial, then so is the polynomial* $x^n \cdot p\left(\frac{1}{x}\right) = \sum_{i=0}^{n} c_{n-i} \cdot x^i$.
3. *(Differentiation:) If* $p(x)$ *is a real-rooted polynomial, then so is* $p'(x)$*, the derivative of* $p(x)$.

We will use this approach to prove that a polynomial is real-rooted, but in the more general multivariate setting which we will define in the next section.

In the remainder of this section, we collect some nice properties of real-rooted polynomials. They will not be used for the method of interlacing family of polynomials, and we refer the reader to the course notes of Oveis Gharan [Ove20] for proofs.

Gauss-Lucas Theorem

The following theorem is a generalization of item (3) in Exercise 13.2. The proof is by considering p'/p and writing a root of p' but not p as a convex combination of the roots of p.

Problem 13.3 (Gauss-Lucas Theorem)**.** *If* $p \in \mathbb{C}[x]$ *is a non-constant polynomial with complex coefficients, then all roots of* p' *are in the convex hull of the set of roots of* p.

Ravichandran used the techniques developed for the restricted invertibility problem in Section 12.4 to prove the following quantitative generalization of the Gauss-Lucas theorem, which bounds the area of the convex hull after many differentiations.

Theorem 13.4 (Quantitative Gauss-Lucas Theorem [Rav18])**.** *Let* $p \in \mathbb{C}[x]$ *be a degree* n *polynomial with complex coefficients. Then, for any* $c \geq 1/2$,

$$\left|\mathcal{K}\left(p^{\lceil cn \rceil}\right)\right| \leq 4(c - c^2)\left|\mathcal{K}(p)\right|,$$

where $\mathcal{K}(p)$ *denotes the convex hull of the roots of* p *and* $|S|$ *denotes the area of the convex set* S *in the plane.*

Note that there are examples where taking the $(\frac{n}{2} - 1)$-th derivative does not decrease the area yet. One such example is $(x+1)^{n/2}(x-1)^{n/2}$.

Generating Polynomials

Given a probability distribution μ over $[n]$, we define its generating polynomial as

$$p_\mu(x) = \sum_{i=1}^{n} \mu_i \cdot x^i.$$

The following is an interesting characterization of when such a generating polynomial is real-rooted.

Proposition 13.5 (Real-Rooted Generating Polynomials)**.** *The generating polynomial* $p_\mu(x)$ *is real-rooted if and only if* μ *is the distribution of a sum of independent Bernoulli random variables.*

We will study in a later chapter about probability distributions with real-stable generating polynomials, and we may discuss the proof of Proposition 13.5 there.

One consequence of Proposition 13.5 is that we can use Chernoff bounds to bound the coefficient $a_i = \Pr[X = i]$ with i far away from the mean $\mathbb{E}[X] = \sum_{i=1}^{n} i \cdot a_i$. From this connection, we expect to see a Bell curve when we plot the numbers $a_1, \ldots, a_n$ of a real-rooted polynomial with non-negative coefficients. This intuition can be made precise by the notion in the next subsection.

Log-Concavity

The following is an analog of a log-concave function for a sequence.

Definition 13.6 (Log-Concave Sequence)**.** *A sequence $a_0, \ldots, a_n$ of non-negative numbers is said to be log-concave if for all $0 < i < n$,*

$$a_{i-1} \cdot a_{i+1} \leq a_i^2 \quad \iff \quad \frac{1}{2}\big(\log(a_{i-1}) + \log(a_{i+1})\big) \leq \log(a_i).$$

A sequence $a_0, \ldots, a_n$ of non-negative numbers is said to be ultra log-concave if for all $0 < i < n$,

$$\frac{a_{i-1}}{\binom{n}{i-1}} \cdot \frac{a_{i+1}}{\binom{n}{i+1}} \leq \left(\frac{a_i}{\binom{n}{i}}\right)^2$$

We can use the operations in Exercise 13.2 to reduce a degree n real-rooted polynomial to a quadratic real-rooted polynomial involving only a_{i-1}, a_i, a_{i+1}, and then consider the discriminant of the resulting quadratic real-rooted polynomial to prove the following result.

Problem 13.7 (Newton Inequalities)**.** *For any real-rooted polynomial $p(x) = \sum_{i=0}^{n} a_i \cdot x^i$ with non-negative coefficients, the sequence $a_0, \ldots, a_n$ is ultra log-concave.*

In the third part of the course, we will study log-concave polynomials and see that some sequences from combinatorial problems are log-concave (such as the number of matchings of size i).

13.2 Real Stable Polynomials

The class of real-stable polynomials is a multivariate generalization of real-rooted polynomials.

Definition 13.8 ($\mathcal{H}$-Stable Polynomials)**.** *A multivariate polynomial $p \in \mathbb{C}[x_1, \ldots, x_n]$ is $\mathcal{H}$-stable if $p(x_1, \ldots, x_n) \neq 0$ whenever $(x_1, \ldots, x_n) \in \mathcal{H}^n$ where $\mathcal{H} = \{y \in \mathbb{C} \mid \Im(y) > 0\}$ is the upper-half of the complex plane.*

In the third part of the course, we may see some other stable polynomials where the root-free region is differently specified (e.g. sector-stable polynomials).

Definition 13.9 (Real Stable Polynomials)**.** *A multivariate polynomial p is called real stable if p is $\mathcal{H}$-stable and all coefficients of p are real numbers.*

Some simple examples of real stable polynomials are $p(x_1, \ldots, x_n) = x_1 x_2 \cdots x_n$ and $p(x_1, \ldots, x_n) = a_1 x_1 + \ldots + a_n x_n$ where $a_i > 0$ for $1 \leq i \leq n$. Some simple non-examples of real stable polynomials are $p(x_1, x_2) = x_1 - x_2$ and $p(x_1, x_2, x_3, x_4) = x_1 x_2 - x_3 x_4$.

Note that it is a generalization of real-rooted univariate polynomials, using that complex roots of a polynomial with real coefficients come in conjugate pairs.

Exercise 13.10 (Univariate Real-Stable Polynomials)**.** *A univariate polynomial $p \in \mathbb{R}[x]$ is real stable if and only if it is real-rooted.*

Sometimes it is more convenient to check whether a multivariate polynomial is real-stable by checking whether certain derived univariate polynomials are real-rooted.

Exercise 13.11 (Univariate Restrictions). *A polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ is real stable if and only if for any $b \in \mathbb{R}^n_+$ and $a \in \mathbb{R}^n$, the univariate polynomial $p(a + yb)$ in y is not identically equal to zero and is real-rooted.*

Using Exercise 13.11, one can draw some pictures to see that the polynomial $1 - xy$ is real-stable while the polynomial $1 + xy$ is not real-stable.

In this course, the source of all real-stable polynomials comes from determinants.

Proposition 13.12 (Source of Real-Stable Polynomials). *If $A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$ are positive semidefinite matrices, then $p(x_0, x_1, \ldots, x_m) := \det(x_0 I + \sum_{i=1}^m x_i A_i)$ is a real stable polynomial.*

Proof. The plan is to show that if $\Im(x_i) > 0$ for all $0 \le i \le m$, then the matrix $x_0 I + \sum_{i=1}^m x_i A_i$ is of full rank, and hence $\det(x_0 I + \sum_{i=1}^m x_i A_i) \neq 0$, implying real stability.

Let $v \in \mathbb{C}^n$, and write $v = c + \imath d$ where $c \in \mathbb{R}^n$ is the real part and $d \in \mathbb{R}^n$ is the imaginary part of v. Let $X = x_0 I + \sum_{i=1}^m x_i A_i$, and write X as $\Re(X) + \imath \Im(X)$ where $\Re(X)$ is the real part and $\Im(X)$ is the imaginary part of X. Note that if $\Im(x_i) > 0$ for all $0 \le i \le m$, then $\Im(X) \succ 0$, as $A_i \succcurlyeq 0$ for $0 \le i \le m$ and $I \succ 0$.

We claim that $Xv = 0$ only if $v = 0$, and hence X is of full rank. To prove this, we show that $v^* X v = (c - \imath d)^T (\Re(X) + \imath \Im(X))(c + \imath d) = 0$ only if $c = d = 0$. Note that the imaginary part of $v^* X v$ is

$$\Im\Big[(c - \imath d)^T (\Re(X) + \imath \Im(X))(c + \imath d)\Big] = c^T \Im(X) c + d^T \Im(X) d,$$

and this is equal to zero only if $c = d = 0$, because $\Im(X) \succ 0$ when $\Im(x_i) > 0$ for $0 \le i \le m$. □

One could also prove Proposition 13.12 using the univariate restrictions in Exercise 13.11; see Oveis Gharan's notes [Ove20].

Later, we will start from the multivariate real-stable polynomials from Proposition 13.12, and then apply the real-stability preserving operations in the next section to prove that a univariate polynomial is real-stable, and hence real-rooted by Exercise 13.10.

13.3 Real Stability Preserving Operations

There are several real-stability preserving operations, with some deep characterizations. We just present the proofs of two operations that we need in this course, and state others without proofs.

The following operation will be useful in reducing the number of variables in the multivariate polynomial.

Proposition 13.13 (Specialization). *Let $p(x_1, \ldots, x_m)$ be a non-zero real-stable polynomial. For any $c \in \mathbb{R}$, $p(c, x_2, \ldots, x_m)$ is a real-stable polynomial.*

Proof. It is clear that $p(c, x_2, \ldots, x_m)$ has real coefficients as p has real coefficients and $c \in \mathbb{R}$. For stability, consider the sequence of polynomials $p_k = p(c + \imath 2^{-k}, x_2, \ldots, x_m)$ for $k \ge 1$. Note that each p_k is a $\mathcal{H}$-stable polynomial (but may have complex coefficients) as p is $\mathcal{H}$-stable. The sequence $\{p_k\}_{k \ge 1}$ is converging uniformly to the polynomial $p(c, x_2, \ldots, x_m)$.

Suppose, by contradiction, that $p(c, x_2, \ldots, x_m)$ has a root $z_2, \ldots, z_m$ with $\Im(z_i) > 0$ for $2 \le i \le m$. By Hurwitz's Theorem 13.14, for any small enough $\epsilon > 0$ and for every large enough k (depending

on ϵ), the polynomial p_k also has a root $y_2, \ldots, y_m$ with $|y_i - z_i| < \epsilon$ for $2 \le i \le m$. By choosing ϵ small enough, we still have $\Im(y_i) > 0$ for $2 \le i \le m$, but this means that p_k has a root with all imaginary parts positive, contradicting the $\mathcal{H}$-stability of p_k. □

Hurwitz's theorem is from complex analysis. The following statement is from Wikipedia.

Theorem 13.14 (Hurwitz's Theorem). *Let $\{f_k\}_{k\ge1}$ be a sequence of holomorphic functions on a connected open set G that converge uniformly on compact subsets of G to a holomorphic function f which is not constantly zero on G. If f has a zero of order l at z_0, then for every small enough $\rho > 0$ and for sufficiently large $k \in \mathbb{N}$ (depending on ρ), f_k has precisely l zeros in the disk defined by $|z - z_0| < \rho$ including multiplicity. Furthermore, these zeroes converge to z_0 as $k \to \infty$.*

The other operation that we need is the differential operator that we have seen a couple of times already. The following proposition is for univariate polynomials.

Proposition 13.15 (Partial Differentiation). *If $p \in \mathbb{C}[x]$ is $\mathcal{H}$-stable, then $p + s \cdot p'$ is $\mathcal{H}$-stable for any $s \in \mathbb{R}$.*

Proof. Since $p(x)$ is stable, it can be written as $c\prod_{j=1}^{n}(x - w_j)$ with $\Im(w_j) \le 0$ for $1 \le j \le n$. Then

$$p(x) + s \cdot p'(x) = p(x)\left(1 + \sum_{j=1}^{n} \frac{s}{x - w_j}\right).$$

For z with $\Im(z) > 0$, $p(z) \neq 0$ as p is $\mathcal{H}$-stable. Furthermore, since $\Im(z) > 0$ and $\Im(w_j) \le 0$ for $1 \le j \le n$, it follows that $\Im\big(\frac{1}{z-w_j}\big) < 0$ for $1 \le j \le n$, and thus $1 + \sum_{j=1}^{n} \frac{s}{x-w_j} \neq 0$. This proves that $g(z) + s \cdot g'(z) \neq 0$ if $\Im(z) > 0$, establishing $\mathcal{H}$-stability. □

This result can be generalized to multivariate polynomials easily by univariate restriction.

Corollary 13.16 (Partial Differentiation). *If $p \in \mathbb{R}[x_1, \ldots, x_m]$ is real-stable, then $(1 + s \cdot \partial_{x_1})p$ is real-stable for any $s \in \mathbb{R}$.*

Proof. It is clear that $(1 + s \cdot \partial_{x_1})p$ has real coefficients if p has. For any $y_2, \ldots, y_m$ with $\Im(y_i) > 0$ for $2 \le i \le m$, the polynomial $q(x_1) := p(x_1, y_2, \ldots, y_m)$ is stable by definition. Proposition 13.15 proves that $(1 + s \cdot \partial_{x_1})q(x_1)$ is also stable. This implies that $(1 + s \cdot \partial_{x_1})p$ has no roots in which all of the variables have positive imaginary part, proving stability. □

The following are some other operations that preserve real-stability, whose proofs are elementary.

Exercise 13.17 (Real-Stability Preserving Operations). *Let $p(x_1, x_2, \ldots, x_m)$ and $q(x_1, \ldots, x_m)$ be real-stable polynomials. Then*

1. *(Product:) $p \cdot q$ is real stable.*
2. *(Symmetrization:) $p(x_1, x_1, x_3, \ldots, x_m)$ is real stable.*
3. *(External Field:) $p(c_1x_1, c_2x_2, \ldots, c_mx_m)$ is real stable for any $c_1, \ldots, c_m \ge 0$.*
4. *(Inversion:) $p\big(-\frac{1}{x_1}, x_2, \ldots, x_m\big) \cdot x_1^{d_1}$ is real stable where d_1 is the degree of x_1 in p.*

5. *(Differentiation:)* $\partial_{x_1} p$ *is real stable.*

Borcea and Brändén characterized a class of differential operators that preserve real stability.

Theorem 13.18 (Borcea-Brändén Theorem)**.** *For vectors* $\alpha, \beta \in \mathbb{N}^m$*, let* $x^\alpha = x^{\alpha(1)} \cdots x^{\alpha(n)}$ *and* $\partial^\beta = \partial_{x_1}^{\beta(1)} \cdots \partial_{x_m}^{\beta(m)}$ *and let* $D = \sum_{\alpha,\beta \in \mathbb{N}^m} c_{\alpha,\beta} \cdot x^\alpha \cdot \partial^\beta$ *be a differential operator with* $c_{\alpha,\beta} \in \mathbb{R}$ *for all* $\alpha, \beta \in \mathbb{N}^m$*. Then* D *is a stability preserving operator (i.e. it maps any real-stable polynomial to a real-stable polynomial) if and only if the polynomial* $\sum_{\alpha,\beta \in \mathbb{N}^m} c_{\alpha,\beta} \cdot x^\alpha \cdot (-w)^\beta \in \mathbb{R}[x_1, \ldots, x_m, w_1, \ldots, w_m]$ *on* $2m$ *variables is real-stable.*

For examples, $1 - \partial x_1 \partial x_2$ is stability preserving because $1 - (-w_1)(-w_2) = 1 - w_1 w_2$ is a real stable polynomial, and similarly $1 + x_1 \partial x_2$ is stability preserving. For non-examples, $1 + \partial_{x_1} \partial_{x_2}$ is not stability preserving as $1 + w_1 w_2$ is not a stable polynomial, and similarly $1 - \partial_{x_1} \partial_{x_2} \partial_{x_3}$ is not stability preserving.

Problem 13.19 (Real Stability Preserving Operators)**.** *Use Theorem 13.18, or otherwise (both are possible), to prove the following results.*

1. *For any* $1 \leq k \leq n$*, the* k*-th elementary symmetric polynomial* $\sum_{S \subseteq \binom{[n]}{k}} x^S$ *is real stable.*
2. *Let MAP be the operator that only retains the multiaffine monomials of a given polynomial, e.g.* $MAP(1 + x + 3x^3 y + 2xy) = 1 + x + 2xy$*. Prove that MAP is stability preserving.*

See [Wag11] for a survey on real-stable polynomials, with a proof of Theorem 13.18.

13.4 Multilinear Formula, Mixed Characteristic Polynomials, and Interlacing Family

In this section, we use the tools from real stable polynomials to prove that a generalization of the family in Example 12.11 is an interlacing family, which will be a key component in constructing bipartite Ramanujan graphs and resolving the Kadison-Singer problem in the next two chapters.

Mixed Characteristic Polynomial and Multilinear Formula

We consider the setting where each A_i is a random symmetric rank-one matrix with finite support (e.g. A_i is aa^T with probability 0.6, bb^T with probability 0.3, cc^T with probability 0.1), and $A = \sum_{i=1}^m A_i$ is a sum of independent rank-one matrices. We are interested in proving that the set of all possible characteristic polynomials $\det(xI - \sum_{i=1}^m A_i)$ forms an interlacing family. The following identity of the expected characteristic polynomial is at the heart of the approach by Marcus, Spielman, and Srivastava.

Theorem 13.20 (Multilinear Formula)**.** *If* $A_1, A_2, \ldots, A_m$ *are independent random symmetric rank-one matrices, then*

$$\mathbb{E}_{A_1,\ldots,A_m}\left[\det\left(\lambda I - \sum_{i=1}^m A_i\right)\right] = \prod_{i=1}^m (1 - \partial_{x_i}) \det\left(\lambda I + \sum_{i=1}^m x_i \cdot \mathbb{E}\left[A_i\right]\right)\Big|_{x_1 = x_2 = \cdots = x_m = 0}.$$

The right hand side of the multilinear formula is called the mixed characteristic polynomial of the expected matrices $\mathbb{E}[A_1], \ldots, \mathbb{E}[A_m]$, which are not of rank one in general.

Definition 13.21 (Mixed Characteristic Polynomial)**.** *The mixed characteristic polynomial of $n \times n$ matrices $B_1, \ldots, B_m$ (not necessarily rank-one) is defined as*

$$\mu[B_1, \ldots, B_m](\lambda) = \prod_{i=1}^{m} (1 - \partial_{x_i}) \det\left(\lambda I + \sum_{i=1}^{m} x_i \cdot B_i\right)\Big|_{x_1 = x_2 = \cdots = x_m = 0}$$

There are different proofs of Theorem 13.20. We first present the proof from [MSS15a] (suggested by James Lee), and then discuss a proof by Tao [Tao13] which shows more clearly why it is a multilinear formula. The original proof by Marcus, Spielman, and Srivastava used the Cauchy-Binet formula in Fact 2.30.

Inductive Proof: The base case is similar to the calculations in Section 10.1 and in Exercise 11.8, with the only difference that $\mathbb{E}[A_i]$ is not necessarily a scaled identity matrix.

Lemma 13.22 (Expected Rank-One Update)**.** *For any square matrix M and a random vector v,*

$$\mathbb{E}_v\big[\det(M - vv^T)\big] = (1 - \partial_x) \det\big(M + x \cdot \mathbb{E}\left[vv^T\right]\big)\Big|_{x=0}.$$

Proof. First, we assume M is invertible. By the matrix determinantal formula in Fact 2.29,

$$\det(M - vv^T) = \det(M) \cdot \big(1 - v^T M^{-1} v\big) = \det(M)\big(1 - \operatorname{Tr}\left(M^{-1} vv^T\right)\big).$$

Taking expectation on both sides,

$$\mathbb{E}_v\big[\det(M - vv^T)\big] = \det(M) - \det(M) \operatorname{Tr}\left(M^{-1}\mathbb{E}\left[vv^T\right]\right).$$

On the other hand, by the Jacobi's formula in Fact 2.39,

$$\partial_x \det\big(M + x \cdot \mathbb{E}\left[vv^T\right]\big)\big|_{x=0} = \det\left(M\right) \operatorname{Tr}\left(M^{-1}\mathbb{E}\left[vv^T\right]\right),$$

and so the lemma follows when M is invertible. When M is not invertible, we can choose a sequence of invertible matrices that approach M. Since the lemma holds for each matrix in the sequence and the two sides are polynomials in the entries of the matrix, a continuity argument implies that the lemma also holds for M as well. □

Then Theorem 13.20 can be proved by applying Lemma 13.22 repeatedly.

Exercise 13.23 (Inductive Proof of Multilinear Formula)**.** *Complete the proof of Theorem 13.20 by using Lemma 13.22 inductively and the assumption that $A_1, \ldots, A_m$ are independent random variables.*

Multilinear Proof: The proof presented by Tao [Tao13] also starts from the matrix determinantal formula, which shows that $\det(\lambda I - \sum_{i=1}^m A_i)$ is multilinear in terms of A_i, when each A_i is a rank one matrix. Then we can understand that the RHS of Theorem 13.20 is just a Taylor expansion of the LHS.

Lemma 13.24 (Taylor Expansion of Multilinear Polynomial)**.** *Let $p(x_1, \ldots, x_m)$ be a multilinear polynomial in $x_1, \ldots, x_m$. Then*

$$p(x_1, \ldots, x_m) = \prod_{i=1}^m (1 + x_i \partial_{y_i}) p(y_1, \ldots, y_m)\Big|_{y_1 = \ldots = y_m = 0}.$$

Proof. As p is a multilinear polynomial, it can be written as $p(x_1, \ldots, x_m) = \sum_{S \subseteq [m]} c_S \prod_{i \in S} x_i$, where c_S is the coefficient of the monomial $\prod_{i \in S} x_i$. Note that $c_S = \prod_{i \in S} \partial_{y_i} p(y_1, \ldots, y_m)\big|_{y_1 = \ldots = y_m = 0}$, as differentiation and substitution kill all the terms except c_S. Therefore,

$$p(x_1, \ldots, x_m) = \sum_{S \subseteq [m]} \Big(\prod_{i \in S} x_i\Big)\Big(\prod_{i \in S} \partial_{y_i} p(y_1, \ldots, y_m)\big|_{y=0}\Big) = \prod_{i=1}^m \Big(1 + x_i \partial_{y_i}\Big) p(y_1, \ldots, y_m)\Big|_{y=0}.$$

□

Putting in $p(x_1, \ldots, x_m) = \det(B + x_1 A_1 + \ldots + x_m A_m)$ in Lemma 13.24 gives the following corollary.

Corollary 13.25 (Determinant of Sum of Rank One Matrices)**.** *If $A_1, \ldots, A_m$ are symmetric rank-one matrices, then*

$$\det(B + x_1 A_1 + \ldots + x_m A_m) = \prod_{i=1}^m (1 + x_i \partial_{y_i}) \det(B + y_1 A_1 + \ldots + y_m A_m)\Big|_{y_1 = \ldots = y_m = 0}.$$

To prove Theorem 13.20, we set $B = \lambda I$ and $x_1 = \ldots = x_m = -1$ in Corollary 13.25. Then, we take the expectation on both sides using the sum of monomials form, and move the expectation inside the summation by linearity of expectation, and then move the expectation inside the products by independence of the random variables $A_1, \ldots, A_m$ to obtain the following result.

Exercise 13.26 (Expansion Proof of Multilinear Formula)**.** *Complete the proof of Theorem 13.20 by proving that*

$$\mathbb{E}_{A_1, \ldots, A_m}\left[\det\left(\lambda I - \sum_{i=1}^m A_i\right)\right] = \mu[\mathbb{E}[A_1], \ldots, \mathbb{E}[A_m]](\lambda),$$

the mixed characteristic polynomial of $\mathbb{E}[A_1], \ldots, \mathbb{E}[A_m]$ in Definition 13.21.

Interlacing Family of Independent Rank-One Matrices

With the multilinear formula in Theorem 13.20, we are now ready to prove that the set of all possible characteristic polynomials $\{\det(\lambda I - \sum_{i=1}^m A_i)\}$ form an interlacing family. The following lemma will be useful in showing that all conditional expectation polynomials are real-rooted.

Proposition 13.27 (Expected Characteristic Polynomial is Real-Rooted)**.** *The expected characteristic polynomial $\mathbb{E}_{A_1, \ldots, A_m}\left[\det\left(\lambda I - \sum_{i=1}^m A_i\right)\right]$ is real-rooted for any independent random symmetric rank-one matrices $A_1, \ldots, A_m$.*

Proof. We start from the RHS of the multilinear formula in Theorem 13.20. Since each A_i is a random symmetric rank-one matrix, the expected matrix $\mathbb{E}[A_i] = \sum_j p_j v_j v_j^T \succcurlyeq 0$ is a positive semidefinite matrix. So, by Proposition 13.12, the multivariate polynomial $\det(\lambda I + \sum_{i=1}^m x_i \cdot \mathbb{E}[A_i])$ is a real-stable polynomial. By the results in stability preserving operations in Corollary 13.16 and Proposition 13.13, applying the differential operator $1 - \partial_{x_i}$ and substituting real numbers preserve stability. Therefore, the LHS of the multilinear formula is a real-stable univariate polynomial in λ, and thus real-rooted by Exercise 13.10. □

The following interlacing family plays a major role in the sequence of papers by Marcus, Spielman and Srivastava [MSS15b, MSS15a, MSS21].

Theorem 13.28 (Interlacing Family of Independent Rank-One Matrices)**.** *Let $A_1, A_2, \ldots, A_m$ be random symmetric rank-one matrices, where each A_i has l_i possibilities $v_{i,1}v_{i,1}^T, \ldots, v_{i,l_i}v_{i,l_i}^T$. The set of all $\prod_{i=1}^m l_i$ polynomials in $\{\det(\lambda I - \sum_{i=1}^m v_{i,j_i}v_{i,j_i}^T)\}$ form an interlacing family, where each $j_i \in \{1, \ldots, l_i\}$ for $1 \le i \le m$. Furthermore, the root polynomial of the interlacing family can be $\mathbb{E}_{A_1,\ldots,A_m}\left[\det\left(\lambda I - \sum_{i=1}^m A_i\right)\right]$ for any independent distributions on $A_1, \ldots, A_m$.*

Proof. The tree has depth m, with the root at depth 0. At depth $0 \le i \le m-1$, each node has l_{i+1} children. Each leaf of the tree is labeled by a sequence $(j_1, j_2, \ldots, j_m)$, representing a path from the root to the tree, where $j_i \in [l_i]$ represents the j_i-th child of an internal node in the $(i-1)$-th level. The polynomial in the leaf node corresponding to $(j_1, j_2, \ldots, j_m)$ is $\det(\lambda I - \sum_{i=1}^m v_{i,j_i}v_{i,j_i}^T)$, a choice $v_{j_i}v_{j_i}^T$ for each A_i for $1 \le i \le m$.

Given the independent distributions on $A_1, \ldots, A_m$, the polynomial in an internal node $(j_1, j_2, \ldots, j_k)$ at depth k is defined as $\mathbb{E}_{A_{k+1},\ldots,A_m}\left[\det\left(\lambda I - \sum_{i=1}^k v_{i,j_i}v_{i,j_i}^T - \sum_{i=k+1}^m A_i\right)\right]$, the conditional expectation polynomial where $A_i = v_{j_i}v_{j_i}^T$ is fixed for $1 \le i \le k$. The root polynomial is then $\mathbb{E}_{A_1,\ldots,A_m}\left[\det\left(\lambda I - \sum_{i=1}^m A_i\right)\right]$.

We need to check that the two conditions of an interlacing family in Definition 12.10 are satisfied. The first condition is satisfied by construction, that the polynomial in each non-leaf node at depth k is a convex combination of the polynomials in its children, where the convex combination is based on the given probability distribution of A_k, which is independent of other random variables.

For the second condition, we need to prove that the polynomials in the children of a non-leaf node have a common interlacing. By Theorem 12.8, it suffices to prove that all convex combinations of the polynomials in the children of a non-leaf node are real-rooted. Consider an internal node $(j_1, \ldots, j_k)$ at depth k, with l_k children $(j_1, \ldots, j_k, 1), \ldots, (j_1, \ldots, j_k, l_k)$. Given any convex combination $\mu_1, \ldots, \mu_{l_k}$ with $\mu_a \ge 0$ for $1 \le a \le l_k$ and $\sum_{a=1}^{l_k} \mu_a = 1$, we need to prove that

$$\sum_{a=1}^{l_k} \mu_a \cdot \mathbb{E}_{A_{k+2},\ldots,A_m}\left[\det\left(\lambda I - \sum_{i=1}^{k} v_{i,j_i}v_{i,j_i}^T - v_{k+1,a}v_{k+1,a}^T - \sum_{i=k+2}^{m} A_i\right)\right]$$

is real-rooted. Observe that this is just the expected characteristic polynomial $\mathbb{E}_{B_1,\ldots,B_m}\det(\lambda I - \sum_{i=1}^m B_i)$ for a related set of independent random symmetric rank-one matrices, where $B_1, \ldots, B_k$ are just the (deterministic) random variables with $B_i = v_{i,j_i}v_{i,j_i}^T$ with probability one, B_{k+1} is the random variable with $B_{k+1} = v_{k+1,a}v_{k+1,a}^T$ with probability μ_a for $1 \le a \le l_{k+1}$, and $B_{k+2}, \ldots, B_m$ are just the same as the random variables $A_{k+2}, \ldots, A_m$. By Proposition 13.27, any such convex combination is real-rooted, and hence the children have a common interlacing by Theorem 12.8. We conclude that the polynomials in the leaves form an interlacing family. □

Note that this generalizes the family in Example 12.11, and thus completes the proof for the restricted invertibility result in Theorem 12.14.

We will use this interlacing family for constructing bipartite Ramanujan graphs and resolving the Kadison-Singer problem in the next two chapters.

13.5 References

[MSS15a] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison-singer problem. *Annals of Mathematics*, pages 327–350, 2015. 133, 135, 145, 146, 148, 150

[MSS15b] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. *Annals of Mathematics*, 182:307–325, 2015. 135, 137, 138, 139

[MSS21] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families iii: Sharper restricted invertibility estimates. *Israel Journal of Mathematics*, pages 1–28, 2021. 104, 111, 119, 122, 123, 135

[Ove20] Shayan Oveis Gharan. Course notes of "Polynomial paradigm in algorithm design", lecture 1 and 2. 2020. https://homes.cs.washington.edu/~shayan/courses/polynomials/. 128, 130, 153, 158, 171, 173, 175, 179, 187, 211

[Rav18] Mohan Ravichandran. Principal submatrices, restricted invertibility, and a quantitative Gauss-Lucas theorem. *International Mathematics Research Notices*, 2020(15):4809–4832, 2018. 122, 124, 125, 128

[Tao13] Terence Tao. Real stable polynomials and the Kadison-Singer problem. *A blogpost of What's New*, 2013. https://terrytao.wordpress.com/2013/11/04/real-stable-polynomials-and-the-kadison-singer-problem/. 133, 134, 150

[Wag11] David Wagner. Multivariate stable polynomials: theory and applications. *Bulletin of the American Mathematical Society*, 48(1):53–84, 2011. 132

Chapter 14

Bipartite Ramanujan Graphs

We will see how Marcus, Spielman, and Srivastava [MSS15] used the method of interlacing family of polynomials to prove the existence of bipartite Ramanujan graphs, using the 2-lift construction proposed by Bilu and Linial [BL06].

The expected characteristic polynomials in this problem are exactly the matching polynomials of graphs, and we will see some classical results of these polynomials.

14.1 Combinatorial Constructions of Ramanujan Graphs

Recall the definition of Ramanujan graphs from Chapter 7.

Definition 14.1 (Ramanujan Graphs). *Let $G = ([n], E)$ be a d-regular graph and $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. We say G is a Ramanujan graph if* $\max\{\alpha_2, |\alpha_n|\} \leq 2\sqrt{d-1}$.

We are interested in constructing an infinitely family of d-regular graphs that are all Ramanujan. This is best possible, as Alon-Boppana Theorem 7.10 proved that for any $\epsilon > 0$, every large engouh d-regular graph has $\max\{\alpha_2, |\alpha_n|\} \geq 2\sqrt{d-1} - \epsilon$. There is a meaning of the value $2\sqrt{d-1}$, which is the bound on the absolute value of the eigenvalues of the infinite d-regular tree, the best possible d-regular expander graphs in the combinatorial sense.

There are known constructions of Ramanujan graphs of constant degree from Cayley graphs. All known graphs are $(q+1)$-regular where q is a prime power. The analyses of these constructions use deep mathematical results and in particular some by Ramanujan (and hence the name). They are explicit in that the neighbors of a vertex can be computed in $O(\log n)$ time. See the survey by Hoory, Linial, and Wigderson [HLW06] for more details.

2-Lifts

It is of interest to find combinatorial constructions of Ramanujan graphs. Bilu-Linial [BL06] proposed a method to construct Ramanujan graphs using 2-lifts.

Definition 14.2 (2-Lift). *Let $G = ([n], E)$ be an undirected graph. A signing of the edges of G is a function $s : E(G) \to \{-1, +1\}$. The 2-lift $\hat{G}_s = (\hat{V}, \hat{E})$ of G associated with a signing s is defined as follows. The vertex set $\hat{V}$ of $\hat{G}_s$ is $\hat{V} = \{1, \ldots, n, 1', \ldots, n'\}$, where each vertex $i \in V(G)$ has two*

copies i and i' in $\hat{G}_s$. For each edge $ij \in E(G)$, if $s(ij) = 1$, then the edges ij and $i'j'$ are in $\hat{E}$; otherwise, if $s(ij) = -1$, then the edges ij' and $i'j$ are in $\hat{E}$.

Bilu and Linial conjectured that if G is Ramanujan, then there is a 2-lift of G that is also Ramanujan. Note that if G is d-regular, then any 2-lift of G is also d-regular with the number of vertices doubled. So, if the conjecture is true, then it implies the existence of an infinite family of d-regular Ramanujan graphs for any degree d. Just start with the complete graph on $d+1$ vertices, which is Ramanujan, and keep doing a good 2-lift to double the graph size. Bilu and Linial used probabilistic method (Lovász local lemma) and the techniques in proving the converse of expander mixing lemma in Theorem 7.5 to prove that there exists a 2-lift with $\max\{\alpha_2, |\alpha_n|\} \lesssim \sqrt{d \log^3 d}$.

Spectrum of 2-Lift

There is a nice formulation to analyze the spectrum of a 2-lift of a graph.

Definition 14.3 (Signed Matrix)**.** *Given a graph $G = ([n], E)$ and a signing $s : E(G) \to \{-1, +1\}$ of the edges, the signed matrix $A_s \in \mathbb{R}^{n \times n}$ is defined as follows. If $ij \in E$, then $(A_s)_{ij} = (A_s)_{ji} = s(ij)$, otherwise $(A_s)_{ij} = 0$.*

The proof of the following statement is left as a homework problem.

Problem 14.4 (Spectrum of 2-Lift)**.** *Given a graph $G = ([n], E)$ and a signing $s : E(G) \to \{-1, +1\}$ of the edges, the spectrum of the adjacency matrix $A(\hat{G}_s)$ of the 2-lift $\hat{G}_s$ is equal to the disjoint union of the spectrum of the adjacency matrix $A(G)$ of G (called the old eigenvalues) and the spectrum of the signed matrix A_s (called the new eigenvalues). That is, the multiplicity of an eigenvalue α of $A(\hat{G}_s)$ is equal to the sum of the multiplicity of α of $A(G)$ and the multiplicity of α of A_s.*

With this statement, to prove that there is a Ramanujan 2-lift of a d-regular Ramanujan graph $G = (V, E)$, it is equivalent to proving that there is a signing $s : E(G) \to \{-1, +1\}$ so that the maximum absolute eigenvalue of the signed matrix A_s is at most $2\sqrt{d-1}$. Bilu and Linial made the following stronger conjecture which does not assume that G is a Ramanujan graph.

Conjecture 14.5 (Bili-Linial [BL06])**.** *For any d-regular graph $G = (V, E)$, there is a signing $s : E(G) \to \{-1, +1\}$ so that all eigenvalues of A_s have absolute value at most $2\sqrt{d-1}$.*

14.2 Bipartite Ramanujan Graphs from Interlacing Family

Marcus, Spielman, Srivastava [MSS15] proved Conjecture 14.5 for bipartite graphs.

Theorem 14.6 (Bili-Linial Conjecture for Bipartite Graphs [MSS15])**.** *Any d-regular bipartite graph $G = (V, E)$ has a signing $s : E(G) \to \{-1, +1\}$ so that the maximum eigenvalue of A_s is at most $2\sqrt{d-1}$.*

Note that for a bipartite graph, bounding the maximum eigenvalue is enough because the spectrum is symmetric (see Lemma 3.4), even for the signed matrix. This is the reason that Theorem 14.6 only holds for bipartite graph, because the new probabilistic method using interlacing polynomials can only bound the maximum eigenvalue (or one eigenvalue), but not the maximum eigenvalue and the minimum eigenvalue at the same time.

As the spectrum of a bipartite graph is symmetric, any d-regular bipartite graph has the maximum eigenvalue equal to d and the minimum eigenvalue equal to $-d$, which are called the trivial eigenvalues. A bipartite graph is Ramanujan if all its non-trivial eigenvalues are at most $2\sqrt{d-1}$.

Definition 14.7 (Bipartite Ramanujan Graphs)**.** *Let $G = ([n], E)$ be a d-regular graph and $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. We say G is a bipartite Ramanujan graph if $\max\{\alpha_2, |\alpha_{n-1}|\} \leq 2\sqrt{d-1}$.*

A corollary of Theorem 14.6 and Problem 14.4 is that any bipartite Ramanujan graph G has a 2-lift that is Ramanujan. Note that a 2-lift of a bipartite graph is bipartite. So, starting from a complete bipartite graph with $2d$ vertices, which is Ramanujan (see Example 3.3), repeatedly applying a good 2-lift proves the following theorem.

Theorem 14.8 (Bipartite Ramanujan Graphs of Every Degree [MSS15])**.** *For every d, there is an infinite family of d-regular bipartite Ramanujan graphs.*

Probabilistic Method

Theorem 14.6 is proved by the method of interlacing family of polynomials developed in Chapter 12 and Chapter 13. Given a d-regular graph $G = (V, E)$ with $m := |E|$ edges, there are totally 2^m different signed matrices of G and we consider the uniform distribution on these 2^m signed matrices. The plan is to prove that there exists a signing $s : E(G) \to \{-1, +1\}$ such that

$$\lambda_{\max}\big(\det(xI - A_s)\big) \leq \lambda_{\max}\Big(\mathbb{E}_{s\in\{\pm1\}^m}\big[\det(xI - A_s)\big]\Big) \leq 2\sqrt{d-1}. \tag{14.1}$$

The first inequality is a relatively straightforward application of Theorem 13.28 and we will prove it in this subsection. For the second inequality, it turns out that the expected characteristic polynomial is exactly the "matching polynomial" of the graph, a well-studied object in the literature, and the upper bound $2\sqrt{d-1}$ is already proved by Heilmann and Lieb in the 70s. We will compute the expected polynomial in the next subsection, and then review some classical results about matching polynomials in Section 14.3.

Theorem 14.9 (Probabilistic Method for Signed Matrices)**.** *For any d-regular graph $G = (V, E)$, there exists a signing $s : E(G) \to \{\pm1\}$ such that $\lambda_{\max}\big(\det(xI - A_s)\big) \leq \lambda_{\max}\big(\mathbb{E}_s\big[\det(xI - A_s)\big]\big)$, where the expectation is over the uniform distribution of all the signings of the edges.*

Proof. To apply Theorem 13.28, we would like to write A_s as a sum of independent random rank one symmetric matrices. Note that we can write $A_s = \sum_{e\in E} A_e$, where each A_e is a random variable with $(A_e)_{ij} = (A_e)_{ji} = s(ij)$ if $e = ij$ and all other entries zero. The issue is that A_e is a rank two matrix, not rank one. Instead, we consider the random variable $L_e = D_e + A_e$, where $(D_e)_{ii} = (D_e)_{jj} = 1$ if $e = ij$ with all other entries zero. So, each L_e is the signed Laplacian matrix of an edge, which is a rank one matrix. Denote $L_s = \sum_{e\in E} L_e$, which is a sum of independent random rank one symmetric matrices. Note that $L_s = dI + A_s$ as the graph is d-regular, and thus $\det(xI - A_s)$ can be written as $\det((x+d)I - L_s)$. Applying Theorem 13.28 on the random variables $\{L_e\}_{e\in E}$, there exists a signing $s : E(G) \to \{\pm1\}$ such that

$$\lambda_{\max}\big(\det(yI - L_s)\big) \leq \lambda_{\max}\big(\mathbb{E}_s\big[\det(yI - L_s)\big]\big)$$

By doing a change of variable $y = x+d$, the same signing s satisfies the statement of the theorem. □

Theorem 14.9 proves the first inequality in Equation 14.1.

Expected Characteristic Polynomial

Perhaps surprisingly, the expected characteristic polynomial is already known to be equal to the matching polynomial of a graph, a well-studied polynomial in Combinatorics.

Definition 14.10 (Matching Polynomials)**.** *Given an undirected graph $G = ([n], E)$, let m_i be the number of matchings in G with i edges with $m_0 = 1$, the matching polynomial of G is defined as*

$$\mu_G(x) := \sum_{i \geq 0} (-1)^i \cdot m_i \cdot x^{n-2i}.$$

The following identity is by Godsil and Gutman (see [God93]).

Theorem 14.11 (Expected Characteristic Polynomial is Matching Polynomial)**.** *Given an undirected graph $G = ([n], E)$, the expected characteristic polynomial of the signed matrices is*

$$\mathbb{E}_{s \in \{\pm 1\}^{|E|}} \big[\det(xI - A_s) \big] = \mu_G(x),$$

where the expectation is over the uniform distribution of all the signings of the edges.

Proof. Let $M_s = xI - A_s$. We expand the determinant of M_s as sum of permutations as in Fact 2.26, so that

$$\mathbb{E}_s \det(xI - A_s) = \mathbb{E}_s \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} (M_s)_{i,\sigma(i)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \mathbb{E}_s \left[\prod_{i=1}^{n} (M_s)_{i,\sigma(i)} \right],$$

where $\operatorname{sgn}(\sigma) = (-1)^{\operatorname{inv}(\sigma)}$ and $\operatorname{inv}(\sigma) := \big|\{(i,j) \mid i < j \text{ and } \sigma(i) > \sigma(j)\}\big|$ is the number of inversion pairs of the permutation σ. Since each edge is independent and $\mathbb{E}\left[(M_s)_{i,j}\right] = 0$ as each edge is equally likely to be ± 1, all the permutations with at least one variable with degree one vanished. Therefore, for the permutations remained, each edge $(M_s)_{ij}$ appears exactly twice and

$$\mathbb{E}_s \left[\prod_{i=1}^{n} (M_s)_{i,\sigma(i)} \right] = x^{n-2k} \cdot \prod_{l=1}^{k} (M_s)_{i_l,j_l}^2 = x^{n-2k}$$

for some k. So, each matching of size k will contribute $\operatorname{sgn}(\sigma)$ to the coefficient of x^{n-2k}. Check that each matching of size k has the same sign, with $\operatorname{sgn}(\sigma) = -1$ if k is odd and $\operatorname{sgn}(\sigma) = +1$ if k is even. We conclude that $\mathbb{E}_s \det(xI - A_s) = \sum_{k \geq 0} (-1)^k \cdot m_k \cdot x^{2n-k} = \mu_G(x)$. □

14.3 Matching Polynomials

Quite amazingly, the maximum root of the matching polynomial was studied by Heilmann and Lieb in 1972 and their result is exactly what is needed for bipartite Ramanujan graphs.

Theorem 14.12 (Heilmann-Lieb)**.** *For any undirected graph G of maximum degree d, the matching polynomial $\mu_G(x)$ is real-rooted with maximum root at most $2\sqrt{d-1}$.*

So, the results by Godsil-Gutman in Theorem 14.11 and Heilmann-Lieb in Theorem 14.12 combined to establish the second inequality in Equation 14.1, and this completes the proof of Theorem 14.6.

The original proof by Heilmann-Lieb uses recursion and induction. We present an approach by Godsil [God93] which consists of three steps:

1. The matching polynomial of a graph G of maximum degree d divides the matching polynomial of an associated tree T (called the path tree) of maximum degree d.
2. The matching polynomial of a tree T is equal to the characteristic polynomial $\det(xI - A_T)$ of its adjacency matrix A_T.
3. The maximum eigenvalue of the adjacency matrix A_T of a tree T of maximum degree d is at most $2\sqrt{d-1}$.

Since the characteristic polynomial of the adjacency matrix of a tree (and more generally a graph) is real-rooted, (1) and (2) imply that the matching polynomial of a graph is real-rooted. Therefore, the maximum root of the matching polynomial of G is

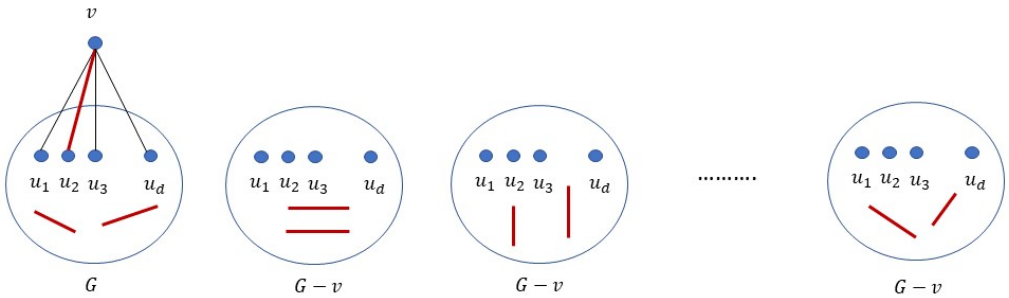
$$\lambda_{\max}\big(\mu_G(x)\big) \le \lambda_{\max}\big(\mu_T(x)\big) = \lambda_{\max}\big(\det(xI - A_T)\big) \le 2\sqrt{d-1},$$

where the first inequality is by (1), the equality is by (2), and the last inequality is by (3). Therefore, proving the three steps would complete the proof of Heilmann and Lieb's result in Theorem 14.12.

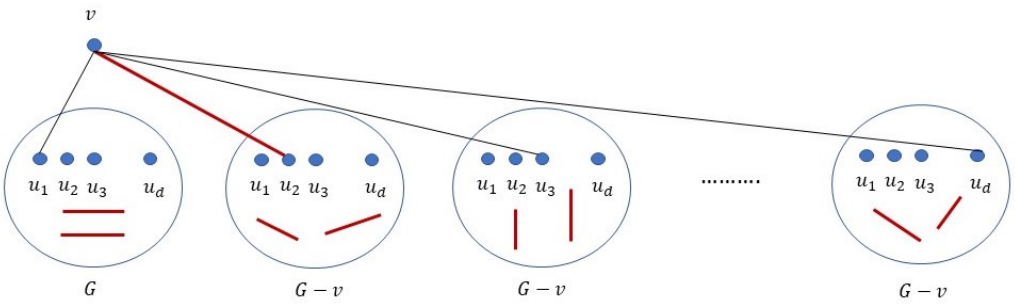
The third step is already done in Problem 3.10. The second step is left as an exercise, as its proof is similar to that in Theorem 14.11, showing that only permutations corresponding to matchings contribute to the characteristic polynomial.

First Step

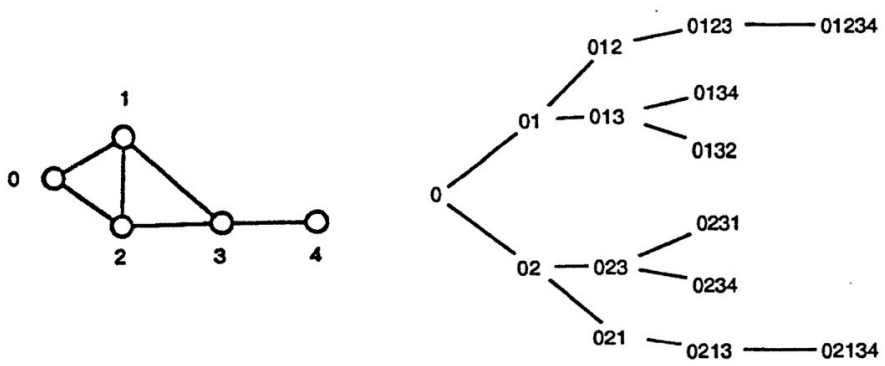
The proof of (1) in [God93] is not long but involves algebraic manipulations. Here we try to give a proof by pictures. Given a graph G, we start from an arbitrary vertex v. Let the degree of v be d, and the neighbors of v be $u_1, \ldots, u_d$. We add $d-1$ copies of $G-v$ to the graph and call the resulting graph H. Check that the matching polynomial of the disjoint union is equal to the product of the matching polynomials, i.e. $\mu_{G_1 \cup G_2}(x) = \mu_{G_1}(x) \cdot \mu_{G_2}(x)$ where $G_1 \cup G_2$ is the disjoint union of two graphs G_1 and G_2. Therefore, $\mu_H(x) = \mu_G(x) \cdot \big(\mu_{G-v}(x)\big)^{d-1}$, and so the matching polynomial of G divides the matching polynomial of H.



Consider the following graph H', where the edge vu_i in G is replaced by vu_i in the i-th copy of $G - v$ for $2 \le i \le d$ (see the picture below). The claim is that the matching polynomials of H and H' are the same. The reason is that there is a one-to-one correspondence between matchings in H and matchings in H', as v can only be matched to one vertex. See the pictures, where the red edges are the edges in a matching. Now, in H', there are no cycles involving v.



Applying the same operations (duplicate and branch, and remove isolated vertices) on u_1 in the first copy of $G-v$, on u_2 in the second copy of $G-v$, and so on, the resulting (big) graph will have no cycles and is a tree. The resulting tree is called the path tree of G, as there is a path in T for each path in G. See the following picture from [God93] for a complete example.



All these operations preserve the property that the matching polynomial of the old graph divides the matching polynomial of the new graph, and so by induction the matching polynomial of the original graph G divides the matching polynomial of the path tree, which has maximum degree at most d. This "proves" the first step. See Chapter 6 of [God93] for the formal proof.

14.4 Discussions and Problems

One obvious open question is whether this approach can be extended to construct a true Ramanujan graph (that satisfies $|\alpha_n| \leq 2\sqrt{d-1}$). There is a trick to get something close.

Problem 14.13 (Twice Ramanujan Graphs)**.** *Show that the current approach can be used to construct a d-regular graph with* $\max\{\alpha_2, |\alpha_n|\} \leq 4\sqrt{d-1}$.

Another obvious open question is whether this approach can be made efficient algorithmically. Note that the natural attempt would not work, as it is NP-hard to compute the coefficients of matching polynomials. Marcus, Spielman, Srivastava [MSS18] gave another construction of bipartite Ramanujan graphs using interlacing families for permutations, and Cohen [Coh16] showed that their construction can be implemented in polynomial time.

The following is an exercise that completes the second step of Godsil's proof.

Exercise 14.14 (Matching Polynomial of a Tree)**.** *Prove that the matching polynomial of a tree T is equal to the characteristic polynomial* $\det(xI - A_T)$ *of its adjacency matrix A_T.*

The following are some identities for matching polynomials, which can be used to give a formal proof that the matching polynomial of a graph G divides the matching polynomial of its path tree. They can be proved by some simple relations between the number of matchings in a graph and its subgraphs.

Problem 14.15 (Identities for Matching Polynomials [God93])**.**

1. $\mu_{G\cup H}(x) = \mu_G(x) \cdot \mu_H(x)$ *for disjoint G and H.*
2. $\mu_G(x) = \mu_{G\setminus e}(x) - \mu_{G\setminus uv}(x)$ *if $e = uv$ is an edge of G.*
3. $\mu_G(x) = x \cdot \mu_{G\setminus u}(x) - \sum_{i\sim u} \mu_{G\setminus ui}(x)$ *if $u \in V(G)$.*

14.5 References

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Comb.*, 26(5):495–519, 2006. 59, 137, 138

[Coh16] Michael B. Cohen. Ramanujan graphs in polynomial time. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 276–281. IEEE Computer Society, 2016. 142

[God93] Chris D. Godsil. *Algebraic combinatorics*. Chapman and Hall mathematics series. Chapman and Hall, 1993. 140, 141, 142, 143

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 26, 29, 57, 60, 61, 62, 63, 67, 68, 137

[MSS15] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. *Annals of Mathematics*, 182:307–325, 2015. 135, 137, 138, 139

[MSS18] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families IV: bipartite ramanujan graphs of all sizes. *SIAM J. Comput.*, 47(6):2488–2509, 2018. 142, 169

Chapter 15

Multivariate Barrier Method

Marcus, Spielman, and Srivastava [MSS15] proved Weaver's conjecture using the method of interlacing family of polynomials. A key component in their proof is a multivariate barrier method to bound the maximum root of the expected characteristic polynomial, which is an extension of the barrier method by Batson, Spielman and Srivastava for spectral sparsification in Chapter 10.

15.1 Weaver's Conjecture

It was observed that the linear-sized spectral sparsification result by Batson, Spielman, and Srivastava in Theorem 10.1 looks similar to the conjecture by Weaver, which is known to be equivalent to the Kadison-Singer problem (see [MSS15, MSS14]), whose positive resolution would have implications in several areas of mathematics.

Conjecture 15.1 (Weaver's Conjecture)**.** *There exist positive constants α and ϵ so that the following holds. For every $m, n \in \mathbb{N}$ and every set of vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ such that*

$$\sum_{i=1}^{m} v_i v_i^T = I_n \quad \text{and} \quad \|v_i\|_2^2 \leq \alpha \quad \text{for} \quad 1 \leq i \leq n,$$

there exists a partition of $\{1, \ldots, m\}$ into two sets S_1 and S_2 so that

$$\lambda_{\max}\left(\sum_{i \in S_j} v_i v_i^T\right) \leq 1 - \epsilon \quad \text{for} \quad 1 \leq j \leq 2.$$

Note that since $\sum_{i \in S_1} v_i v_i^T + \sum_{i \in S_2} v_i v_i^T = I_n$, the conclusion in Weaver's conjecture is equivalent to

$$\epsilon I_n \preccurlyeq \sum_{i \in S_1} v_i v_i^T \preccurlyeq (1-\epsilon) I_n \quad \text{and} \quad \epsilon I_n \preccurlyeq \sum_{i \in S_2} v_i v_i^T \preccurlyeq (1-\epsilon) I_n,$$

and so the vectors in S_1 and in S_2 form spectral approximators of the identity matrix. In Theorem 10.1 by Batson, Spielman, and Srivastava, the task was to find scalars $s_1, \ldots, s_m$ with few nonzeros so that

$$(1-\epsilon) I_n \preccurlyeq \sum_{i=1}^{m} s_i v_i v_i^T \preccurlyeq (1+\epsilon) I_n \quad \iff \quad \frac{1}{2}(1-\epsilon) I_n \preccurlyeq \sum_{i=1}^{m} \frac{s_i}{2} v_i v_i^T \preccurlyeq \frac{1}{2}(1+\epsilon) I_n$$

So, if all the scalars $s_i/2$ are either zero or one, then Theorem 10.1 would have given a positive resolution to Weaver's conjecture. This is not always possible, however, since if there is a long vector (say $\|v_i\| = 1$), then setting $s_i/2$ to be zero or one would violate the minimum eigenvalue and the maximum eigenvalue bounds. This is why there is an additional condition $\|v_i\|_2^2 \leq \alpha$ in Weaver's conjecture, which says that long vectors are the only obstructions to finding such a partitioning.

Graph Sparsification

In terms of graph sparsification, the question in Weaver's conjecture corresponds to finding an unweighted sparsifier. Recall the reduction in Lemma 9.11 and the discussions in Subsection 9.2, the length $\|v_i\|_2^2$ is equal to the effective resistance of the i-th edge in the graph. So, Weaver's conjecture in the graph setting states that if the maximum effective resistance of an edge is small enough, then there is a partitioning of the edges into two groups so that the subgraph formed by each group is a spectral approximator of the original graph. Some examples of graphs with small maximum effective resistance are expander graphs and edge-transitive graphs (such as hypercubes and Cayley graphs).

One could apply the matrix Chernoff bound in Theorem 9.13 to this problem, and it works for $\alpha \lesssim 1/\log^2 n$ with high probability, but this is not enough for Weaver's conjecture. The approach by Batson, Spielman, and Srivastava for spectral sparsification heavily depends on a careful choice of scalars and does not seem applicable for constructing unweighted sparsifiers. See [BST19] for a recent paper on constructing unweighted sparsifiers.

15.2 Probabilistic Formulation

Marcus, Spielman, and Srivastava formulated and proved a probabilistic statement that implies Weaver's conjecture.

Theorem 15.2 (Marcus-Spielman-Srivastava [MSS15]). *Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be independent random vectors with finite support such that*

$$\mathbb{E}\left[\sum_{i=1}^{m} v_i v_i^T\right] = I_n \quad \text{and} \quad \mathbb{E}\left[\|v_i\|_2^2\right] \leq \epsilon \text{ for } 1 \leq i \leq m.$$

Then

$$\Pr\left[\lambda_{\max}\left(\sum_{i=1}^{m} v_i v_i^T\right) \leq \left(1+\sqrt{\epsilon}\right)^2\right] > 0.$$

Reduction

Weaver's conjecture is about partitioning and Theorem 15.2 is about sum of random variables, but there is a simple reduction from the former to the latter.

For each vector $u_i \in \mathbb{R}^n$ in Weaver's problem, define a random vector $v_i \in \mathbb{R}^{2n}$ with two choices:

$$v_i = \sqrt{2}\begin{pmatrix} u_i \\ 0 \end{pmatrix} \text{ with probability } 1/2 \quad \text{and} \quad v_i = \sqrt{2}\begin{pmatrix} 0 \\ u_i \end{pmatrix} \text{ with probability } 1/2.$$

The first choice corresponds to putting u_i in the first group, and the second choice corresponds to putting u_i in the second group. Then, by the assumption that $\sum_{i=1}^m u_i u_i^T = I_n$,

$$\mathbb{E}\left[v_i v_i^T\right] = \begin{pmatrix} u_i u_i^T & 0 \\ 0 & u_i u_i^T \end{pmatrix} \quad \Longrightarrow \quad \sum_{i=1}^m \mathbb{E}\left[v_i v_i\right]^T = \sum_{i=1}^m \begin{pmatrix} u_i u_i^T & 0 \\ 0 & u_i u_i^T \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ 0 & I_n \end{pmatrix} = I_{2n}.$$

Similarly, by the assumption that $\|u_i\|_2^2 \leq \alpha$,

$$\mathbb{E}\left[\|v_i\|_2^2\right] = \mathbb{E}\left[v_i^T v_i\right] = 2\|u_i\|_2^2 \leq 2\alpha.$$

By Theorem 15.2, there exists a choice of $v_1, \ldots, v_m$ such that $\lambda_{\max}(\sum_{i=1}^m v_i v_i^T) \leq (1+\sqrt{2\alpha})^2$. As intended, we put vector u_i into S_1 if we select the first choice for v_i, and put u_i into S_2 otherwise. Then the conclusion from Theorem 15.2 implies that

$$\lambda_{\max}\left(\begin{pmatrix} 2\sum_{i\in S_1} u_i u_i^T & 0 \\ 0 & 2\sum_{i\in S_2} u_i u_i^T \end{pmatrix}\right) \leq (1+\sqrt{2\alpha})^2 \quad \Longrightarrow \quad \lambda_{\max}\left(\sum_{i\in S_j} v_i v_i^T\right) \leq \frac{1}{2}(1+\sqrt{2\alpha})^2$$

for $1 \leq j \leq 2$. So, when α is small enough (say $\alpha \leq \frac{1}{32}$), then $\frac{1}{2}(1+\sqrt{2\alpha})^2 < 1$ and thus Weaver's conjecture follows. We record the following corollary for future references.

Corollary 15.3 (Solution to Weaver's Conjecture)**.** *Under the same setting in Conjecture 15.1, there exists a partition of $[m]$ into two sets S_1 and S_2 such that for $1 \leq j \leq 2$,*

$$\left(\frac{1}{2} - \sqrt{2\alpha} - \alpha\right) \cdot I_n \preccurlyeq \sum_{i\in S_j} v_i v_i^T \preccurlyeq \frac{1}{2}(1+\sqrt{2\alpha})^2 \cdot I_n.$$

Corollary 15.3 is quantitatively stronger than Weaver's formulation, as when α is small enough, we can bound how far is the solution from the ideal partitioning $\frac{1}{2}I$, which will be useful in applications.

Proof Overview

The plan of the proof is to show that there exists a choice of the random variables $v_1, \ldots, v_m$ such that

$$\lambda_{\max}\left(\det\left(xI_n - \sum_{i=1}^m v_i v_i^T\right)\right) \leq \lambda_{\max}\left(\mathbb{E}_{v_1,\ldots,v_m}\left[\det\left(xI_n - \sum_{i=1}^m v_i v_i^T\right)\right]\right) \leq \left(1+\sqrt{\epsilon}\right)^2. \quad (15.1)$$

We have established in Theorem 13.28 that the set of all possible characteristic polynomials $\det(xI_n - \sum_{i=1}^m v_i v_i^T)$ forms an interlacing family, and the root polynomial can be set to be $\mathbb{E}_{v_1,\ldots,v_m}\left[\det\left(xI_n - \sum_{i=1}^m v_i v_i^T\right)\right]$ where the expectation is taken over the independent uniform random distributions on $v_1, \ldots, v_m$. Therefore, by the new probabilistic method for interlacing family in Theorem 12.12, we have already proved the first inequality, using the techniques from real stable polynomials described in Chapter 13.

The main goal of this chapter is to prove the second inequality in Equation 15.1, given the assumptions that $\mathbb{E}\left[\|v_i\|_2^2\right] \leq \epsilon$ for $1 \leq i \leq m$ and $\mathbb{E}\left[\sum_{i=1}^m v_i v_i^T\right] = I_n$. In Chapter 14, when we construct bipartite Ramanujan graphs, the expected characteristic polynomial turns out to be exactly the matching polynomial and there were known results bounding the maximum root. For Weaver's

conjecture, bounding the maximum root of the expected polynomial is a major technical challenge (that took Marcus, Spielman, and Srivastava four years to solve).

Recall the multilinear formula in Theorem 13.20 that

$$\mathbb{E}_{v_1,\ldots,v_m}\left[\det\left(\lambda I-\sum_{i=1}^m v_iv_i^T\right)\right]=\prod_{i=1}^m(1-\partial_{x_i})\det\left(\lambda I+\sum_{i=1}^m x_i\cdot\mathbb{E}\left[v_iv_i^T\right]\right)\Big|_{x_1=x_2=\cdots=x_m=0}.$$

This formula plays an important role in the first step, by showing that the expected characteristic polynomial is real-rooted to establish common interlacing for the new probabilistic method to work. Perhaps unexpectedly, the formula also plays an important role in the second step. Their idea is to first prove an upper bound of the "maximum root" of the multivariate polynomial $\det\left(\lambda I+\sum_{i=1}^m x_i\cdot\mathbb{E}\left[v_iv_i^T\right]\right)$, and then maintain a good upper bound after each $(1-\partial_{x_i})$ differential operator is applied. To establish the upper bound, they finally realized that the barrier method developed for linear-sized spectral sparsification in Chapter 10 can be extended to the multivariate setting in a syntatically similar way!

15.3 Multivariate Approach

To bound the maximum root of the univariate polynomial $\mathbb{E}\left[\det\left(\lambda I-\sum_{i=1}^m v_iv_i^T\right)\right]$, the approach taken is to bound the "maximum root" of the multivariate polynomial $\prod_{i=1}^m(1-\partial_{x_i})\det\left(\lambda I+\sum_{i=1}^m x_i\cdot\mathbb{E}\left[v_iv_i^T\right]\right)$, which will be defined in a moment.

First, we use the assumption and define some notations to slightly simplify the statement. Using the assumption that $\mathbb{E}\left[\sum_{i=1}^m v_iv_i^T\right]=I$, we rewrite the RHS of the multilinear formula as

$$\prod_{i=1}^m(1-\partial_{x_i})\det\left(\sum_{i=1}^m(\lambda+x_i)\cdot\mathbb{E}\left[v_iv_i^T\right]\right)\Big|_{x_1=\ldots=x_m=0}=\prod_{i=1}^m(1-\partial_{x_i})\det\left(\sum_{i=1}^m x_i\cdot\mathbb{E}\left[v_iv_i^T\right]\right)\Big|_{x_1=\ldots=x_m=\lambda}.$$

Denote the matrix $\mathbb{E}\left[v_iv_i^T\right]=B_i$ and note that $B_i\succcurlyeq 0$ for $1\le i\le m$. Denote the polynomial after applying the differential operator k times by

$$p_k(x_1,\ldots,x_m):=\prod_{i=1}^k(1-\partial_{x_i})\det\left(\sum_{i=1}^m x_iB_i\right),\tag{15.2}$$

so that $p_0(x_1,\ldots,x_m)=\det(\sum_{i=1}^m x_iB_i)$ and $p_m(x_1,\ldots,x_m)=\prod_{i=1}^m(1-\partial_{x_i})\det\left(\sum_{i=1}^m x_iB_i\right)$.

Definition 15.4 (Above the Roots)**.** *Given a multivariate polynomial $p(x_1,\ldots,x_m)$, we say a point $y\in\mathbb{R}^m$ is "above the roots" of p if $p(y+t)>0$ for all $t=(t_1,\ldots,t_m)\in\mathbb{R}^m_{\ge 0}$.*

Our goal is to prove that the point $(1+\sqrt{\epsilon})^2\cdot\vec{1}$ is above the roots of the multivariate polynomial $p_m(x_1,\ldots,x_m)$. Note that this implies that the maximum root of the univariate polynomial $p_m(\lambda,\ldots,\lambda)=\prod_{i=1}^m(1-\partial_{x_i})\det\left(\sum_{i=1}^m x_iB_i\right)|_{x_1=\ldots=x_m=\lambda}$ is at most $(1+\sqrt{\epsilon})^2$, and thus by the multilinear formula the maximum root of the univariate polynomial $\mathbb{E}_{v_1,\ldots,v_m}\left[\det\left(\lambda I-\sum_{i=1}^m v_iv_i^T\right)\right]$ is at most $(1+\sqrt{\epsilon})^2$.

Initially, since $\sum_{i=1}^m B_i=I_n$ by assumption, it follows that $p_0(t,t,\ldots,t)=\det(tI)>0$ for any $t>0$, and so the point $t\cdot\vec{1}$ is above the roots of p_0 for any $t>0$. The strategy in [MSS15] is to prove inductively that $(\underbrace{t+\delta,\ldots,t+\delta}_{k\text{ coordinates}},t\ldots,t)$ is above the roots of p_k for some δ for all $1\le k\le m$.

Multivariate Barrier Functions

To execute the above inductive proof strategy, a similar approach as in Chapter 10 for spectral sparsification is used, to establish a "soft/comfortable" upper bound for the induction to go through.

Recall that in Definition 10.3, the potential function $\Phi^u(A) = \mathrm{Tr}(uI - A)^{-1}$ is defined, and the invariant $\Phi^u(A) \leq \phi$ is maintained to guarantee that u is well above the roots. Also recall from Remark 10.6 and Remark 11.6 that $\Phi^u(A) = p'_A(u)/p_A(u)$ where $p_A(x) = \det(xI - A)$ is the characteristic polynomial of A, and so the potential function has a natural interpretation in terms of polynomials. This univariate barrier function is generalized to the multivariate setting as follows.

Definition 15.5 (Multivariate Barrier Functions)**.** *Given a real-stable polynomial $p \in \mathbb{R}[x_1, \ldots, x_m]$ and a point $y \in \mathbb{R}^m$ above the roots of p, for $1 \leq i \leq m$, the barrier function of p in direction i at y is defined as*

$$\Phi_p^i(y) := \frac{\partial_{x_i} p(y)}{p(y)}.$$

Equivalently, we can define

$$\Phi_p^i(y) = \frac{q'_{y,i}(y_i)}{q_{y,i}(y_i)} = \sum_{j=1}^{d} \frac{1}{y_i - \lambda_j},$$

where $q_{y,i}(t)$ is the univariate restriction $q_{y,i}(x) = p(y_1, \ldots, y_{i-1}, t, y_{i+1}, \ldots, y_m)$ where $\lambda_1, \ldots, \lambda_d$ are the roots of this univariate polynomial, which is real-rooted as substituting real numbers preserve real-stability by Proposition 13.13.

For spectral sparsification, we maintain one potential function $\Phi^A(x)$ to show that $u \in \mathbb{R}$ is well above the roots by showing that $\Phi^A(u) \leq \phi$ for some small $\phi \in \mathbb{R}$. For Weaver's problem, we maintain m potential functions $\Phi_p^1(x), \ldots, \Phi_p^m(x)$ to show that $y \in \mathbb{R}^m$ is well above the roots by showing that $\Phi_p^i(y) \leq \phi$ for $1 \leq i \leq m$ for some small $\phi \in \mathbb{R}$.

Definition 15.6 (Induction Hypothesis)**.** *Let $x_0 = (t, \ldots, t) \in \mathbb{R}^m$ be the initial point above the roots of p_0 in Equation 15.2 for some $t > 0$, such that $\Phi_p^i(x_0) \leq \phi$ for some $\phi \in \mathbb{R}$ for $1 \leq i \leq m$. This is the base case. Let $x_k = (t + \delta, \ldots, t + \delta, t, \ldots, t)$ with the first k coordinates being $t + \delta$. The induction hypothesis is to maintain that $\Phi_{p_k}^i(x_k) \leq \phi$ for $1 \leq i \leq m$, for $1 \leq k \leq m$ where p_k is defined in Equation 15.2. The parameters t, ϕ, δ will be chosen at the end.*

15.4 Bounding the Maximum Root

In this section, we do the calculations to carry out the induction as described in Definition 15.6.

Monotonicity and Convexity

The following monotonicity and convexity properties are generalizations of Exercise 10.4 in the univariate case to the multivariate setting.

Proposition 15.7 (Monotonicity and Convexity)**.** *Suppose $p \in \mathbb{R}[x_1, \ldots, x_m]$ is real-stable and y is above the roots of p. Then, for all $i, j \in [m]$ and $\delta \geq 0$, the following two properties hold.*

1. *Monotonicity: $\Phi_p^i(y + \delta \cdot e_j) \leq \Phi_p^i(y)$ where e_j is the j-th vector in the standard basis.*

2. *Convexity:* $\Phi_p^i(y+\delta\cdot e_j)-\Phi_p^i(y)\le\delta\cdot\partial_{x_j}\Phi_p^i(y+\delta\cdot e_j)$.

The proof of the univariate case in Exercise 10.4 is easy, but the proof of the multivariate case in Proposition 15.7 is not. The proof in [MSS15] uses a deep result that any bivariate real-stable polynomial $p(x_1,x_2)$ can be written as $\pm\det(x_1A+x_2B+C)$ for some $A,B\succcurlyeq 0$ and some symmetric C. Then they do some explicit computations from this representation to prove monotonicity and convexity.

Tao [Tao13] gave a more elementary proof using complex analysis. We gave a proof sketch of Tao's proof in L15 in the previous offering of CS 860. We won't give a proof of this proposition in this offering, and refer the reader to [MSS15, Tao13]. One reason is that the arguments are different and independent from the rest of the proofs and also not self-contained, and another reason is that I don't understand the proofs well enough to provide any further explanations.

Inductive Proof

As a warm up, we first see that when a point y is well above the roots, then y is still above the roots after the operation $1-\partial_{x_j}$.

Lemma 15.8 (Above the Roots after One Operation)**.** *Suppose that $p\in\mathbb{R}[x_1,\dots,x_m]$ is real stable and $y\in\mathbb{R}^m$ is above the roots of p, with the additional property that $\Phi_p^i(y)<1$ for $1\le i\le m$. Then y is still above the roots of $(1-\partial_{x_j})p$ for $1\le j\le m$.*

Proof. Let $z\in\mathbb{R}^m$ be a point above y such that $z\ge y$. We would like to prove that $(1-\partial_{x_j})p(z)\neq 0$ for any $1\le j\le m$, and this would imply that y is still above the roots of $(1-\partial_{x_j})p$. By the monotonicity property in Proposition 15.7, $\Phi_p^j(z)\le\Phi_p^j(y)<1$ for $1\le j\le m$. This implies that

$$1>\Phi_p^j(z)=\frac{\partial_{x_j}p(z)}{p(z)}\quad\Longrightarrow\quad 0\neq p(z)-\partial_{x_j}p(z)=(1-\partial_{x_j})p(z).$$

□

The lemma shows that y is still above the roots after one differential operation, but we cannot repeat this argument because the condition $\Phi^j_{(1-\partial_{x_i})p}(y)<1$ may no longer hold. To maintain the invariant, we will increase the upper bound in the corresponding coordinate to decrease the potential value. Assuming the monotonicity and convexity properties of the multivariate barrier functions in Proposition 15.7, the proof in the following inductive step is actually very similar to the univariate case as presented in Lemma 11.9 (and also in Problem 11.13). Basically, we can do exact calculation to compute the increase of the potential function after the $1-\partial_{x_i}$ operation, and then use convexity to bound the decrease of the potential value by shifting up the barrier to $y+\delta\cdot e_j$.

Lemma 15.9 (Maintaining the Potential Values)**.** *Suppose that $p\in\mathbb{R}[x_1,\dots,x_m]$ is real stable and y is above the roots of p. Suppose further that $\Phi_p^i(y)\le 1-\frac{1}{\delta}$ for $1\le i\le m$ for some $\delta>0$. Then, for $1\le i,j\le m$,*

$$\Phi^i_{(1-\partial_{x_j})p}(y+\delta\cdot e_j)\le\Phi_p^i(y),$$

and $y+\delta\cdot e_j$ is still above the roots of $(1-\partial_{x_j})p$.

Proof. By Definition 15.5,

$$\Phi^i_{(1-\partial_{x_j})p} = \frac{\partial_{x_i}((1-\partial_{x_j})p)}{(1-\partial_{x_j})p} = \frac{\partial_{x_i}((1-\Phi^j_p)p)}{(1-\Phi^j_p)p} = \frac{(1-\Phi^j_p)\partial_{x_i}p}{(1-\Phi^j_p)p} + \frac{(\partial_{x_i}(1-\Phi^j_p))p}{(1-\Phi^j_p)p} = \Phi^i_p - \frac{\partial_{x_i}\Phi^j_p}{1-\Phi^j_p}.$$

Therefore,

$$\Phi^i_{(1-\partial_{x_j})p}(y+\delta e_j) = \Phi^i_p(y+\delta e_j) - \frac{\partial_{x_i}\Phi^j_p(y+\delta e_j)}{1-\Phi^j_p(y+\delta e_j)}.$$

To prove $\Phi^i_{(1-\partial_{x_j})p}(y+\delta e_j) \le \Phi^i_p(y)$, it is equivalent to proving that

$$\Phi^i_p(y) - \Phi^i_p(y+\delta e_j) \ge -\frac{\partial_{x_i}\Phi^j_p(y+\delta e_j)}{1-\Phi^j_p(y+\delta e_j)}. \tag{15.3}$$

By convexity of the multivariate barrier function in Proposition 15.7, we have that

$$\Phi^i_p(y) - \Phi^i_p(y+\delta e_j) \ge -\delta\cdot\partial_{x_j}\Phi^i_p(y+\delta e_j).$$

So, Equation 15.3 holds if we could prove that

$$\delta\cdot\partial_{x_j}\Phi^i_p(y+\delta e_j) \le \frac{\partial_{x_i}\Phi^j_p(y+\delta e_j)}{1-\Phi^j_p(y+\delta e_j)} \quad\iff\quad \delta \ge \frac{1}{1-\Phi^j_p(y+\delta e_j)}, \tag{15.4}$$

where the equivalence is by noting that $\partial_{x_j}\Phi^i_p = \partial_{x_i}\Phi^j_p$ and so the numerators are the same, and $\partial_{x_j}\Phi^i_p(y+\delta e_j) \le 0$ as the barrier function is monotonically decreasing above the roots by Proposition 15.7 and so the inequality is reversed when we cancel the numerators. Our assumption implies that

$$\delta \ge \frac{1}{1-\Phi^j_p(y)} \ge \frac{1}{1-\Phi^j_p(y+\delta e_j)}$$

as desired, where the second inequality is again by monotonicity. We conclude that $\Phi^i_{(1-\partial_{x_j})p}(y+\delta\cdot e_j) \le \Phi^i_p(y)$ for $1\le i,j\le m$, and $y+\delta\cdot e_j$ is still above the roots of $(1-\partial_{x_j})p$ by Lemma 15.8. □

We will choose $\phi = 1-\frac{1}{\delta}$ in the induction hypothesis in Definition 15.6.

Choosing the Parameters

By Lemma 15.9, if we choose the initial $x_0 = (t,\dots,t)$ such that $\Phi^i_p(x_0) \le 1-\frac{1}{\delta}$ for $1\le i\le m$ for some $\delta > 0$. Then, by induction, $x_m = (t+\delta,\dots,t+\delta)$ is above the roots of p_m. This would imply that $t+\delta$ is above the roots of the univariate polynomial $\mathbb{E}\left[\det\left(\lambda I - \sum_{i=1}^m v_i v_i^T\right)\right]$ by the multilinear formula. It remains to optimize t and δ to prove the best upper bound.

Note that for the induction step to go through, we only used the property that the polynomial is real stable, and the general properties of monotonicity and convexity. We have not used the specific form of p_k in Equation 15.2. Also, we have not used the crucial assumption that $\mathbb{E}\left[\|v_i\|_2^2\right] \le \epsilon$. These are (only) needed in the following computation of the initial value.

Recall that $p_0(x_1,\dots,x_m) = \det(\sum_{i=1}^m x_i B_i)$ where $B_i = \mathbb{E}\left[v_i v_i^T\right] \succcurlyeq 0$. The assumptions of Theorem 15.2 are that $\sum_{i=1}^m B_i = \sum_{i=1}^m \mathbb{E}\left[v_i v_i^T\right] = I_n$ and $\mathbb{E}\left[\|v_i\|_2^2\right] = \mathrm{Tr}(B_i) \le \epsilon$. The initial potential function is

$$\Phi^j_{p_0}(x) = \frac{\partial_{x_j}\det\left(\sum_{i=1}^m x_i B_i\right)}{\det\left(\sum_{i=1}^m x_i B_i\right)} = \frac{\det\left(\sum_{i=1}^m x_i B_i\right)\mathrm{Tr}\left(\left(\sum_{i=1}^m x_i B_i\right)^{-1} B_j\right)}{\det\left(\sum_{i=1}^m x_i B_i\right)} = \mathrm{Tr}\left(\left(\sum_{i=1}^m x_i B_i\right)^{-1} B_j\right),$$

where the second equality is by the Jacobi formula in Fact 2.39. Put in $x_0 = (t, \ldots, t)$, using the assumptions that $\sum_{i=1}^m B_i = I$ and $\mathrm{Tr}(B_i) \leq \epsilon$, the initial potential value is

$$\Phi^j_{p_0}(x_0) = \mathrm{Tr}\left((tI)^{-1} B_j\right) = \frac{1}{t}\mathrm{Tr}(B_j) \leq \frac{\epsilon}{t}.$$

If we set t so that $\Phi^j_{p_0}(x_0) \leq \frac{\epsilon}{t} \leq 1 - \frac{1}{\delta}$, then by Lemma 15.9, we will get the final bound $t + \delta$. So, we should set t so that $\frac{\epsilon}{t} = 1 - \frac{1}{\delta}$, and the final bound is

$$t + \frac{1}{1 - \frac{\epsilon}{t}}.$$

This is minimized when $t = \sqrt{\epsilon} + \epsilon$, and the final bound is $(1 + \sqrt{\epsilon})^2$. This completes the proof of Theorem 15.2 using the induction hypothesis in Definition 15.6.

15.5 Discussions

The major open problem is to design a polynomial time algorithm to find the solution in Theorem 15.2.

It is interesting to reflect on the journey to the solution to the Kadison-Singer problem. First, it started from the nice formulation of the spectral sparsification problem, as an intermediate step to design a fast Laplacian solver. Then, the reduction to the isotropy case, which helps to match the cut sparsification result by Benczur and Karger. Then, the barrier method was developed, starting with the heuristic argument from expected characteristic polynomial. Then, the interlacing property was observed, and the heuristic argument was developed as a new probabilistic method of interlacing family. The theory of real-stable polynomials was used to establish that the family in Theorem 15.2 is an interlacing family, with the crucial multilinear formula. Finally, the barrier method was understood as a way to bound roots, and it was extended to the multivariate setting through the multilinear formula to solve the problem. It is an amazing line of work with so many great ideas and techniques developed.

15.6 References

[BST19] Nikhil Bansal, Ola Svensson, and Luca Trevisan. New notions and constructions of sparsification for graphs and hypergraphs. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 910–928. IEEE Computer Society, 2019. 146

[MSS14] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Ramanujan graphs and the solution of the Kadison-Singer problem. In *Proceedings of the international Congress of Mathematicians*, volume 3, pages 363–386. 2014. 117, 145

[MSS15] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison-singer problem. *Annals of Mathematics*, pages 327–350, 2015. 133, 135, 145, 146, 148, 150

[Tao13] Terence Tao. Real stable polynomials and the Kadison-Singer problem. *A blogpost of What's New*, 2013. https://terrytao.wordpress.com/2013/11/04/real-stable-polynomials-and-the-kadison-singer-problem/. 133, 134, 150

Chapter 16

Strongly Rayleigh Measure

We see some applications of the theory of real-stable polynomials in probability theory. Some results will be used in the next chapter to extend the method of interlacing family of polynomials.

We have mentioned in Section 13.1 that the real-rootedness of the generating polynomial of a probability distribution implies some strong properties of the probability distribution. In this chapter, we study the following generalization of this concept to the multivariate setting.

Definition 16.1 (Strongly Rayleigh Measure)**.** *Given a probability distribution $\mu : \{0,1\}^m \to \mathbb{R}$, the generating polynomial is defined as*

$$g_\mu(x_1, \ldots, x_m) := \sum_{S \subseteq [m]} \mu(S) \prod_{i \in S} x_i.$$

We say μ is strongly Rayleigh if its generating polynomial g_μ is a real-stable polynomial.

We will see some interesting examples in Section 16.1, and some useful properties in Section 16.2. Some parts of this chapter are from the course notes of Oveis Gharan [Ove15, Ove20].

16.1 Determinantal Measure

An important example of strongly Rayleigh measure is determinantal measure. This is also called the determinantal point process in the literature (see [KT12]).

Definition 16.2 (Determinantal Measure)**.** *Let X be a random variable over $\{0,1\}^m$ with probability distribution $\mu : \{0,1\}^m \to \mathbb{R}$. We say μ is determinantal if there exists a matrix $A \in \mathbb{R}^{m \times m}$ such that*

$$\Pr(S \subseteq X) = \sum_{R : R \supseteq S} \mu(R) = \det(A_{S,S})$$

for every subset $S \subseteq [m]$, where $A_{S,S}$ is the $|S| \times |S|$-submatrix of A restricting to the rows and columns corresponding to S.

There is a compact formula to write the generating polynomial of μ in terms of A.

Proposition 16.3 (Generating Polynomial of Determinantal Measure)**.** *If $\mu : \{0,1\}^m \to \mathbb{R}$ is determinantal with an $m \times m$ matrix $0 \preccurlyeq A \preccurlyeq I$, then the generating polynomial is*

$$g_\mu(x) = \det(I - A + A \cdot \mathrm{diag}(x)).$$

Proof. Let $h(x) = \det(I - A + A \cdot \mathrm{diag}(x))$ where $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$. We will prove that $h(\chi_S) = g_\mu(\chi_S) = \sum_{R \subseteq S} \mu(R)$ for every subset $S \subseteq [m]$ where χ_S is the characteristic vector of the subset S, and this will imply that $h(x) = g_\mu(x)$. Note that

$$h(\chi_S) = \det\left(I - A + A \cdot \mathrm{diag}(\chi_S)\right) = \det \begin{pmatrix} I_{|S|} & -A_{S,\bar{S}} \\ 0 & I_{m-|S|} - A_{\bar{S},\bar{S}} \end{pmatrix} = \det(I_{m-|S|} - A_{\bar{S},\bar{S}}).$$

Let $\det(A_{\emptyset,\emptyset}) = 1$. Recall the expansion of the characteristic polynomial in Fact 2.31 that

$$\det(\lambda I_n - A) = \sum_{k=0}^{n} \lambda^{n-k}(-1)^k \sum_{S \in \binom{[n]}{k}} \det(A_{S,S}).$$

Applying this formula on h,

$$h(\chi_S) = 1 + \sum_{k=1}^{m-|S|} (-1)^k \sum_{R: R \subseteq \bar{S}, |R|=k} \det(A_{R,R}) = 1 + \sum_{k=1}^{m-|S|} (-1)^k \sum_{R: R \subseteq \bar{S}, |R|=k} \Pr(X \supseteq R)$$

where the second equality is by the definition of determinantal measure where X denotes the random outcome. Using the inclusion-exclusion principle that

$$\Pr(X \cap Y \neq \emptyset) = \sum_{k=1}^{|Y|} (-1)^{k+1} \sum_{R: R \subseteq Y, |R|=k} \Pr(X \supseteq R)$$

for a fixed subset Y and plugging in $Y = \overline{S}$, the above expression can be simplified to

$$h(\chi_S) = 1 - \Pr(X \cap \overline{S} \neq \emptyset) = \Pr(X \cap \overline{S} = \emptyset) = \Pr(X \subseteq S) = \sum_{R \subseteq S} \mu(R) = g(\chi_S).$$

□

Then the proof that determinantal measure is strongly Rayleigh follows from the results of real-stable polynomials in Chapter 13.

Theorem 16.4 (Determinantal Measure is Strongly Rayleigh)**.** *If $\mu : \{0,1\}^m \to \mathbb{R}$ is determinantal, then μ is strongly Rayleigh.*

Proof. Using Proposition 16.3, we just need to prove that $h(x) = \det(I - A + A \cdot \mathrm{diag}(x))$ is a real-stable polynomial. We prove the claim when $0 \prec A \preccurlyeq I$, and the claim for $0 \preccurlyeq A \preccurlyeq I$ will follow from a continuity argument using Hurwitz's Theorem 13.14 as in Proposition 13.13. Note that $h(x) = \det(I - A + A \cdot \mathrm{diag}(x)) = \det(A) \cdot \det(A^{-1} - I + \mathrm{diag}(x))$, where we used that $A \succ 0$ so that A^{-1} exists and also $\det(A) > 0$. Since $0 \prec A \preccurlyeq I$, it follows that $B := A^{-1} - I \succcurlyeq 0$, and so $\det(A^{-1} - I + \mathrm{diag}(x))$ can be written as $\det(B + \sum_{i=1}^m x_i \mathrm{diag}(e_i))$ where $B \succcurlyeq 0$ and $\mathrm{diag}(e_i) \succcurlyeq 0$ for $1 \leq i \leq m$. By Proposition 13.12, $\det(B + \sum_{i=1}^m x_i \mathrm{diag}(e_i))$ is a real-stable polynomial and so is $h(x)$. (Actually, the proof in Proposition 13.12 only proves the case when $B \succ 0$, but the case $B \succcurlyeq 0$ follows from the same continuity argument using Hurwitz's Theorem 13.14.) □

Volume Measure

One interesting example of determinantal measure is the volume measure.

Definition 16.5 (Volume Measure)**.** *Given vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ that satisfies $\sum_{i=1}^m v_i v_i^T = I_n$, the volume measure μ is defined as $\mu(S) = \det\left(\sum_{i \in S} v_i v_i^T\right)$ for each subset $S \subseteq [m]$ with $|S| = n$. Note that μ is well-defined by the Cauchy-Binet formula in Fact 2.30, as*

$$1 = \det(I_n) = \det\left(\sum_{i=1}^m v_i v_i^T\right) = \sum_{S: S \subseteq [m], |S| = n} \det\left(\sum_{i \in S} v_i v_i^T\right) = \sum_{S: S \subseteq [m], |S| = n} \mu(S).$$

The following will be the base case of the proof that volume measure is determinantal.

Lemma 16.6 (Marginal Probability of Volume Measure)**.** *Let X be a random output of the volume measure μ. Then $\Pr(j \in X) = \|v_j\|^2$.*

Proof. Let V be the $n \times m$ matrix with the j-th column being v_j for $1 \leq j \leq m$. Let $V_{\overline{j}}$ be the $n \times (m-1)$ matrix where the j-th column of V is removed. By the Cauchy-Binet formula in Fact 2.30,

$$\Pr(j \notin X) = \sum_{S: j \notin S, |S| = n} \det\left(\sum_{i \in S} v_i v_i^T\right) = \det(V_{\overline{j}} V_{\overline{j}}^T).$$

By the matrix determinantal formula in Fact 2.29 and the assumption that $VV^T = I_n$,

$$\det(V_{\overline{j}} V_{\overline{j}}^T) = \det(VV^T - v_j v_j^T) = \det(VV^T)\left(1 - v_j^T (VV^T) v_j\right) = 1 - \|v_j\|^2.$$

□

The Gram matrix of the vectors $v_1, \ldots, v_m$ shows that the volume measure is determinantal. The proof of the following theorem uses that the formula for the characteristic polynomial and the inclusion-exclusion principle are the same.

Theorem 16.7 (Volume Measure is Determinantal)**.** *Let $Y = V^T V$ be the $m \times m$ Gram matrix of the vectors $v_1, \ldots, v_m \in \mathbb{R}^n$. Let X be a random output of the volume measure μ. For any $S \subseteq [m]$,*

$$\Pr_{X \sim \mu}[S \subseteq X] = \det(Y_{S,S}).$$

Proof. We prove by induction on the size of S. The base case when $|S| = 1$ was done in Lemma 16.6. For the induction step, as in the proof of Lemma 16.6, note that

$$\Pr_{X \sim \mu}[X \cap S = \emptyset] = \frac{\det\left(VV^T - \sum_{i \in S} v_i v_i^T\right)}{\det(VV^T)} = \det\left(I_n - \sum_{i \in S} v_i v_i^T\right).$$

On one hand, by the inclusion-exclusion principle,

$$\Pr_{X \sim \mu}[X \cap S = \emptyset] = 1 - \Pr_{X \sim \mu}[X \cap S \neq \emptyset] = 1 + \sum_{k=1}^{|S|} (-1)^k \sum_{R: R \subseteq S, |R| = k} \Pr_{X \sim \mu}[X \supseteq R].$$

On the other hand, let V_S be the $n \times |S|$ submatrix of V with the columns in S. By $\det(I+XY) = \det(I+YX)$ in Fact 2.28 and the formula for the characteristic polynomial in Fact 2.31,

$$\det\left(I_n - \sum_{i\in S} v_i v_i^T\right) = \det(I_n - V_S V_S^T) = \det(I_{|S|} - V_S^T V_S) = 1 + \sum_{k=1}^{|S|}(-1)^k \sum_{R:R\subseteq S,|R|=k} \det(V_R^T V_R).$$

By the induction hypothesis, $\Pr(R \subseteq X) = \det(Y_{R,R}) = \det(V_R^T V_R)$ for all $R \subset S$. So, there is a one-to-one correspondence between the (inner) summands in the above two equations for $|R| \leq k-1$, and hence we must have $\Pr_{X\sim\mu}[S \subseteq X] = \det(V_S^T V_S) = \det(Y_{S,S})$ as stated. □

Combining Theorem 16.7 and Proposition 16.3 gives a formula for the generating polynomial of the volume measure.

Corollary 16.8 (Generating Polynomial of Volume Measure)**.** *Let μ be a volume measure as defined in Definition 16.5, and Y be the Gram matrix of the vectors as defined in Theorem 16.7. Then the generating polynomial is*

$$g_\mu(x) = \det\left(I - Y + Y \cdot \mathrm{diag}(x)\right).$$

Spanning Tree Measure

An interesting consequence of Theorem 16.7 is that the uniform distribution on spanning trees is determinantal.

Definition 16.9 (Spanning Tree Measure)**.** *Let $G = (V, E)$ be an undirected graph with edge weight w_e on each edge $e \in E$. Let the edge set E be the ground set. Let $\mu : \{0,1\}^{|E|} \to \mathbb{R}$ be the probability distribution with $\mu(T) \propto \prod_{e\in T} w_e$ if T is a spanning tree and zero otherwise.*

The uniform distribution of the spanning trees in a graph is a special case when $w_e = 1$ for all $e \in E$. Using the proof in the matrix tree theorem in Problem 3.24 and the reduction to the identity matrix as in Lemma 9.11, one can show that the spanning tree measure is a volume measure.

Problem 16.10 (Burton-Pemantle Theorem)**.** *Prove that the spanning tree measure for any weighted undirected graph is a volume measure (and hence determinantal and strongly Rayleigh).*

(It may be more convenient to consider the matrix $L(G) + \vec{1}\vec{1}^T$ so that it is invertible and do the matrix tree theorem with this modified Laplacian matrix.)

A nice corollary of Problem 16.10 is that we have a nice formula from Theorem 16.7 to compute the probability that a subset of edges $F \subseteq E$ is contained in a random spanning tree.

16.2 Properties of Strongly Rayleigh Measure

Some useful properties of strongly Rayleigh measures follow from closure properties of real stable polynomials in Chapter 13.

Exercise 16.11 (Strongly Rayleigh Preserving Operations)**.** *Suppose $\mu : \{0,1\}^m \to \mathbb{R}$ is a strongly Rayleigh measure. Prove the following distributions are also strongly Rayleigh.*

1. *(Conditioning:) The conditional probability distributions $\mu|_{x_i=0}$ and $\mu|_{x_i=1}$ where the i-th variable is fixed to 0 or 1 for some $1 \le i \le m$.*

2. *(Projection:) For a subset $S \subseteq [m]$, the projection of μ onto S, denoted by $\mu|_S$, is the distribution supported on subsets of S, where for any $R \subseteq S$,*

$$\mu|_S(R) = \sum_{T \subseteq [m]: T \cap S = R} \mu(T).$$

3. *(External Field:) Given a non-negative vector $(\lambda_1, \ldots, \lambda_m)$, $\mu * \lambda$ is the distribution where*

$$\mu * \lambda(S) = \mu(S) \cdot \prod_{i \in S} \lambda_i.$$

The following exercise shows that some concentration property holds for strongly Rayleigh distributions.

Exercise 16.12 (Rank Sequence)**.** *Suppose $\mu : \{0,1\}^m \to \mathbb{R}$ is a strongly Rayleigh measure. For $0 \le i \le m$, let $a_i = \Pr_{S \sim \mu}[|S| = i]$. Use Problem 13.7 to show that the sequence $a_0, \ldots, a_m$ is ultra log-concave as defined in Definition 13.6.*

Truncation

We will use the following result from real stable polynomials.

Lemma 16.13 (Homogenization)**.** *Given a polynomial $p \in \mathbb{R}[x_1, \ldots, x_m]$, the homogenized version of p, denoted by p_H, is defined as*

$$p_H(x_1, \ldots, x_m, x_{m+1}) = x_{m+1}^{\deg p} \cdot p\Big(\frac{x_1}{x_{m+1}}, \ldots, \frac{x_m}{x_{m+1}}\Big).$$

For any real stable polynomial $p \in \mathbb{R}[x_1, \ldots, x_m]$ with non-negative coefficients, p_H is real stable.

The following truncation operation is quite useful.

Theorem 16.14 (Truncation)**.** *Given a distribution μ and an integer $k \ge 1$, the truncation of μ is defined as the distribution μ_k where $\mu_k(S) \propto \mu(S)$ if $|S| = k$ and zero otherwise. For any strongly Rayleigh distribution μ and any $1 \le k \le n$, μ_k is strongly Rayleigh.*

Proof. Let $g_\mu(x_1, \ldots, x_m)$ be the generating polynomial of μ. As μ is strongly Rayleigh, g_μ is real stable. Consider the homogenized version $(g_\mu)_H$ of g_μ. By Lemma 16.13, $(g_\mu)_H$ is also real stable. Let the degree of g_μ be d. Observe that the generating polynomial of μ_k is

$$g_{\mu_k} \propto \partial_{x_{m+1}}^{d-k} g_{\mu_H}\Big|_{x_{m+1}=0},$$

because only the terms with the degree of x_{m+1} being $d-k$ remained, and those terms have total degree exactly k in other variables $x_1, \ldots, x_m$ in the homogenized polynomial $(g_\mu)_H$. By Exercise 13.17 and Proposition 13.13, differentiation and specialization preserves real stability and so g_{μ_k} is real stable. Therefore, by Definition 16.1, μ_k is strongly Rayleigh. □

This provides an alternative proof that the volume measure is strongly Rayleigh.

Problem 16.15. *Let $L \succcurlyeq 0$ be an $m \times m$ matrix. Prove that the polynomial $\sum_{S:S\subseteq[m]} \det(L_{S,S}) \cdot z^S$ is real stable. Conclude that the volume measure in Definition 16.5 is strongly Rayleigh.*

One useful implication is that the determinantal point process restricted to size k subsets, called k-DPP, is still strongly Rayleigh, even though it is no longer determinantal. So, in particular, k-DPP still enjoys the nice properties of strongly Rayleigh distributions, including the negative correlation property in the following subsection.

Negative Correlation

This is probably the most important property of strongly Rayleigh distributions, as for instance it allows us to apply Chernoff bounds on the variables to prove concentration results.

The simplest form of negative dependency is $\Pr(x_i = 1 \mid x_j = 1) \leq \Pr(x_i = 1)$. Note that the probability $\Pr(x_i = 1)$ can be read from the generating probability as

$$\Pr(x_i = 1) = \partial_{x_i} g(x_1, \ldots, x_m)\Big|_{x_1=\ldots=x_m=1} = \sum_{S:S\ni i} \mu(S),$$

the sum of the coefficients containing i. Therefore, we can rewrite the negative correlation inequality as $\Pr(x_i = 1 \cap x_j = 1) \leq \Pr(x_i = 1) \cdot \Pr(x_j = 1)$, and then express it using generating polynomial as

$$\big(\partial_{x_i}\partial_{x_j} g(\vec{1})\big) \cdot g(\vec{1}) \leq \big(\partial_{x_i} g(\vec{1})\big)\big(\partial_{x_j} g(\vec{1})\big).$$

Strongly Rayleigh measures satisfy this inequality for any $y \in \mathbb{R}^m$, not just for $y = \vec{1}$.

Theorem 16.16 (Negative Correlation)**.** *Let $g(x_1, \ldots, x_m)$ be a multi-linear real stable polynomial. Then, for all $i \neq j$, for all $y \in \mathbb{R}^m$,*

$$\big(\partial_{x_i}\partial_{x_j} g(y)\big) \cdot g(y) \leq \big(\partial_{x_i} g(y)\big)\big(\partial_{x_j} g(y)\big).$$

Proof. For any $y \in \mathbb{R}^m$, consider the bivariate restriction

$$f(s,t) = g(y_1, \ldots, y_{i-1}, y_i + s, y_{i+1}, \ldots, y_{j-1}, y_j + t, y_{j+1}, \ldots, y_m).$$

Then f is a bivariate real stable polynomial. Since g is multi-linear, note that

$$f(s,t) = g(y) + s \cdot \partial_{x_i} g(y) + t \cdot \partial_{x_j} g(y) + st \cdot \partial_{x_i}\partial_{x_j} g(y).$$

The univariate polynomial $h(s) = f(s, \imath)$ is $\mathcal{H}$-stable (but not necessarily real). Let $a + \imath b$ be a root of $h(s)$. Then

$$\Re(h(a + \imath b)) = g(y) + a \cdot \partial_{x_i} g(y) - b \cdot \partial_{x_i}\partial_{x_j} g(y) = 0,$$
$$\Im(h(a + \imath b)) = b \cdot \partial_{x_i} g(y) + \partial_{x_j} g(y) + a \cdot \partial_{x_i}\partial_{x_j} g(y) = 0.$$

Solving the two equations by eliminating a, we get

$$\big(\partial_{x_i}\partial_{x_j} g(y)\big) \cdot g(y) - \big(\partial_{x_i}\partial_{x_j} g(y)\big)^2 \cdot b = \big(\partial_{x_i} g(y)\big)^2 \cdot b + \big(\partial_{x_i} g(y)\big) \cdot \big(\partial_{x_j} g(y)\big).$$

By stability of h, we have $b \leq 0$. This implies that $(\partial_{x_i}\partial_{x_j} g(y)) \cdot g(y) - (\partial_{x_i} g(y)) \cdot (\partial_{x_j} g(y)) \leq 0$. □

The converse of Theorem 16.16 is also true; see [Ove20].

Negative Association

A stronger form of negative dependency is called negative association.

Definition 16.17 (Negative Association)**.** *The binary random variables $\{x_1, \ldots, x_m\}$ are negatively associated if for any two non-decreasing functions $f, g \in \{0,1\}^m \to \mathbb{R}$ that depend on disjoint set of variables, it holds that*

$$\mathbb{E}\left[f(x_1, \ldots, x_m) \cdot g(x_1, \ldots, x_m)\right] \leq \mathbb{E}\left[f(x_1, \ldots, x_m)\right] \cdot \mathbb{E}\left[g(x_1, \ldots, x_m)\right],$$

where a function f is nondecreasing if $f(\vec{x}) \geq f(\vec{y})$ if $\vec{x} \geq \vec{y}$.

Note that negative correlation is a special case of negative association. Feder and Mihail [FM92] used negative correlation as the base case in an induction to prove that the random variables of a strongly Rayleigh measure are negatively associated. Borcea, Brändén and Liggett [BBL09] developed the theory of strongly Rayleigh measure and use it to answer many questions about negatively dependent random variables (see [Pem11]).

One consequence of negative association is that Chernoff bounds apply on the random variables from a strongly Rayleigh distribution, even though they are not independent.

Theorem 16.18 (Concentration of Strongly Rayleigh Distribution)**.** *Let $\mu : \{0,1\}^m \to \mathbb{R}$ be a k-homogeneous strongly Rayleigh distribution. Let $f : \{0,1\}^m \to \mathbb{R}$ be a 1-Lipchitz function where $|f(S_1) - f(S_2)| \leq 1$ for any two sets $S_1, S_2 \subseteq [m]$ with $|S_1 \triangle S_2| = 1$. Then, for any $a \geq 0$,*

$$\Pr\left[|f - \mathbb{E}\left[f\right]| > a\right] \leq e^{-\frac{a^2}{k}}.$$

Oveis Gharan used this theory to prove many interesting properties of random spanning trees, and used these properties to design and analyze approximation algorithms for traveling salesman problems. We refer the reader to his notes [Ove15] for some interesting examples.

16.3 References

[BBL09] Julius Borcea, Petter Brändén, and Thomas Liggett. Negative dependence and the geometry of polynomials. *Journal of the American Mathematical Society*, 22(2):521–567, 2009. 159

[FM92] Tomás Feder and Milena Mihail. Balanced matroids. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 26–38. ACM, 1992. 159

[KT12] Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Found. Trends Mach. Learn.*, 5(2-3):123–286, 2012. 153

[Ove15] Shayan Oveis Gharan. Course notes of "Recent developments in approximation algorithms", lecture 10–12. 2015. https://homes.cs.washington.edu/~shayan/courses/approx/index.html. 153, 159, 226

[Ove20] Shayan Oveis Gharan. Course notes of "Polynomial paradigm in algorithm design", lecture 4. 2020. https://homes.cs.washington.edu/~shayan/courses/polynomials/. 128, 130, 153, 158, 171, 173, 175, 179, 187, 211

[Pem11] Robin Pemantle. Hyperbolicity and stable polynomials in combinatorics and probability. *Current Developments in Mathematics*, 2011(1), 2011. 159

Chapter 17

More Interlacing Families

We see some further developments in the method of interlacing family of polynomials in this chapter.

17.1 Interlacing Family for Strongly Rayleigh Measure

Motivated by the thin tree problem, Anari and Oveis Gharan [AO14] developed an interlacing family for strongly Rayleigh distributions and applied it to the asymmetric traveling salesman problem. We will first study the interlacing family and then discuss the application.

Recall that the interlacing family and the probabilistic formulation by Marcus, Spielman, and Srivastava in Theorem 13.28 and Theorem 15.2 crucially depend on the random variables being independent. Anari and Oveis Gharan proved a beautiful generalization of the probabilistic formulation to strongly Rayleigh measures. In the following, a strongly Rayleigh measure μ is homogeneous if every non-zero monomial in the generating polynomial g_μ is of the same degree.

Theorem 17.1 (Probabilistic Method for Strongly Rayleigh Measure). *Let $\mu : \{0,1\}^m \to \mathbb{R}$ be a homogeneous strongly Rayleigh measure such that the marginal probability $\Pr_{S\sim\mu}[i \in S]$ of each element $1 \le i \le m$ is at most ϵ_1. Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be vectors in isotropic condition $\sum_{i=1}^m v_i v_i^T = I_n$ and $\|v_i\|_2^2 \le \epsilon_2$ for $1 \le i \le m$. Then*

$$\Pr_{S\sim\mu}\left[\left\|\sum_{i\in S} v_i v_i^T\right\| \le 4(\epsilon_1+\epsilon_2) + 2(\epsilon_1+\epsilon_2)^2\right] > 0.$$

Product distributions are strongly Rayleigh distributions, so Theorem 17.1 should be more general than Theorem 15.2, but the leading constant is just slightly larger that it cannot be used to prove Weaver's Conjecture 15.1. It can still be used to prove a multi-partitioning version of Weaver's conjecture for any $r \ge 5$, using a similar reduction as in Section 15.2.

Problem 17.2 (Multi-Partitioning Weaver's Problem). *Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be vectors in isotropic condition $\sum_{i=1}^m v_i v_i^T = I_n$ and $\|v_i\|_2^2 \le \epsilon$ for $1 \le i \le m$. Then, for any r, there is an r partitioning of $[m]$ into $S_1, \ldots, S_r$ such that for any $1 \le j \le r$,*

$$\left\|\sum_{i\in S} v_i v_i^T\right\| \le 4\left(\frac{1}{r}+\epsilon\right) + 2\left(\frac{1}{r}+\epsilon\right)^2.$$

The proof of Theorem 17.1 is based on the same two key steps as in Chapter 14 and Chapter 15:

1. Prove that the family of polynomials $\left\{\det\left(xI-\sum_{i\in S}v_iv_i^T\right)\right\}_{S\in\mathrm{supp}(\mu)}$ forms an interlacing family, and apply the probabilistic method in Theorem 12.12 to show that there exists $S\in\mathrm{supp}(\mu)$ with
$$\lambda_{\max}\left(\det\left(xI-\sum_{i\in S}v_iv_i^T\right)\right)\le\lambda_{\max}\left(\mathbb{E}_{S\sim\mu}\left[\det\left(xI-\sum_{i\in S}v_iv_i^T\right)\right]\right).$$
2. Bound the maximum root of $\mathbb{E}_{S\sim\mu}\left[\det\left(xI-\sum_{i\in S}v_iv_i^T\right)\right]=\sum_S\mu(S)\cdot\det\left(xI-\sum_{i\in S}v_iv_i^T\right)$.

Expected Characteristic Polynomial

Recall that in the solution to the Weaver's conjecture in Chapter 15, both steps depend crucially on the multilinear formula in Theorem 13.20. Anari and Oveis Gharan proved a generalization incorporating the probability measure μ.

Theorem 17.3 (Expected Characteristic Polynomial of Strongly Rayleigh Distribution). *Let g_μ be the homogeneous generating polynomial of a measure $\mu:\{0,1\}^m\to\mathbb{R}$ with degree d. Let $v_1,v_2,\ldots,v_m\in\mathbb{R}^n$. For any $c\in\mathbb{R}$,*
$$c^{d-n}\sum_{R\subseteq[m]}\mu(R)\det\left(c\lambda I-\sum_{i\in R}2v_iv_i^T\right)=\prod_{i=1}^m(1-\partial_{x_i}^2)g_\mu(c\vec{1}+x)\cdot\det\left(\lambda I+\sum_{i=1}^m x_iv_iv_i^T\right)\Big|_{\vec{x}=0}.$$

Proof. Start with the LHS. Let $A_i=v_iv_i^T$, which is rank one so that $\det(\lambda I+\sum_i x_iA_i)$ is multilinear in x_i. Then
$$\begin{aligned}\sum_R\mu(R)\det\left(\lambda I+\sum_{i\in R}x_iA_i\right)&=\sum_R\mu(R)\sum_{S\subseteq R}x^S\left(\prod_{i\in S}\partial_{x_i}\det\left(\lambda I+\sum_{j\in R}x_jA_j\right)|_{\vec{x}=0}\right)\\&=\sum_R\mu(R)\sum_{S\subseteq R}x^S\left(\prod_{i\in S}\partial_{x_i}\det\left(\lambda I+\sum_{j=1}^m x_jA_j\right)|_{\vec{x}=0}\right)\\&=\sum_S x^S\Big(\sum_{R:R\supseteq S}\mu(R)\Big)\left(\prod_{i\in S}\partial_{x_i}\det\left(\lambda I+\sum_{j=1}^m x_jA_j\right)|_{\vec{x}=0}\right)\end{aligned}\tag{17.1}$$
where the first equality is by the multilinear expression of $\det(\lambda I+\sum_{i\in R}x_iA_i)$ as described in Subsection 13.4.

The idea is to come up with one polynomial g with coefficient $\sum_{R:R\supseteq S}\mu(R)$ on the monomial x^S, and another polynomial f with coefficient $\prod_{i\in S}\partial_{x_i}\det(\lambda I+\sum_{j=1}^m x_jA_j)|_{\vec{x}=0}$ on the monomial x^S. Clearly $f(x):=\det(\lambda I+\sum_{i=1}^m x_iA_i)$.

Consider $g_\mu(c\vec{1}+x)=\sum_R\mu(R)\prod_{i\in R}(c+x_i)$. Each R with $R\supseteq S$ contributes $\mu(R)\cdot c^{|R|-|S|}$ to x^S. Therefore, since μ is homogeneous, the coefficient of x^S in g_μ is $\sum_{R:R\supseteq S}\mu(R)\cdot c^{|R|-|S|}=c^{d-|S|}\sum_{R:R\supseteq S}\mu(R)$. Let $g(x):=g_\mu(c\vec{1}+x)$.

Since both f and g are multilinear in x_i, the coefficient of $\prod_{i\in S}x_i^2$ in $f\cdot g$ is the product of the coefficients of x^S in f and in g. We can read the coefficient of $\prod_{i\in S}x_i^2$ in $f\cdot g$ using the formula $2^{-|S|}\prod_{i\in S}\partial_{x_i}^2 f\cdot g|_{\vec{x}=0}$. So,
$$2^{-|S|}\prod_{i\in S}\partial_{x_i}^2\ f\cdot g\big|_{\vec{x}=0}=c^{d-|S|}\sum_{R:R\supseteq S}\mu(R)\prod_{i\in S}\partial_{x_i}\det\left(\lambda I+\sum_{j=1}^m x_jA_j\right)|_{\vec{x}=0}.$$

Therefore,

$$\begin{aligned}\prod_{i=1}^{m}(1-\partial_{x_i}^2)f\cdot g\Big|_{\vec{x}=0} &= \sum_S(-1)^{|S|}\prod_{i\in S}\partial_{x_i}^2 f\cdot g\Big|_{\vec{x}=0}\\ &= c^d\sum_S(-1)^{|S|}\cdot\Big(\frac{2}{c}\Big)^{|S|}\Big(\sum_{R:R\supseteq S}\mu(R)\Big)\Big(\prod_{i\in S}\partial_{x_i}\det\Big(\lambda I+\sum_{j=1}^m x_jA_j\Big)|_{\vec{x}=0}\Big).\end{aligned}$$

Note that the LHS of this equality is equal to the RHS of the statement, and the RHS of this equality is equal to the LHS of the statement by plugging in $x_i = -2/c$ in Equation 17.1 for $1 \le i \le m$. □

Using Theorem 17.3 and Corollary 16.8, we have the following formula for the expected characteristic polynomial for the volume measure.

Problem 17.4. *Let $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$ with $\sum_{i=1}^m v_iv_i^T = I_n$ and let μ be its volume measure as in Definition 16.5. Then, for any $c \in \mathbb{R}$,*

$$c^{d-n}\sum_{S\subseteq[m]}\mu(S)\det\Big(c\lambda I-\sum_{i\in S}2v_iv_i^T\Big)=\prod_{i=1}^m\Big(1-\partial_{x_i}^2\Big)\det\Big(cI+\sum_{i=1}^m x_iv_iv_i^T\Big)\cdot\det\Big(\lambda I+\sum_{i=1}^m x_iv_iv_i^T\Big)\Big|_{\vec{x}=0}.$$

Interlacing Family

Once the formula in Theorem 17.3 is established, the proof that the family of polynomials $\{\det(xI-\sum_{i\in S}v_iv_i^T)\}_{S\in\mathrm{supp}(\mu)}$ forms an interlacing family is similar to that of Theorem 13.28. First, using that g_μ is real stable as μ is strongly Rayleigh, we can prove that the expected characteristic polynomial is real-rooted.

Exercise 17.5 (Expected Characteristic Polynomial is Real-Rooted)**.** *The expected characteristic polynomial $\mathbb{E}_{S\sim\mu}[\det(xI-\sum_{i\in S}v_iv_i^T)]$ is real-rooted for any strongly Rayleigh distribution μ.*

Then, using a tree with depth m where each internal node has at most 2 children, and associating each non-leaf node with the conditional expected polynomial, we can use a similar argument as in Theorem 13.28 to establish an interlacing family for strongly Rayleigh measure.

Problem 17.6 (Interlacing Family of Strongly Rayleigh Measure)**.** *Let $\mu : \{0,1\}^m \to \mathbb{R}$ be a strongly Rayleigh measure with homogeneous generating polynomial g_μ. The set of all polynomials in $\{\det(xI-\sum_{i\in S}v_iv_i^T)\}_{S\in\mathrm{supp}(\mu)}$ form an interlacing family, where the root polynomial can be chosen to be $\mathbb{E}_{S\sim\mu}[\det(xI-\sum_{i\in S}v_iv_i^T)]$.*

Multivariate Barrier Argument

The second step is to upper bound the maximum root of the expected characteristic polynomial. The proof structure is similar to that in the induction hypothesis in Definition 15.6. The same multivariate barrier functions $\Phi_p^i(y) = \partial_{x_i}p(y)/p(y)$ as in Definition 15.5 are used. Starting with a point $(t,\ldots,t)$ which is above the roots of the multivariate polynomial $g_\mu(\lambda\vec{1}+x)\cdot\det(\lambda I+\sum_{i=1}^m x_iv_iv_i^T)$, Anari and Oveis Gharan proved that $(t+\delta,\ldots,t+\delta,t,\ldots,t)$ with the first k coordinates being $t+\delta$ for a small δ is still above the roots after applying the differential operator $(1-\partial_{x_i}^2)$ for $1 \le i \le k$.

Also, the monotoncity and the convexity of the barrier functions in Proposition 15.7 are important in the analysis.

The main difference is that the differential operator $(1-\partial_{x_i}^2)$ is different. So, not only they need to keep track of $\Phi_p^i(y) = \partial_{x_i} p(y)/p(y)$, but also the second derivative $\Psi_p^i(z) = \partial_{x_i}^2 p(y)/p(y)$ as well. They prove a new lemma that $\partial_{x_i}\Psi_p^j(y)/\partial_{x_i}\Phi_p^j(y) \leq 2\Phi_p^j(y)$, also using the result that a bivariate real-stable polynomial can be written as $\det(x_1A + x_2B + C)$ for $A, B \succcurlyeq 0$ and C Hermitian. As in Chapter 15, the assumptions $\Pr(i \in S) \leq \epsilon_1$ and $\|v_j\|_2^2 \leq \epsilon_2$ are (only) used in the computation of the initial value of the barrier functions. Because of the differential operator $(1-\partial_{x_i}^2) = (1+\partial_{x_i})(1-\partial_{x_i})$, they could prove that the shift δ is much smaller as $(1-\partial_{x_i})$ shifts the root up while $(1+\partial_{x_i})$ shifts the roots down, hence getting a final bound that is much smaller than that in Theorem 15.2.

We refer the reader to [AO14] for details. It would be very nice if one could strengthen their result to prove Weaver's Conjecture 15.1.

Thin Tree and Asymmetric Traveling Salesman Problem

The main motivation of their work is the thin tree problem and its application to the asymmetric traveling salesman problem (ATSP).

Definition 17.7 (Thin Tree)**.** *Given an undirected graph $G = (V, E)$ and $0 < \alpha < 1$, we say a spanning tree T is α-thin if $|\delta_T(S)| \leq \alpha \cdot |\delta_G(S)|$ for all $S \subseteq V$. In words, a spanning tree is α-thin if it uses at most α-fraction of edges in every cut.*

There is a strong conjecture about the existence of a thin tree.

Conjecture 17.8 (Goddyn's Conjecture)**.** *Every k-edge-connected graph has a $O(\frac{1}{k})$-thin tree.*

If the conjecture is true and a $O(\frac{1}{k})$-thin tree can be found in polynomial time, then it would imply a constant factor approximation algorithm for ATSP [AGM+17].

It can be proved that a random spanning tree is a $O\big(\frac{\log n}{\log\log n}\cdot\frac{1}{k}\big)$-thin tree [AGM+17]. The argument is similar to that in cut sparsification, using Chernoff bound and careful union bound. The reason that we can apply Chernoff bound is that the edges in a random spanning tree are negatively associated as shown in Chapter 16.

As in graph sparsification, one can define a stronger spectral notion of a thin tree.

Definition 17.9 (Spectrally Thin Tree)**.** *Given an undirected graph $G = (V, E)$ and $0 < \alpha < 1$, we say a spanning tree T is α-spectrally-thin if $L(T) \preccurlyeq \alpha \cdot L(G)$, where $L(T)$ and $L(G)$ are the Laplacian matrices of T and G respectively.*

Exercise 17.10 (Spectral Thin Tree is Combinatorially Thin Tree)**.** *Prove that an α-spectrally thin tree is also an α-thin tree.*

One advantage of this stronger notion is that it is easier to work with. For example, given a tree, it is easy to check whether it is α-spectrally thin, while it is not known how to check whether it is (combinatorially) α-thin. Moreover, the solution to the Weaver's conjecture in Corollary 15.3 provides a non-trivial sufficient condition for the existence of a spectrally thin tree.

Proposition 17.11 (Sufficient Condition for Spectrally Thin Tree [HO14])**.** *If the maximum effective resistance of an edge in a graph G is α, then G has a $O(\alpha)$-spectrally thin tree.*

The solution to Weaver's conjecture in Corollary 15.3 implies that if the maximum effective resistance of an edge in G is α, then the edge set of G can be partitioned into two subgraphs H_1 and H_2 such that for $i \in \{1,2\}$,

$$\frac{1}{2}\big(1-\sqrt{2\alpha}\big)L_G \preccurlyeq L_{H_i} \preccurlyeq \frac{1}{2}\big(1+\sqrt{2\alpha}\big)L_G.$$

The idea in Proposition 17.11 is to recursively apply this partitioning in each subgraph (with slightly weaker bound on the maximum effective resistance of an edge) until we cannot apply again, by that time there will be $O(\frac{1}{\alpha})$ edge-disjoint subgraphs of G, each is connected and $O(\alpha)$-spectrally thin.

This gives hope that the techniques developed in the method of interlacing family of polynomials can be used to prove Goddyn's conjecture. Proposition 17.11 gives us a spectrally thin tree, which is combinatorially thin, but it requires the assumption that the maximum effective resistance of an edge is small, which is not necessarily satisfied in a k-edge-connected graph. The breakthrough by Anari and Oveis Gharan [AO15], in a high level, is to reduce the combinatorial problem to the spectral problem, and use Theorem 17.1 to prove the following result.

Theorem 17.12 (Anari, Oveis Gharan [AO15]). *Every k-edge-connected graph has a $O\big(\log\log n \cdot \frac{1}{k}\big)$-thin tree.*

The reduction, however, is very complicated and technically challenging. Also, there is now a constant factor approximation algorithm for ATSP. So we just highlight some underlying mathematics of the thin tree result. First, check that the probabilistic formulation for strongly Rayleigh measure can be used to prove Proposition 17.11 without using recursion.

Exercise 17.13 (Direct Proof of Sufficient Condition). *Show that Theorem 17.1 can be used to prove Proposition 17.11. You may use the fact that the probability that an edge is in a uniform spanning tree is equal to the effective resistance of an edge, i.e.* $\Pr_T[e \in T] = \mathrm{Reff}(e)$.

The main advantage of Theorem 17.1 is that the output is guaranteed to be a spanning tree, so that we get connectivity for free (without worrying about the minimum eigenvalue), which is important for the thin tree problem. The following is the fundamental building block of their approach.

Theorem 17.14 (Spectral Thin Tree from Subgraph). *Given a graph $G = (V,E)$ and a subset of edges $F \subseteq E$ such that (V,F) is k-edge-connected, if $\mathrm{Reff}(e) \le \epsilon$ for all $e \in F$, then G has a $O\big(\frac{1}{k}+\epsilon\big)$-spectrally thin tree in F.*

Proof Idea: Since F is k-edge-connected, there are at least $k/2$ edge-disjoint spanning trees by Tutte or Nash-Williams' theorem. This implies that there is a point in the spanning tree polytope with maximum edge value $O(1/k)$. By expressing this point using "maximum entropy distribution", it can be proved that there is a weighting of the edges, so that the weighted random spanning tree distribution μ (which is still homogeneous strongly Rayleigh by Exercise 16.11) has maximum marginal probability of an edge $O(1/k)$, i.e. $\epsilon_1 = O(1/k)$ in Theorem 17.1. The assumption about effective resistance implies that $\epsilon_2 \le \epsilon$, and so Theorem 17.1 can be applied. □

With Theorem 17.14, their strategy is to add "short-cut" edges in the graph, so that they don't change the cut structures much, while creating many edges with small effective resistance. They do it in $O(\log\log n)$ iterations so that the edges with small effective resistance form a k-edge-connected subgraph. The most difficult step is to prove the existence of good short-cut edges, which they proved by using an involved analysis of a semidefinite program. They managed to prove Theorem 17.12 after 80 pages of work after Theorem 17.14.

17.2 Interlacing Family for Matrix Discrepancy

The result by Marcus, Spielman, and Srivastava on Weaver's conjecture can be interpreted as an improved discrepancy bound over the matrix Chernoff bound in Theorem 9.13. Taking this perspective, Kyng, Luh, and Song [KLS20] considered the following more refined matrix concentration result.

Theorem 17.15 (Matrix Concentration with Variance [Tro12]). *Let $\xi_i \in \{\pm 1\}$ be independent random signs, and let $A_1, \dots, A_m \in \mathbb{R}^{n\times n}$ be symmetric matrices. Let $\sigma^2 = \|\sum_{i=1}^m \mathrm{var}[\xi_i] A_i^2\|$. Then,*

$$\Pr\left(\left\|\sum_{i=1}^m \mathbb{E}\left[\xi_i\right] A_i - \sum_{i=1}^m \xi_i A_i\right\| \geq t\cdot\sigma\right) \leq 2ne^{-\frac{t^2}{2}}.$$

This theorem implies that with high probability the discrepancy is at most $O(\sqrt{\log n})\cdot\sigma$. The main result of Kyng, Luh, and Song is to prove that there exists a signing with a stronger discrepancy bound.

Theorem 17.16 (Matrix Discrepancy of Rank One Matrices [KLS20]). *Consider any independent scalar random variables $\xi_1, \dots, \xi_m$ with finite support. Let $u_1, \dots, u_m \in \mathbb{R}^n$ and*

$$\sigma^2 = \left\|\sum_{i=1}^m \mathrm{var}[\xi_i](u_i u_i^T)^2\right\|.$$

Then there exists a choice of outcomes $\epsilon_1, \dots, \epsilon_m$ in the support of $\xi_1, \dots, \xi_m$ such that

$$\left\|\sum_{i=1}^m \mathbb{E}\left[\xi_i\right] u_i u_i^T - \sum_{i=1}^m \epsilon_i u_i u_i^T\right\| \leq 4\sigma.$$

Note that if $\|u_i\|_2^2 \leq \epsilon$ and $\sum_{i=1}^m u_i u_i^T = I_n$, then $\sigma^2 \leq \epsilon$, and the conclusion is that there is a signing $\epsilon_1, \dots, \epsilon_m \in \{\pm 1\}$ with $\left\|\sum_{i=1}^m \epsilon_i u_i u_i^T\right\| \leq O(\sqrt{\epsilon})$. Check that this matches the result of Marcus, Spielman and Srivastava in Corollary 15.3 applied to the same setting (i.e. with $\{\pm 1\}$ instead of $\{0,1\}$). Theorem 17.16 is more flexible that allow arbitrary biased ± 1 random variables, instead of only zero mean random variables. Also, Theorem 17.16 is more refined in that it proves stronger bounds when there are only a few vectors with $\|u_i\|_2^2 = \epsilon$ while other vectors are much shorter.

Two-Sided Spectral Rounding

One interesting application of Theorem 17.16 is the two-sided spectral rounding problem from Chapter 11.

Problem 17.17 (Two-Sided Spectral Rounding). *Let $v_1, ..., v_m \in \mathbb{R}^n$ and $x \in [0,1]^m$. Suppose $\sum_{i=1}^m x_i v_i v_i^T = I_n$ and $\|v_i\|_2^2 \leq \epsilon$ for all $i \in [m]$. Prove that there exists a subset $S \subseteq [m]$ satisfying*

$$\left(1 - O(\sqrt{\epsilon})\right)\cdot I_n \preccurlyeq \sum_{i\in S} v_i v_i^T \preccurlyeq \left(1 + O(\sqrt{\epsilon})\right)\cdot I_n.$$

This result can be slightly extended to incorporate one non-negative linear constraint, which has some applications in network design.

Theorem 17.18 (Two-Sided Spectral Rounding with Costs [LZ20]). *Let $v_1, ..., v_m \in \mathbb{R}^n$, $x \in [0,1]^m$ and $c \in \mathbb{R}^m_{\geq 0}$. Suppose $\sum_{i=1}^m x_i v_i v_i^T = I_n$, $\|v_i\|_2^2 \leq \epsilon$ for all $i \in [m]$ and $c_\infty \leq \epsilon^2 \langle c, x \rangle$. Then there exists $z \in \{0,1\}^m$ such that*

$$\big(1-O(\sqrt{\epsilon})\big)\cdot I_n \preccurlyeq \sum_{i=1}^m z_i v_i v_i^T \preccurlyeq \big(1+O(\sqrt{\epsilon})\big)\cdot I_n \quad \text{and} \quad \big(1-O(\sqrt{\epsilon})\big)\cdot\langle c,x\rangle \leq \langle c,z\rangle \leq \big(1+O(\sqrt{\epsilon})\big)\cdot\langle c,x\rangle.$$

Proof Ideas

Theorem 17.16 needs to bound the maximum eigenvalue and the minimum eigenvalue of the difference. Their main idea is to consider the polynomial

$$\det\left(x^2 I - \Big(\sum_{i=1}^m \xi_i u_i u_i^T\Big)^2\right) = \det\left(xI - \sum_{i=1}^m \xi_i u_i u_i^T\right)\cdot \det\left(xI + \sum_{i=1}^m \xi_i u_i u_i^T\right)$$

Note that the largest root of this polynomial is

$$\lambda_{\max}\left(\det\left(x^2 I - \Big(\sum_{i=1}^m \xi_i u_i u_i^T\Big)^2\right)\right) = \left\|\sum_{i=1}^m \xi_i u_i u_i^T\right\|.$$

They found a nice formula for the expected characteristic polynomial.

Proposition 17.19 (Expected Characteristic Polynomial for Matrix Discrepancy). *Let $u_1, \ldots, u_m \in \mathbb{R}^n$. Consider independent random variables ξ_i with means μ_i and variances γ_i^2. Let $Q \in \mathbb{R}^{m\times m}$ be a symmetric matrix. Then*

$$\begin{aligned}&\mathbb{E}_\xi\left[\det\left(x^2 I - \Big(Q + \sum_{i=1}^m (\xi_i-\mu_i)u_i u_i^T\Big)^2\right)\right]\\ =\ &\prod_{i=1}^m\left(1-\frac{1}{2}\partial_{z_i}^2\right)\det\left(xI - Q + \sum_{i=1}^m z_i\gamma_i u_i u_i^T\right)\cdot\det\left(xI + Q + \sum_{i=1}^m z_i\gamma_i u_i u_i^T\right)\Big|_{z_1=\ldots=z_m=0}.\end{aligned}$$

This formula is obtained inductively by the following lemma, as in the inductive proof of the multilinear formula in Subsection 13.4.

Problem 17.20 (Expected Characteristic Polynomial after One Step). *For positive semidefinite matrices $M, N \in \mathbb{R}^{m\times m}$, $v \in \mathbb{R}^m$ and ξ a random variable with zero mean and variance γ^2,*

$$\mathbb{E}_\xi\left[\det\left(M-\xi vv^T\right)\cdot\det\left(N+\xi vv^T\right)\right] = \left(1-\frac{1}{2}\frac{d^2}{dt^2}\right)\det\left(M+t\gamma vv^T\right)\det\left(N+t\gamma vv^T\right)\Big|_{t=0}.$$

Proposition 17.19 implies that the expected characteristic polynomial is real-rooted. Then, using a similar argument as in Theorem 13.28 and Problem 17.6, one can prove that the set of all possible characteristic polynomials form an interlacing family. Therefore, by the probabilistic method in Theorem 12.12, there exists a choice of outcomes $\epsilon_1, \ldots, \epsilon_m$ in the finite support of $\xi_1, \ldots, \xi_m$ such that

$$\left\|\sum_{i=1}^m \epsilon_i u_i u_i^T - \sum_{i=1}^m \mathbb{E}\left[\xi_i\right] u_i u_i^T\right\| \leq \lambda_{\max}\left(\mathbb{E}_{\xi_1,\ldots,\xi_m}\left[\det\left(x^2 I - \Big(\sum_{i=1}^m (\xi_i - \mathbb{E}\left[\xi_i\right])u_i u_i^T\Big)^2\right)\right]\right).$$

Then the second step is to bound the maximum root of the expected polynomial. The differential operator $1 - \partial_{z_i}^2$ in Proposition 17.19 is the same as in the formula for strongly Rayleigh measure in Theorem 17.3. It turned out that the same setup (including the induction hypothesis and the multivariate barrier functions) and many calculations in [AO14] can be reused. The base case is different, and once again the assumptions are (only) used in computing the initial values of the barrier functions.

17.3 Interlacing Family for Higher Rank Matrices

The interlacing families that we have seen so far all involve sum of rank one matrices. It was remarked that the same approach would fail spectacularly for sum of higher rank matrices, e.g. the multilinear formula in Theorem 13.20 does not hold and the expected characterisitic polynomials may not even be real-rooted anymore. Cohen [Coh16] found a very clever solution to bypass expected characteristic polynomials and proved the following generalization of Theorem 15.2.

Theorem 17.21 (Cohen [Coh16]). *Let $A_1, \dots, A_m \in \mathbb{R}^{n\times n}$ be independent random positive semidefinite matrices with finite support such that*

$$\mathbb{E}\left[\sum_{i=1}^m A_i\right] = I_n \quad \text{and} \quad \mathbb{E}\big[\operatorname{Tr}(A_i)\big] \le \epsilon \text{ for } 1 \le i \le m.$$

Then

$$\Pr\left[\lambda_{\max}\left(\sum_{i=1}^m A_i\right) \le \left(1+\sqrt{\epsilon}\right)^2\right] > 0.$$

The insight of Cohen is to focus on the RHS of the multilinear formula, the mixed characteristic polynomial in Definition 13.21. The following is a generalization with a "non-mixed" matrix M, where the mixed characteristic polynomial in Definition 13.21 is a special case with $M = 0$.

Definition 17.22 (Generalized Mixed Characteristic Polynomial). *The generalized mixed characteristic polynomial of $n \times n$ matrices $M, B_1, \dots, B_m$ (not necessarily rank-one) is defined as*

$$\mu[M; B_1, \dots, B_m](\lambda) = \prod_{i=1}^m (1 - \partial_{x_i}) \det\left(\lambda I - M + \sum_{i=1}^m x_i \cdot B_i\right)\Bigg|_{x_1=x_2=\cdots=x_m=0}$$

The multivariate barrier argument in Chapter 15 proved that

$$\lambda_{\max}\Big(\mu\big[\mathbb{E}\left[A_1\right], \dots, \mathbb{E}\left[A_m\right]\big](\lambda)\Big) = \lambda_{\max}\left(\prod_{i=1}^m (1-\partial_{x_i}) \det\left(\lambda I + \sum_{i=1}^m x_i \cdot \mathbb{E}\left[A_i\right]\right)\Bigg|_{\vec{x}=0}\right) \le (1+\sqrt{\epsilon})^2.$$

Cohen's proof has two steps. The first step is to prove that the set of all possible mixed characteristic polynomials form an interlacing family and so the probabilistic method in Theorem 12.12 applies.

Problem 17.23 (Interlacing Family for Mixed Characteristic Polynomials). *Let $A_1, \dots, A_m \in \mathbb{R}^{n\times n}$ be independent random positive semidefinite matrices, where each A_j has k possibilities $M_{j,1}, \dots, M_{j,k}$. Prove that the set of all k^m possible mixed characteristic polynomials $\mu[M_{1,i_1}, \dots, M_{m,i_m}](\lambda)$ where each $1 \le i_j \le k$ for $1 \le j \le m$ form an interlacing family, and the root polynomial can be chosen to be $\mu\big[\mathbb{E}\left[A_1\right], \dots, \mathbb{E}\left[A_m\right]\big]$. Conclude that there exists a choice $M_j \in \operatorname{supp}(A_j)$ for $1 \le j \le m$ such that*

$$\lambda_{\max}\Big(\mu\big[M_1, \dots, M_m\big](\lambda)\Big) \le \lambda_{\max}\Big(\mu\big[\mathbb{E}\left[A_1\right], \dots, \mathbb{E}\left[A_m\right]\big](\lambda)\Big).$$

The second step is to prove that the maximum root of the characteristic polynomial can only be smaller than the maximum root of the mixed characteristic polynomial.

Proposition 17.24 (Maximum Root of Mixed Characteristic Polynomials). *For any positive semidefinite matrices $M_1, \dots, M_m$,*

$$\lambda_{\max}\Big(\det\Big(\lambda I - \sum_{i=1}^{m} M_i\Big)\Big) \le \lambda_{\max}\Big(\mu\big[M_1, \dots, M_m\big](\lambda)\Big)$$

The proof of Proposition 17.24 is by applying the following lemma repeatedly, where we move the mixed part to the non-mixed part one at a time. The proof of the following lemma nicely uses the convexity of real-stable polynomials for points above the roots.

Lemma 17.25 (Inductive Step for Proposition 17.24). *Let $M, M_1, \dots, M_m \in \mathbb{R}^{n\times n}$ be positive semidefinite matrices. Then*

$$\lambda_{\max}\Big(\mu\big[M + M_m; M_1, \dots, M_{m-1}\big](\lambda)\Big) \le \lambda_{\max}\Big(\mu\big[M; M_1, \dots, M_m\big](\lambda)\Big)$$

Proof Sketch: Consider the bivariate polynomial

$$p(\lambda, x) := \prod_{i=1}^{m-1}(1 - \partial_{x_i}) \det\left(\lambda I - M + xM_m + \sum_{j=1}^{m-1} x_i M_i\right)\Big|_{x_1=\dots=x_{m-1}=0}.$$

Note that

$$p(\lambda, -1) = \mu\big[M + M_m; M_1, \dots, M_{m-1}\big](\lambda) \quad \text{and} \quad (1 - \partial_x)p(\lambda, x)|_{x=0} = \mu\big[M; M_1, \dots, M_m\big](\lambda).$$

Let λ^* be the maximum root of $p(\lambda, -1)$. Note that both $p(\lambda, -1)$ and $(1 - \partial_x)p(\lambda, x)|_{x=0}$ are real-rooted with positive leading coefficients. So, to prove the statement of the lemma, it suffices to prove that $(1 - \partial_x)p(\lambda^*, x)|_{x=0} \le 0$.

Using the result that any bivariate real-stable polynomial $p(x_1, x_2)$ can be written as $\pm\det(x_1 A + x_2 B + C)$ for some $A, B \succcurlyeq 0$ and some symmetric C (or the complex analysis argument in Tao's blogpost), it can be shown that the roots of the univariate polynomial $p_x(\lambda)$ can only decrease when we increase x. This implies that the point $(\lambda^*, -1)$ is above the roots of $p(\lambda, x)$. Since $(\lambda^*, -1)$ is above the roots of $p(\lambda, x)$, the point $(\lambda^*, 0)$ is also above the roots of $p(\lambda, x)$. Again, by the result of bivariate real-stable polynomial, it follows that $p(\lambda^*, x)$ is convex along the interval $x \in [-1, 0]$, which implies that

$$p(\lambda^*, 0) - p(\lambda^*, -1) \le \partial_x p(\lambda^*, 0) \quad \Longrightarrow \quad (1 - \partial_x)p(\lambda^*, x)|_{x=0} \le p(\lambda^*, -1) = 0.$$

□

More Results

The interlacing family for permutations in [MSS18] and the interlacing family for the paving problem in [RL20] are also very interesting.

17.4 References

[AGM+17] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An $O(\log n/\log\log n)$-approximation algorithm for the asymmetric traveling salesman problem. *Oper. Res.*, 65(4):1043–1061, 2017. 164

[AO14] Nima Anari and Shayan Oveis Gharan. The Kadison-Singer problem for strongly Rayleigh measures and applications to asymmetric TSP. *arXiv preprint arXiv:1412.1143*, 2014. 161, 164, 168

[AO15] Nima Anari and Shayan Oveis Gharan. Effective-resistance-reducing flows, spectrally thin trees, and asymmetric TSP. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 20–39. IEEE, 2015. 165

[Coh16] Michael Cohen. Talk of "Improved spectral sparsification and kadison-singer for sums of higher-rank matrices" at Banff. 2016. https://www.birs.ca/events/2016/5-day-workshops/16w5111/videos/watch/201608011534-Cohen.html. 168

[HO14] Nicholas J. A. Harvey and Neil Olver. Pipage rounding, pessimistic estimators and matrix concentration. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 926–945. SIAM, 2014. 164

[KLS20] Rasmus Kyng, Kyle Luh, and Zhao Song. Four deviations suffice for rank 1 matrices. *Advances in Mathematics*, 375:107366, 2020. 166

[LZ20] Lap Chi Lau and Hong Zhou. A spectral approach to network design. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 826–839. ACM, 2020. 109, 110, 114, 167

[MSS18] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families IV: bipartite Ramanujan graphs of all sizes. *SIAM J. Comput.*, 47(6):2488–2509, 2018. 142, 169

[RL20] Mohan Ravichandran and Jonathan Leake. Mixed determinants and the Kadison–Singer problem. *Mathematische Annalen*, 377(1):511–541, 2020. 169

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012. 93, 166

Chapter 18

Real-Stability and Log-Concavity

In this chapter, we study the pioneer work of Gurvits on using real stable polynomials in combinatorial problems that inspired the recent developments, and we see some applications of real-stable polynomials in designing approximation algorithms for combinatorial optimization problems. This concludes the second part of the course about real stable polynomials, and connects to the third part of the course that involves log-concave polynomials. Our presentation follows closely that of the course notes by Oveis Gharan [Ove20].

18.1 Gurvits' Capacity Inequality

An influential concept defined by Gurvits is the capacity function [Gur04, Gur06].

Definition 18.1 (Capacity of a Polynomial). *Let $p \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial. The capacity of p is defined as*

$$\mathrm{cap}(p) := \inf_{x>0} \frac{p(x_1, \ldots, x_n)}{x_1 \cdots x_n}.$$

The main theorem in [Gur06] is to use the capacity of a real-stable polynomial p to estimate the coefficient of the monomial $x_1 \cdots x_n$ in p.

Theorem 18.2 (Gurvits [Gur06]). *Let $p \in \mathbb{R}[x_1, \ldots, x_n]$ be a real-stable polynomial with non-negative coefficients. Then*

$$e^{-n} \cdot \mathrm{cap}(p) \leq \partial_{x_1} \cdots \partial_{x_n} p\big|_{x=0} \leq \mathrm{cap}(p).$$

The proof of the second inequality is easy and holds for any non-negative polynomial. The proof of the first inequality is deferred to the next section after we introduced log-concavity.

The optimization problem in the capacity function can be formulated as a convex program, which can be solved in polynomial time using only a value oracle.

Proposition 18.3 (Computing Capacity). *Given a real stable polynomial p with non-negative coefficients, and an oracle that for any $x \in \mathbb{R}^n$ returns the value $p(x)$, there is an algorithm to compute the capacity of p up to $(1+\epsilon)$ factor in time* $\mathrm{poly}\left(\langle p \rangle, 1/\epsilon\right)$, *where $\langle p \rangle := n + \deg(p) + |\log c_{\max}| + |\log c_{\min}|$ and $c_{\max}, c_{\min}$ are defined as the maximum and minimum coefficients in p.*

Proof Sketch: The idea is to do a change of variables to replace x_i by e^{y_i}, which is valid since $x > 0$. Then consider the logarithm of the objective function

$$\log \operatorname{cap}(p) = \inf_{y \in \mathbb{R}^n} \left\{ \log \left(p\big(e^{y_1}, \ldots, e^{y_n}\big) \right) - \sum_{i=1}^{n} y_i \right\}.$$

Since all the coefficients of p are non-negative, the term $\log \left(p\big(e^{y_1}, \ldots, e^{y_n}\big) \right)$ can be written as $\log \sum_i a_i e^{\langle b_i, y \rangle}$ where $a_i \geq 0$ and $b_i \in \mathbb{R}^n$ for all i (one term for each monomial). This is known as the log-sum-exponential function, which is a convex function in y (often used as a soft-max function in convex optimization).

Using the ellipsoid method to compute $\log \operatorname{cap}(p)$, one can implement the separation oracle using only a value oracle and analyze the time complexity by bounding the volumes of the outer ellipsoid and the inner ellipsoid. See [AO17] for the details. □

Permanent

Gurvits [Gur06] used Theorem 18.2 to approximate the permanent of a matrix and to give a beautiful proof of the Van der Waerden's conjecture.

Definition 18.4 (Permanent of a Matrix)**.** *The permanent of a matrix $A \in \mathbb{R}^{n \times n}$ is defined as*

$$\operatorname{per}(A) = \sum_{\sigma \in S^n} \prod_{i=1}^{n} A_{i,\sigma(i)},$$

where the summation runs over the set of all permutations of n elements.

Gurvits' idea is to read the permanent of a matrix from the following real-stable polynomial.

Exercise 18.5 (Permanent Polynomial)**.** *Given a non-negative matrix $A \in \mathbb{R}^{n \times n}_{\geq 0}$, the permanent polynomial is defined as*

$$p_A(x_1, \ldots, x_n) := \prod_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} \cdot x_j.$$

Show that p_A is a homogeneous real-stable polynomial with non-negative coefficients and $\operatorname{per}(A) = \partial_{x_1} \cdots \partial_{x_n} p_A(x_1, \ldots, x_n)\big|_{x=0}$.

It follow from Theorem 18.2 and Exercise 18.5 that there is a determininstic e^n-approximation algorithm for the permanent of a non-negative matrix. The best known deterministic approximation ratio is $(\sqrt{2})^n$ by Anari and Rezaei [AR19]. A major breakthrough in approximate counting is a randomized $(1+\epsilon)$-approximation algorithm for the permanent of a non-negative matrix by Jerrum, Sinclair and Vigoda [JSV04] with running time polynomial in n and $1/\epsilon$. Their method is to use Markov chains to sample a uniform random perfect matching of the bipartite graph of the input matrix. It has been a long standing open question to match this result with a deterministic algorithm.

Van der Waerden conjectured that the permanent of an $n \times n$ non-negative doubly stochastic matrix is at least e^{-n}. This conjecture was proven in the 80s by Gyires and by Egorychev and Falikman. Gurvits provided a simple and elegant proof using Theorem 18.2 and the AM-GM inequality (which follows from the concavity of the logarithmic function).

Exercise 18.6 (Weighted AM-GM Inequality). *Let $a_1, \ldots, a_n \geq 0$ and $\mu_1, \ldots, \mu_n \geq 0$ with $\sum_{i=1}^n \mu_i = 1$. Then*

$$\sum_{i=1}^n \mu_i a_i \geq \prod_{i=1}^n a_i^{\mu_i}.$$

Theorem 18.7 (Van der Waerden's Conjecture). *The permanent of any non-negative doubly stochastic matrix $A \in \mathbb{R}^{n \times n}$ is at least e^{-n}, where a matrix is called doubly stochastic if every row sum and every column sum is equal to 1.*

Proof. Let p_A be the permanent polynomial in Exercise 18.5. By Exercise 18.5 and Theorem 18.2,

$$\operatorname{per}(A) = \partial_{x_1} \cdots \partial_{x_n} p_A(x_1, \ldots, x_n)\big|_{x=0} \geq e^{-n} \cdot \operatorname{cap}(p_A),$$

and so the statement would follow if we could prove that $\operatorname{cap}(p_A) \geq 1$ for any non-negative doubly stochastic matrix A. For any $x \in \mathbb{R}^n_+$,

$$p_A(x_1, \ldots, x_n) = \prod_{i=1}^n \sum_{j=1}^n A_{i,j} x_j \geq \prod_{i=1}^n \prod_{j=1}^n x_j^{A_{i,j}} = \prod_{j=1}^n x_j^{\sum_{i=1}^n A_{i,j}} = \prod_{j=1}^n x_j,$$

where the inequality is by the weighted AM-GM inequality in Exercise 18.6 and the assumption that every row sum is equal to one, and the last equality is by the assumption that every column sum is equal to one. This implies that $\operatorname{cap}(p_A) \geq 1$ and completes the proof. □

Gurvits' result can also be used to give a simple proof of the following bound by Schrijver, whose original proof is combinatorial and highly complicated.

Problem 18.8 (Schrijver's Bound). *Let G be a k-regular bipartite graph with n vertices on each side. Prove that the number of perfect matchings in G is at least $\left(\frac{k}{e}\right)^n$.*

18.2 Log-Concavity

We follow the proof of Theorem 18.2 by Oveis Gharan [Ove20] that highlights the role of log-concavity. The following simple lemma about univariate polynomials will be used in the base case.

Lemma 18.9 (Log-Concavity of Non-Negative Real-Rooted Polynomial). *Let $f \in \mathbb{R}[x]$ be a real-rooted polynomial with non-negative coefficients then $\log f$ is a concave function on $\mathbb{R}_{\geq 0}$.*

Proof. Let $\alpha_1, \ldots, \alpha_n$ be the roots of f. As all coefficients of f are non-negative, it follows that $f(x) > 0$ for all $x > 0$ as long as $f \not\equiv 0$, and so all the roots of f must be non-positive. Assume without loss that the leading coefficient of f is one, then

$$\log f = \log\Big(\prod_{i=1}^n (x - \alpha_i)\Big) = \sum_{i=1}^n \log(x - \alpha_i).$$

Since $\alpha_i \leq 0$ for $1 \leq i \leq n$, we conclude that $\log f$ is concave on $\mathbb{R}_{\geq 0}$ as each $\log(x - \alpha_i)$ is well-defined and concave on $\mathbb{R}_{\geq 0}$ and the sum of concave functions is a concave function. □

The following is Theorem 18.2 in the univariate case.

Lemma 18.10 (Univariate Case of Theorem 18.2)**.** *For any real-rooted polynomial $f \in \mathbb{R}[x]$ with non-negative coefficients,*

$$f'(0) \geq \frac{1}{e} \inf_{x>0} \frac{f(x)}{x}.$$

Proof. If $f(0) = 0$, then $f'(0) = \inf_{x>0} \frac{f(x)}{x}$ as $f(x)$ is a convex function in x, and so the inequality holds trivially. Henceforth we assume that $f(0) > 0$. By log-concavity of f from Lemma 18.9, for any $x \geq 0$,

$$\log f(x) \leq \log f(0) + x(\log f(0))' \quad \Longrightarrow \quad \log \frac{f(x)}{x} \leq \log f(0) + x\frac{f'(0)}{f(0)} - \log x$$

The RHS is minimized when $x = f(0)/f'(0)$, and this implies that

$$\inf_{x>0} \log \frac{f(x)}{x} \leq \log f(0) + 1 - \log \frac{f(0)}{f'(0)} \quad \Longrightarrow \quad 1 + \log f'(0) \geq \inf_{x>0} \log \frac{f(x)}{x},$$

which implies the lemma. □

We are ready to prove Theorem 18.2.

Proof of Theorem 18.2. The proof is by induction on the number of variables n. Let

$$q(x_1, \ldots, x_{n-1}) := \partial_{x_n} p|_{x_n=0}.$$

Note that q is real-stable as differentiation and substituting real number preserve real stability by Exercise 13.17 and Proposition 13.13, and also q has non-negative coefficients as p has.

For any $x_1, \ldots, x_{n-1} > 0$, consider the univariate polynomial $f(x_n) := p(x_1, \ldots, x_{n-1}, x_n)$. Note that f is real-stable and thus real-rooted, and also $f'(0) = q(x_1, \ldots, x_{n-1})$. Applying Lemma 18.10 on f,

$$q(x_1, \ldots, x_{n-1}) = f'(0) \geq \frac{1}{e} \inf_{x_n>0} \frac{f(x_n)}{x_n} = \frac{1}{e} \inf_{x_n>0} \frac{p(x_1, \ldots, x_n)}{x_n}.$$

Using this and applying the induction hypothesis on q, we conclude that

$$\begin{aligned} \partial_{x_1} \cdots \partial_{x_n} p\big|_{x_1=\ldots=x_n=0} &= \partial_{x_1} \cdots \partial_{x_{n-1}} q\big|_{x_1=\ldots=x_{n-1}=0} \\ &\geq e^{-n-1} \inf_{x_1,\ldots,x_{n-1}>0} \frac{q(x_1, \ldots, x_{n-1})}{x_1 \cdots x_{n-1}} \geq e^{-n} \inf_{x_1,\ldots,x_n>0} \frac{p(x_1, \ldots, x_n)}{x_1 \cdots x_n}. \end{aligned}$$

□

Finally, we prove the following generalization of Lemma 18.9, which can be seen as a generalization of the well-known fact that $\det(X)$ is log-concave over the space of positive semidefinite matrices. This result will be useful in the next section for designing approximation algorithms.

Theorem 18.11 (Log-Concavity of Homogeneous Real-Stable Polynomials)**.** *Let $p \in \mathbb{R}[x_1, \ldots, x_n]$ be a homogeneous real-stable polynomial with non-negative coefficients. Then p is log-concave on $\mathbb{R}^n_{\geq 0}$.*

Proof. To prove the statement, we will prove that $\log p(a+tb)$ is concave along the line $a+tb$, for any $a \in \mathbb{R}^n_+$ and $b \in \mathbb{R}_n$ such that $a+tb \in \mathbb{R}^n_+$ for all $t \in [0,1]$. Let p be homogeneous of degree k. Then

$$p(a+tb) = p\Big(t\Big(\frac{a}{t}+b\Big)\Big) = t^k \cdot p\Big(\frac{a}{t}+b\Big).$$

Since p is real-stable and $a \in \mathbb{R}^n_+$, $p(at+b)$ is real rooted. Let $p(at+b) = c\prod_{i=1}^k (t-\lambda_i)$ where $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ are the roots. Then $p(\frac{a}{t}+b) = c\prod_{i=1}^k (\frac{1}{t}-\lambda_i)$, and so

$$p(a+tb) = t^k \cdot p\Big(\frac{a}{t}+b\Big) = c\prod_{i=1}^k (1-t\lambda_i)$$

Note that $\lambda_i < 1$ for $1 \le i \le k$, as otherwise there exists $t \in [0,1]$ such that $p(a+tb) = 0$, contradicting to our assumption that the line $a+tb \in \mathbb{R}^n_+$ for $t \in [0,1]$ and so $p(a+tb) > 0$ as p has non-negative coefficients. Therefore,

$$\log p(a+tb) = \log(c) + \sum_{i=1}^k \log(1-t\lambda_i),$$

which is a concave function as each $\log(1-t\lambda_i)$ is a concave function of t for $t \in [0,1]$ when $\lambda_i < 1$. □

18.3 Determinant Maximization

In this section, we see some applications of Gurvits' capacity inequality in Theorem 18.2 and the log-concavity of real-stable polynomial in designing approximation algorithms for some combinatorial optimization problems.

The determinant maximization problem is closely related to the D-design problem that we have discussed in Section 11.1.

Definition 18.12 (Determinant Maximization Problem)**.** *Given a positive semidefinite matrix $M \in \mathbb{R}^{n\times n}$ and an integer k, the goal is to output a set $S \subseteq [n]$ with $|S| = k$ that maximizes* $\det(M_{S,S})$.

Oveis Gharan [Ove20] give a simple proof of the following result by Nikolov [Nik15] using the theory of real-stable polynomials.

Theorem 18.13 (Nikolov [Nik15])**.** *There is a polynomial time algorithm that gives a e^{-k} approximation to the determinant maximization problem.*

Proof Sketch: Consider the following mathematical program for the problem:

$$\begin{aligned} \max \quad & \log \sum_{S\subseteq[n]:|S|=k} \det(M_{S,S}) \prod_{i\in S} x_i \\ \text{subject to} \quad & \sum_{i=1}^n x_i = k \\ & x_i \ge 0 \quad \text{for } 1 \le i \le n. \end{aligned}$$

First we argue that the program is convex and can be solved in polynomial time. We know from Problem 16.15 that the polynomial $\sum_{S\subseteq[n]:|S|=k} \det(M_{S,S}) \prod_{i\in S} x_i$ is real-stable when $M \succcurlyeq 0$. As

this polynomial is homogeneous with non-negative coefficients as $M \succcurlyeq 0$, it follows from Theorem 18.11 that the objective function is a concave function, and thus the program is a convex program. We remark that the proof of Problem 16.15 also provides a compact representation of the polynomial in the objective function, so that we can evaluate the objective function in time polynomial in terms of the size of the input M.

Note that the convex program is a relaxation of the determinant maximization problem, and so its objective value $\sum_{S\subseteq[n]:|S|=k} \det(M_{S,S}) \prod_{i\in S} x_i$ is at least the optimal value opt. Given an optimal solution x to the convex program, we consider the following simple randomized rounding algorithm. Let μ be the distribution on $[n]$ where $\mu(i) = x_i/k$. Choose k indexes $i_1, i_2, \ldots, i_k \in [n]$, where each index i_j is sampled from μ independently. If $i_1, \ldots, i_k$ are all distinct, then we output $S = \{i_1, \ldots, i_k\}$, otherwise we output "failed". Then, the expected objective value of the output is

$$\mathbb{E}[\mathsf{alg}] = \sum_{S\in\binom{[n]}{k}} \Pr[S \text{ is sampled}] \cdot \det(M_{S,S}) = \sum_{S\in\binom{[n]}{k}} \left(k! \cdot \prod_{i\in S} \frac{x_i}{k}\right) \cdot \det(M_{S,S}) \geq e^{-k} \cdot \mathsf{opt},$$

where the second equality is because there are $k!$ permutations to choose the same subset S, each permutation with probability $\prod_{i\in S} \frac{x_i}{k}$. This bounds the integrality gap of the convex program, but note that the randomized rounding algorithm is not a polynomial time algorithm (see Exercise 18.14). Nikolov [Nik15] derandomized this analysis using conditional expectation to give a deterministic polynomial time algorithm with the same guarantee. □

Exercise 18.14 (Exponential Running Time)**.** *Show an example where the randomized rounding algorithm in the proof of Theorem 18.13 runs in time $\Omega(e^k)$.*

Problem 18.15 (Sampling by Volume)**.** *Suppose there is a polynomial time algorithm that outputs a random size-k subset S with probability proportional to $\det(M_{S,S})$. Show that this algorithm can be used to give a randomized polynomial time e^{-k}-approximation algorithm for the determinant maximization problem.*

Determinant Maximization with Partition Constraints

Nikolov and Singh [NS16] considered the generalization of the determinant maximization problem with partition constraints. The following is a simple version of their problem.

Definition 18.16 (Determinant Maximization with Partition Constraints)**.** *Given a positive semidefinite matrix $M \in \mathbb{R}^{n\times n}$, an integer k and a partition of the ground set $[n]$ into $P_1 \cup P_2 \cup \ldots \cup P_k$, the goal is to output a set S with $|S \cap P_i| = 1$ for $1 \leq i \leq k$ that maximizes $\det(M_{S,S})$.*

We briefly discuss the main ingredients in [NS16]. The natural relaxation in the proof of Theorem 18.13 (with suitable modification) has unbounded integrality gap. The key contribution by Nikolov and Singh is to come up with a very interesting convex relaxation for the problem. Let $\mathcal{B} := \{S \subseteq [n] \mid |S \cap P_i| = 1 \; \forall 1 \leq i \leq k\}$ be the set of subsets that satisfy the partition constraints.

Write $M = V^T V$ and let v_i be the i-th column of V. The relaxation in [NS16] is

$$\begin{aligned} \mathsf{opt} \quad &= \quad \sup_x \inf_y \det\Big(\sum_{i=1}^n x_i y_i v_i v_i^T\Big) \\ \text{subject to} \quad & \sum_{j\in P_i} x_j = 1 \qquad \forall 1 \le i \le k \\ & 0 \le x_j \le 1 \qquad \forall 1 \le j \le n, \\ & \prod_{i\in S} y_i = 1 \qquad \forall S \in \mathcal{B}. \end{aligned}$$

Nikolov and Singh showed that it is indeed a relaxation and it can be solved in polynomial time. They analyzed the simple rounding algorithm where we choose one vector in P_i with probability distribution $\{x_j\}_{j\in P_i}$. The analysis uses the real-stability of the polynomial $p(y) := \det(\sum_{i=1}^n x_i y_i v_i v_i^T)$. They reduced the problem of bounding the expected value of the output to bounding the coefficient of the monomial $z_1 \dots z_m$ of a related real-stable polynomial, and then they applied Gurvits' inequality in Theorem 18.2 to prove that the convex program gives a e^{-k}-approximation to the problem.

Generalized Permanent Inequality

Extending the convex program in [NS16], Anari and Oveis Gharan [AO17] obtained an elegant generalization of Gurvits' permanent inequality. The following is a simpler version of their main theorem.

Theorem 18.17 (Generalized Permanent Inequality [AO17]). *For any two multi-linear real-stable polynomial $p, q \in \mathbb{R}[x_1, \dots, x_n]$ with non-negative coefficients,*

$$\sup_{\alpha\ge 0} \inf_{x,y>0} e^{-\alpha} \frac{p(x)q(y)}{\big(xy/\alpha\big)^\alpha} \le \sum_\kappa c_p(\kappa)\cdot c_q(\kappa) \le \sup_{\alpha\ge 0} \inf_{x,y>0} \frac{p(x)q(y)}{\big(xy/\alpha\big)^\alpha},$$

where $\alpha, x, y \in \mathbb{R}^n$ are vectors, and κ is over the set of monomials of p and q with coefficients $c_p(\kappa)$ and $c_q(\kappa)$ respectively. For two vectors $a, b \in \mathbb{R}^n$, $ab \in \mathbb{R}^n$ denotes the vector with the i-th entry being $a_i b_i$, and $a^b \in \mathbb{R}$ denotes the number $\prod_{i=1}^n a_i^{b_i}$.

They showed that this result generalizes Gurvits' inequality and Nikolov-Singh's result, and provides deterministic polynomial time algorithms for several counting problems.

Concluding Remark

Besides the applications we discussed, Gurvits' concept of capacity has found important applications in analyzing scaling problems including matrix scaling, frame scaling and operator scaling, and there are various applications of these scaling problems (see e.g. [GGOW20]).

18.4 References

[AO17] Nima Anari and Shayan Oveis Gharan. A generalization of permanent inequalities and applications in counting and optimization. In *Proceedings of the 49th Annual ACM*

SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017, pages 384–396. ACM, 2017. 172, 177

[AR19] Nima Anari and Alireza Rezaei. A tight analysis of bethe approximation for permanent. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1434–1445. IEEE Computer Society, 2019. 172

[GGOW20] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. Operator scaling: theory and applications. *Foundations of Computational Mathematics*, 20(2):223–290, 2020. 177

[Gur04] Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, 2004. 171

[Gur06] Leonid Gurvits. Hyperbolic polynomials approach to Van der Waerden/Schrijver-Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 417–426, 2006. 171, 172

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004. 53, 172

[Nik15] Aleksandar Nikolov. Randomized rounding for the largest simplex problem. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 861–870. ACM, 2015. 175, 176

[NS16] Aleksandar Nikolov and Mohit Singh. Maximizing determinants under partition constraints. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 192–201. ACM, 2016. 176, 177

[Ove20] Shayan Oveis Gharan. Course notes of "Polynomial paradigm in algorithm design", lecture 3,6. 2020. https://homes.cs.washington.edu/~shayan/courses/polynomials/. 128, 130, 153, 158, 171, 173, 175, 179, 187, 211

Chapter 19

High Dimensional Expanders

We begin the third part of the course about high dimensional expanders and log-concave polynomials. In this chapter, we study a definition of high dimensional expanders through local spectral expansion. We will introduce the necessary concepts and then prove a fundamental result by Oppenheim called the trickling down theorem. See [Ove20, HL21] for similar notes which are used in the preparation of this chapter.

High Dimensional Expanders

As we have seen in Chapter 7, expander graphs have nice combinatorial, probabilistic, and algebraic properties, which is an important reason that there is a rich theory with connections and applications in diverse areas. While it may be easy to generalize the definition of expander graphs to higher dimensions for some properties (e.g. combintorial expansion in hypergraphs), it is not easy to find a definition in higher dimension that generlizes all nice properties of expander graphs. There are various definitions of high dimensional expanders, some using concepts from algebraic topology; see [Lub18] for a survey with motivations and applications.

In this course, the main application of high dimensional expanders is in analyzing mixing time of Markov chains. We will study a more recent and elementary definition developed in [KM17, DK17, KO20], which was motivated by the study of random walks.

19.1 Simplicial Complexes

A simplicial complex is a high dimensional generalization of a graph.

Definition 19.1 (Simplicial Complex)**.** *A set system is a pair $X = (U, \mathcal{F})$ with U as the ground set and $\mathcal{F}$ is a set of subsets of U. A simplicial complex is a set system that is downward closed, such that if $\tau \in \mathcal{F}$ and $\sigma \subset \tau$, then $\sigma \in \mathcal{F}$.*

We follow the convention of using Greek letters $\sigma, \tau, \eta, \alpha, \beta$ for subsets in $\mathcal{F}$. The following are some basic definitions about simplicial complexes.

Definition 19.2 (Face, Dimension, Pure Simplicial Complex)**.** *Any subset $\sigma \in \mathcal{F}$ is called a face of the simplicial complex $X = (U, \mathcal{F})$. A face σ is of dimension k if its size is $|\sigma| = k + 1$, e.g. a 0-dimensional face is a singleton (a vertex), a 1-dimensional face is a pair (an edge), a 2-dimensional face is a triple, etc. Given a simplicial complex $X = (U, \mathcal{F})$, we use $X(k)$ to denote the set of*

faces of dimension k. A simplicial complex is d-dimensional if the maximum face size is $d+1$. A d-dimensional simplicial complex is pure if every maximal face is of size $d+1$.

Simplicial complex is a very general definition. We can associate a simplicial complex to many classes of combinatorial objects.

Example 19.3 (Simplicial Complex from Spanning Trees)**.** *Given a graph $G = (V, E)$, we can define a simplicial complex $X = (E, \mathcal{F})$ where the ground set in X is the edge set E of G. A subset of edges $E' \subseteq E$ is in $\mathcal{F}$ if and only if E' forms an acyclic subgraph in G. It should be clear that X is a pure simplicial complex. When G is connected, the maximal faces correspond to spanning tree, which are of size $|V| - 1$ and so X is $(|V| - 2)$-dimensional.*

More generally, every matroid naturally corresponds to a simplicial complex.

Example 19.4 (Simplicial Complex from Matroids)**.** *A matroid $M = (U, \mathcal{I})$ is a set system where U is the ground set and $\mathcal{I}$ is the set of subsets of U which satisfies the following two properties:*

1. *$\mathcal{I}$ is downward close, i.e. $S \in \mathcal{I}$ and $T \subseteq S$ implies $T \in \mathcal{I}$.*
2. *If $S, T \in \mathcal{I}$ and $|T| > |S|$, then there exists $x \in T \setminus S$ such that $S \cup \{x\} \in \mathcal{I}$.*

So, by (1), $M = (U, \mathcal{I})$ is a simplicial complex. And, by (2), $M = (U, \mathcal{I})$ is a pure simplicial complex. The sets in $\mathcal{I}$ are usually called the independent sets, and the maximal sets are usually called bases. It is not difficult to check that the simplicial complex from spanning trees is a matroid.

A more general example is the class of linear matroids. Given a matrix $A \in \mathbb{F}^{m \times n}$, the linear matroid of A is defined as $M = ([n], \mathcal{I})$ where the ground set $[n]$ is the set of columns of A, and a subset S of columns is in $\mathcal{F}$ if and only if the columns in S are linearly independent. We leave it as an exercise to check that it is a matroid and includes the matroid from spanning trees as a special case.

There are many more simplicial complexes that one can define, e.g. simplicial complexes from cliques of graphs, simplicial complexes for graph coloring, etc. We may discuss some of these in later chapters.

Weighted Simplicial Complexes

We will consider pure simplicial complexes with weights on its faces. We follow the convention in [DDFH18] that the weights form a probability distribution on the faces of the same dimension.

Definition 19.5 (Weighted Simplicial Complexes)**.** *A weighted pure simplicial complex (X, Π) is a pure simplicial complex with a probability distribution Π on the faces of maximal dimension.*

In applications of random sampling, the probability distribution in the maximal faces are usually the uniform distribution, so they are simply unweighted simplicial complexes, but the following definition of induced distributions will be important in our study. An alternative way to think about these induced distributions is to think of them as weighted degrees of a subset in the simplicial complex.

Definition 19.6 (Induced Distributions)**.** *Given a d-dimensional weighted pure simplicial complex (X, Π), a probability distribution Π_k on $X(k)$ for $0 \le k \le d$ is defined inductively as follows. The base case is $\Pi_d = \Pi$. For $d-1 \ge k \ge 0$, the probability distribution $\Pi_k : X(k) \to \mathbb{R}$ is defined by considering the marginal distributions such that*

$$\Pi_k(\alpha) = \frac{1}{k+2} \sum_{\beta \in X(k+1) : \beta \supset \alpha} \Pi_{k+1}(\beta) \tag{19.1}$$

for each face $\alpha \in X(k)$. Equivalently, we can understand Π_k as the probability distribution of the following random process. Sample a random face $\beta \in X(d)$ using the probability distribution Π_d, and then sample a uniform random subset α of β in $X(k)$, so that

$$\Pi_k(\alpha) = \frac{1}{\binom{d+1}{k+1}} \sum_{\beta \in X(d) : \beta \supset \alpha} \Pi_d(\beta) = \frac{1}{\binom{d+1}{k+1}} \Pr_{\beta \sim \Pi_d}[\beta \supset \alpha].$$

We will often drop the subscript about the dimension of the face. Just keep in mind that each Π_k is a probability distribution.

19.2 Local Spectral Expanders

We first define links and graphs of simplicial complexes, and then define local spectral expanders.

Links

The following is the key definition that enables a local-to-global approach for simplicial complexes.

Definition 19.7 (Links)**.** *Let $X = (U, \mathcal{F})$ be a simplicial complex. For a face $\alpha \in \mathcal{F}$, the link X_α is defined as*

$$X_\alpha := \{\beta \setminus \alpha \mid \beta \in \mathcal{F}, \beta \supset \alpha\}.$$

In words, X_α is defined by the faces τ that can be used to extend α such that $\alpha \cup \tau \in \mathcal{F}$.

If X is a pure d-dimensional simplicial complex and $\alpha \in X(k)$, then X_α is a pure $(d-|\alpha|)$-dimensional simplicial complex (where the empty set is a face of dimension -1). In the spanning tree complex $X = (E, \mathcal{I})$, given a subset of acyclic edges $F \in \mathcal{I}$, the link X_F is defined such that a subset of edges F' is a face in X_F if and only if $F \cup F'$ is an acyclic subgraph. In matroid terminology, the link X_F is obtained by "contracting" the elements in F. A general approach to study a simplicial complex is to decompose it into its links, as we will see later in this chapter.

The probability distributions $\Pi_0, \dots, \Pi_d$ on X in Definition 19.6 can be used to define $\Pi^\alpha_0, \dots, \Pi^\alpha_{d-k-1}$ on X_α using conditional probability, where $\Pi^\alpha(\tau) \propto \Pr_{\beta \sim \Pi_d}[\beta \supset \tau \mid \beta \supset \alpha]$

Definition 19.8 (Induced Distributions on Links)**.** *Let (X, Π) be a d-dimensional weighted pure simplicial complex. For any face α and any $\tau \in X_\alpha$,*

$$\Pi^\alpha(\tau) := \Pr_{\beta \sim \Pi_{|\tau|+|\alpha|-1}} \left[\beta = \alpha \cup \tau \mid \beta \supset \alpha\right] = \frac{\Pi(\alpha \cup \tau)}{\sum_{\beta : |\beta| = |\tau|+|\alpha|, \beta \supset \alpha} \Pi(\beta)} = \frac{\Pi(\alpha \cup \tau)}{\binom{|\alpha \cup \tau|}{|\alpha|} \cdot \Pi(\alpha)}, \tag{19.2}$$

where the last equality follows from Definition 19.6.

Often, it is enough to understand that $\Pi^\alpha(\tau) \propto \Pi(\alpha \cup \tau)$, and just see the denominator in Definition 19.8 as a normalizing constant.

Exercise 19.9. *Verify that Π_k^α is a probability distribution for every $0 \le k \le d - |\alpha|$.*

Skeletons and Graphs

Definition 19.10 (k-Skeletons)**.** *Given $X = (U, \mathcal{F})$, the k-skeleton of X is the simplicial complex $X_k = (U, \mathcal{F}_k)$ where $\mathcal{F}_k$ is the set of faces of $\mathcal{F}$ with dimension at most k. When there are weights on the faces in $\mathcal{F}$, we use the same weight on the faces in $\mathcal{F}_k$.*

The special case of 1-skeleton will be of particular interest, which could be thought of as the underlying graph of the simplicial complex.

Definition 19.11 (Graph of Links)**.** *For a link X_α, the graph $G_\alpha = (X_\alpha(0), X_\alpha(1), \Pi_1^\alpha)$ is defined as the 1-skeleton of X_α. More explicitly, each singleton $\{v\}$ in X_α is a vertex v in G_α, each pair $\{u, v\}$ in X_α is an edge uv in G_α, and the weight of uv in G_α is equal to $\Pi_1^\alpha(\{u, v\})$.*

A simple observation is that if X is a pure d-dimensional simplicial complex and Π is the uniform distribution on $X(d)$, then for any $\alpha \in X(d-2)$ the weighting Π_1^α on the edges of G_α is uniform. We will use this simple observation later.

Random Walk Matrices

The definition of local spectral expanders will be based on the random walk matrices of the links.

Definition 19.12 (Random Walk Matrix of a Link)**.** *Given the graph $G_\alpha = (X_\alpha(0), X_\alpha(1), \Pi_1^\alpha)$ of a link X_α, let A_α be the adjacency matrix of G_α and let D_α be the diagonal degree matrix where*

$$D_\alpha(x,x) = \sum_{y \in X_\alpha(0)} A_\alpha(x,y) = \sum_{y \in X_\alpha(0)} \Pi_1^\alpha(\{x,y\}) = 2\Pi_0^\alpha(\{x\}).$$

Check that the last equality follows from Definition 19.6 and Definition 19.8. The random walk matrix W_α of G_α is defined as

$$W_\alpha := D_\alpha^{-1} A_\alpha \quad \text{where} \quad W_\alpha(x,y) = \frac{\Pi_1^\alpha(\{x,y\})}{2\Pi_0^\alpha(\{x\})} = \frac{\Pi(\alpha \cup \{x,y\})}{(|\alpha|+2) \cdot \Pi(\alpha \cup \{x\})} \text{ for all } \{x,y\} \in X_\alpha(1).$$

Check that the last equality follows from Definition 19.8. Note that the distribution Π_0^α is the stationary distribution of W_α as

$$(\Pi_0^\alpha)^T W_\alpha = (\Pi_0^\alpha)^T D_\alpha^{-1} A_\alpha = \frac{1}{2}(\vec{1})^T A_\alpha = (\Pi_0^\alpha)^T.$$

Recall from Chapter 6 that the random walk matrix and the normalized adjacency matrix of a graph are similar matrices, and so the eigenvalues of the random walk matrices are real. The largest eigenvalue of W_α is one and the all-one vector is a corresponding eigenvector.

Local Spectral Expanders

Finally, we can state the definition of high dimensional expanders that we will use.

Definition 19.13 (Local Spectral Expanders [KM17, DK17, KO20])**.** *Let (X, Π) be a pure d-dimensional simplicial complex. We say (X, Π) is a γ-local-spectral expander if $\lambda_2(W_\alpha) \leq \gamma$ for all faces $\alpha \in X$, where $\lambda_2(W_\alpha)$ is the second largest eigenvalue of the random walk matrix W_α.*

More generally, given $\gamma_{-1}, \ldots, \gamma_{d-2}$, we say (X, Π) is a $(\gamma_{-1}, \ldots, \gamma_{d-2})$-local-spectral expander if $\lambda_2(W_\alpha) \leq \gamma_k$ for all faces $\alpha \in X(k)$ for all $-1 \leq k \leq d-2$.

The definitions in [KM17, DK17] require a lower bound on the minimum eigenvalue of W_α as well. The above definition is from [KO20] where Kaufman and Oppenheim realized that only upper bounding λ_2 is enough for fast mixing of higher order random walks that we will define in the next chapter.

We can understand the above definition as requiring the "local" random walks in each link graph are fast mixing. As the random walk matrix has the same spectrum as the normalized adjacency matrix, we can also understand that the above definition as requiring the "local" weighted graphs of the links have large edge conductance through Cheeger's inequality.

Example 19.14 (Complete Complex)**.** *Consider the complete complex $X_d = (U, \mathcal{F}_d)$ where every subset $S \subseteq U$ with $|S| \leq d+1$ is in $\mathcal{F}_d$, equipped with the uniform distribution on the faces of dimension d. Then the graph of every link of dimension k is an unweighted complete graph with $d-k$ vertices, with second largest eigenvalue of the random walk matrix being $-1/(d-k-1)$.*

It is not surprising that a complete complex is a good high dimensional expander (if not, what is?). As in expander graphs in Chapter 7, the goal is usually to construct high dimensional expanders with few maximal faces. Unlike in the graph case, however, random simplicial complexes are *not* high dimensional expanders with high probability. It is difficult to construct sparse high-dimensional expanders, with only a few known algebraic constructions [Lub18]. This is a topic of great interest but is out of the scope of this course.

19.3 Oppenheim's Trickling Down Theorem

To show that a simplicial complex is a γ-local-spectral expander, we need to bound the second largest eigenvalue of the random walk matrix for every link up to dimension $d-2$. In applications where the goal is to do uniform sampling of the maximal faces, it is usually much easier to work with the random walk matrix of the "top" links of dimension $d-2$, because the graphs of these links are unweighted as we mentioned before. For "lower" links, just determining the edge weights may already involve some difficult counting problems. So, it would be very nice if we could bound the second largest eigenvalues of the lower links by the second largest eigenvalues of the top links, and Oppenheim's trickling down theorem [Opp18] provides such a general bound for any pure simplicial complex.

Theorem 19.15 (Oppenheim's Trickling Down Theorem [Opp18])**.** *Let (X, Π) be a pure d-dimensional weighted simplicial complex where Π satisfies Equation 19.1 and Equation 19.2. Suppose the graph $G_\emptyset = (X(0), X(1), \Pi_1)$ is connected and $\lambda_2(W_v) \leq \gamma$ for all $v \in X(0)$. Then*

$$\lambda_2(W_\emptyset) \leq \frac{\gamma}{1-\gamma}.$$

Note that the condition that the graph $G_\emptyset$ is connected is necessary, as the example of two disjoint cliques shows. Applying Theorem 19.15 inductively would give us the following bound.

Exercise 19.16 (Oppenheim's Bound [Opp18]). *Let (X, Π) be a pure d-dimensional weighted simplicial complex where Π satisfies Equation 19.1 and Equation 19.2. If G_α is connected for every $\alpha \in X(k)$ for every $k \leq d-2$, then for any $-1 \leq j \leq d-2$,*

$$\gamma_j \leq \frac{\gamma_{d-2}}{1-(d-2-j)\gamma_{d-2}}.$$

In general, when $\gamma_{d-2} > 0$, the bound deteriorates as we go to lower links. But if we could prove that $\gamma_{d-2} \leq 0$, then Oppenheim's bound in Exercise 19.16 would allow us to conclude that the simplicial complex is a 0-local-spectral expander, which is almost as strong as the complete complex in Example 19.14. An important example of 0-local-spectral expander is the matroid complex in Example 19.4.

Matroid Complex

The following result is proved in [ALOV19], as an important step in proving the matroid expansion conjecture that we will explain in the next chapter.

Theorem 19.17 (Matroid Complex is 0-Local-Spectral Expander [ALOV19]). *The simplicial complex of any matroid in Example 19.4 with the uniform distribution on the maximal faces is a 0-local-spectral expander.*

Proof. Let X be a pure d-dimensional simplicial complex from a matroid M. By Oppenheim's bound in Exercise 19.16, we just need to prove that the graph of every link is connected and the second largest eigenvalue of the random walk matrix of the links of diemnsion $d-2$ is at most 0.

The first claim that the graph of every link is connected follows from the second axiom of matroids stated in Example 19.4, and is left as a simple exercise.

For the second claim, we first consider the adjacency matrix A_α of a face α of dimension $d-2$. Since the probability distribution on the maximal faces is the uniform distribution, every non-zero entry of the adjacency matrix has the same weight. For bounding the spectrum, without loss of generality, we rescale the matrix such that $A_\alpha(i,j) = 1$ if $\alpha \cup \{i,j\}$ is a maximal face and $A_\alpha(i,j) = 0$ otherwise. We would like to argue that A_α has at most one positive eigenvalue, and this would imply that the normalized adjacency matrix $\mathcal{A}_\alpha$ has at most one positive eigenvalue by the Courant-Fischer Theorem 2.12, and this would imply that the random walk matrix W_α has at most one positive eigenvalue as W_α and $\mathcal{A}_\alpha$ are similar matrices.

To argue that A has at most one positive eigenvalue, let us start with the spanning tree complex in Example 19.3. In the spanning tree complex $X = (E, \mathcal{F})$ of a graph $G = ([n], E)$, the maximal faces are of size $n-1$ and thus of dimension $d := n-2$. Given a face $F \subseteq E$ of dimension $d-2$, with $|F| = n-3$, the subgraph formed by the edges in F has exactly three components left. Note that the edges remained in the link X_F are the edges with endpoints in different components. Two edges e, f in X_F form a face of size 2 if and only if $F \cup \{e,f\}$is a spanning tree if and only if e and f are not parallel edges if we contract the three components into single vertices. In other words, the edges in X_F can be partitioned into three equivalent classes E_1, E_2, E_3 such that two edges e, f form a face of size 2 in X_F if and only if they do not belong to the same subset. So, the adjacency matrix A_F can be written as $J - \chi_{E_1}\chi_{E_1}^T - \chi_{E_2}\chi_{E_2}^T - \chi_{E_3}\chi_{E_3}^T$, where J is the all-one matrix and

χ_{E_i} is the characteristic vector of E_i for $1 \le i \le 3$. Therefore, A_F is a rank-one matrix minus three positive semidefinite matrices. It follows from Courant-Fischer Theorem 2.12 or Cauchy interlacing Theorem 2.13 that A_F has at most one positive eigenvalue, and this concludes the proof for spanning tree complexes.

The same proof works for linear matroids, where two columns i, j form a face of size 2 if and only if they are parallel in the linear algebraic sense, and so again the columns can be partitioned into equivalence classes $E_1, E_2, \ldots, E_l$ (with l not necessarily equal to 3) so that $A = J - \sum_{i=1}^{l} \chi_{E_i} \chi_{E_i}^T$. In general, this holds for arbitrary matroids and is known as the matroid partition property and so the same proof works. □

The proof can be generalized so that the probability distribution on the maximal faces are product distributions.

Exercise 19.18 (Product Distributions). *Let $X = (E, \mathcal{I})$ be a matroid complex. Suppose each element $e \in E$ has a weight w_e. Consider the probability distribution Π where each maximal face F has probability $\Pi(F)$ proportional to $\prod_{e \in F} w_e$. Prove that (X, Π) is still a 0-local-spectral expander.*

One may wonder what are other 0-local-spectral expanders. The following problem shows that they have very restrictive structures such that the graphs of the top links must be complete multipartite graphs.

Problem 19.19 (Complete Multi-Partite Graphs). *The adjacency matrix of a graph G has at most one positive eigenvalue if and only if G is a complete multi-partite graph.*

19.4 Garland's Method

The main goal in this section is to prove Oppenheim's Theorem 19.15. We will first prepare by introducing different inner products for the calculations in the proof. Then we will introduce the Garland's method which decomposes a structure of a simplicial complex to the corresponding structure of its links. Then we will present the proof of Oppenheim's theorem.

Inner Products and Rayleigh Quotients

Recall from Chapter 6 that the random walk matrix of a graph and the normalized adjacency matrix of a graph are similar matrices, and so the eigenvalues are real, but the eigenvectors may not be orthonormal using the standard inner product. It will be convenient to work with a different inner product so that the eigenvectors are orthonormal with respect to this inner product. Given a random walk matrix $W = D^{-1}A$, we have shown in Lemma 6.18 that the eigenvectors $u_1, \ldots, u_n \in \mathbb{R}^n$ satisfies $\langle u_i, u_j \rangle_D := \sum_{l=1}^{n} D(l,l) \cdot u_i(l) \cdot u_j(l) = 0$. For simplicial complexes, we will use the probability distribution Π_0 to define the inner product, which is equivalent to the degree distribution as shown in Definition 19.12.

Definition 19.20 (Inner Products using Π). *Given a simplicial complex (X, Π), for two functions $f, g : X(0) \to \mathbb{R}$, define*

$$\langle f, g \rangle_{\Pi_0} := \mathbb{E}_{i \sim X(0)} \big[f(i) g(i) \big] = \sum_{i \in X(0)} \Pi_0(i) f(i) g(i).$$

Similarly, given a link X_α and two functions $f, g : X_\alpha(0) \to \mathbb{R}$, define $\langle f, g\rangle_{\Pi_0^\alpha} := \mathbb{E}_{i\sim X_\alpha(0)}[f(i)g(i)]$. Note that W_α is self-adjoint with respect to this inner product, as

$$\langle f, W_\alpha g\rangle_{\Pi_0^\alpha} = \langle f, D_\alpha^{-1}A_\alpha g\rangle_{\Pi_0^\alpha} = \frac{1}{2}\langle f, A_\alpha g\rangle = \frac{1}{2}\langle A_\alpha f, g\rangle = \langle W_\alpha f, g\rangle_{\Pi_0^\alpha}. \tag{19.3}$$

Check that this implies that all eigenvalues of W_α are real with corresponding eigenvectors orthonormal with respect to this inner product.

We also define Rayleigh quotients using the inner product in Definition 19.20.

Definition 19.21 (Rayleigh Quotients using Π)**.** *Given a simplicial complex (X, Π) and a link (X_α, Π^α), for a function $g : X_\alpha(0) \to \mathbb{R}$, the Rayleigh quotient of g is defined as*

$$\frac{\langle g, W_\alpha g\rangle_{\Pi_0^\alpha}}{\langle g, g\rangle_{\Pi_0^\alpha}}.$$

Check that there is a one-to-one correspondence between the Rayleigh quotients of W_α and the Rayleigh quotients of $\mathcal{A}_\alpha$ defined as $f^T\mathcal{A}_\alpha f/f^T f$, and in particular the second largest eigenvalue of W_α can be characterized as

$$\lambda_2(W_\alpha) = \max_{g:\langle g,\vec{1}\rangle_{\Pi_0^\alpha}=0} \frac{\langle g, W_\alpha g\rangle_{\Pi_0^\alpha}}{\langle g, g\rangle_{\Pi_0^\alpha}}. \tag{19.4}$$

The advantage of working with W_α (instead of A_α or $\mathcal{A}_\alpha$) is that we know that the vector $\vec{1}/\|1\|_{\Pi_0^\alpha}$ is an eigenvector of W_α with eigenvalue 1 for every link α. Let $u_1, \ldots, u_n$ be the eigenvectors of W_α that are Π_0^α-orthonormal. Given any $y \in \mathbb{R}^n$, note that we can write $y = c_1u_1 + \ldots + c_nu_n$ with $c_i = \langle y, u_i\rangle_{\Pi_0^\alpha}$, and in particular

$$c_1 = \langle y, u_1\rangle_{\Pi_0^\alpha} = \frac{\langle y, \vec{1}\rangle_{\Pi_0^\alpha}}{\|\vec{1}\|_{\Pi_0^\alpha}}. \tag{19.5}$$

Garland's Method

Our plan is to bound the second largest eigenvalue of W using the Rayleigh quotient formulation in Equation 19.4. Garland's method is a well-known technique in high dimensional expanders that decompose a term (Rayleigh quotient in this case) into the corresponding terms over the links, so that we can apply the properties (second largest eigenvalue in this case) in the links to bound the terms over the links in order to bound the original term. We first define a notation for the localization of a function into a link.

Definition 19.22 (Localization to Link)**.** *Given a simplicial complex (X, Π), a function $f : X(k) \to \mathbb{R}$ and a face τ, the localization of f to $X_\tau(k)$ is defined as $f_\tau : X_\tau(k) \to \mathbb{R}$ such that $f_\tau(\sigma) = f(\sigma)$ for all $\sigma \in X_\tau(k)$.*

The following two lemmas show how to decompose the denominator and the numerator of Equation 19.4 respectively.

Lemma 19.23 (Decomposition of Denominator)**.** *Given a simplicial complex (X, Π), for any two functions $f, g : X(0) \to \mathbb{R}$,*

$$\langle f, g\rangle_{\Pi_0} = \mathbb{E}_{v\sim\Pi_0}\left[\langle f_v, g_v\rangle_{\Pi_0^v}\right],$$

where f_v is the localization of f to $X_v(0)$ as defined in Definition 19.22.

Proof. The proof is by showing that the distribution Π_0 can be written as $\mathbb{E}_{v\sim\Pi_0}[\Pi_0^v]$ by using conditional probability. Note that $\langle f, g\rangle_{\Pi_0} = \sum_{w\in X(0)} \Pi_0(w) f(w) g(w)$ and

$$\mathbb{E}_{v\sim\Pi_0}\left[\langle f_v, g_v\rangle_{\Pi_0^v}\right] = \sum_{v\in X(0)} \Pi_0(v) \sum_{w\in X_v(0)} \Pi_0^v(w) f_v(w) g_v(w) = \sum_{w\in X(0)} \left(\sum_{v\in X(0)} \Pi_0(v)\tilde{\Pi}_0^v(w)\right) f(w)g(w)$$

where $\tilde{\Pi}_0^v$ is just an extension of Π_0^v (with zero entries) so that it has the same dimension as Π_0. We prove the statement by showing that $\Pi_0(w) = \sum_{v\in X(0)} \Pi_0(v)\cdot\tilde{\Pi}_0^v(w)$ as

$$\Pi_0(w) = \frac{1}{2}\sum_{v:\{w,v\}\in X(1)} \Pi_1(\{w,v\}) = \sum_{v:\{w,v\}\in X(1)} \Pi_0(v)\cdot\Pi_0^v(w) = \sum_{v\in X(0)} \Pi_0(v)\cdot\tilde{\Pi}_0^v(w).$$

where the first equality is by Equation 19.1 and the second equality is by Equation 19.2.

An alternative succinct way [Ove20] to write the above proof is

$$\mathbb{E}_{w\sim\Pi_0}\left[f(w)g(w)\right] = \mathbb{E}_{vw\in\Pi_1}\mathbb{E}_{w|\{v,w\}}\left[f(w)g(w)\right] = \mathbb{E}_{v\in\Pi_0}\mathbb{E}_{\{v,w\}|v}\left[f(w)g(w)\right] = \mathbb{E}_{v\sim\Pi_0}\left[\langle f_v, g_v\rangle_{\Pi_0^v}\right],$$

where in the first equality to sample a random vertex w we choose a random pair $\{v,w\}$ and then drop a random vertex with probability $1/2$, in the second equality we use an equivalent process of first choosing a random vertex v then choose a random edge $\{v,w\}$ incident on it and then choose the other vertex w. □

Lemma 19.24 (Decomposition of Numerator)**.** *Given a simplicial complex (X,Π), for two functions $f, g : X(0)\to\mathbb{R}$,*

$$\langle f, Wg\rangle_{\Pi_0} = \mathbb{E}_{v\sim\Pi_0}\left[\langle f_v, W_v g_v\rangle_{\Pi_0^v}\right],$$

where W and W_v are the random walk matrices of the empty link $X_\emptyset$ and the link X_v respectively.

Proof. The proof is by showing that the adjacency matrix can be written as the expected matrix of the adjacency matrices of the links. We use Equation 19.3 to write the terms using the adjacency matrices so that $\langle f, Wg\rangle_{\Pi_0} = \frac{1}{2}\langle f, Ag\rangle$ and

$$\mathbb{E}_{v\sim\Pi_0}\left[\langle f_v, W_v g_v\rangle_{\Pi_0^v}\right] = \mathbb{E}_{v\sim\Pi_0}\left[\frac{1}{2}\langle f_v, A_v g_v\rangle\right] = \mathbb{E}_{v\sim\Pi_0}\left[\frac{1}{2}\langle f, \tilde{A}_v g\rangle\right] = \frac{1}{2}\left\langle f, \left(\mathbb{E}_{v\in\Pi_0}\tilde{A}_v\right) g\right\rangle$$

where $\tilde{A}_v$ is just the extended matrix of A_v (with zero rows and columns) so that it has the same dimension as A. We prove the statement by showing that $A = \mathbb{E}_{v\in\Pi_0}\tilde{A}_v$. Using conditional probability, for each entry (u,w),

$$\begin{aligned} A_{u,w} &= \Pi_1(\{u,w\}) = \frac{1}{3}\sum_{v:\{u,v,w\}\in X(2)} \Pi_2(\{u,v,w\}) = \sum_{v:\{u,v,w\}\in X(2)} \Pi_0(v)\cdot\Pi_1^v(\{u,w\}) \\ &= \sum_{v\in X(0)} \Pi_0(v)\cdot\Pi_1^v(\{u,w\}) = \sum_{v\in X(0)} \Pi_0(v)\cdot\left(\tilde{A}_v\right)_{u,w}, \end{aligned}$$

where the first line is by Equation 19.1 and Equation 19.2.

An alternative succinct way [Ove20] to write the above proof is to rewrite the first step as $\langle f, Wg\rangle_{\Pi_0} = \mathbb{E}_{\{v,w\}\sim\Pi_1}[f(v)g(w)]$ (exercise) and the remaining steps as

$$\begin{aligned} \mathbb{E}_{\{u,w\}\sim\Pi_1}[f(u)g(w)] &= \mathbb{E}_{\{u,v,w\}\sim\Pi_2}\mathbb{E}_{\{u,w\}|\{u,v,w\}}[f(u)g(w)] = \mathbb{E}_{v\sim\Pi_0}\mathbb{E}_{\{u,v,w\}|v}[f(u)g(w)] \\ &= \mathbb{E}_{v\sim\Pi_0}\mathbb{E}_{\{u,w\}\sim\Pi_1^v}[f(u)g(w)] = \mathbb{E}_{v\sim\Pi_0}\left[\langle f_v, W_v g_v\rangle_{\Pi_0^v}\right]. \end{aligned}$$

□

Proof of Oppenheim's Theorem

We are ready to prove Oppenheim Theorem 19.15. Let G be the graph of X and W be its random walk matrix. Since G is connected by assumption, the second largest eigenvalue of W is less than one. Let $\lambda < 1$ be the second largest eigenvalue of W and f be a corresponding eigenvector achieving the maximum of the Rayleigh quotient in Equation 19.4 with $\langle f, \vec{1}\rangle_{\Pi_0} = 0$. We assume without loss of generality that $\langle f, f\rangle_{\Pi_0} = 1$. Therefore, by Lemma 19.24,

$$\lambda = \langle f, Wf\rangle_{\Pi_0} = \mathbb{E}_{v\in\Pi_0}\left[\langle f_v, W_v f_v\rangle_{\Pi_0^v}\right]$$

As each W_v is a random walk matrix, the largest eigenvalue of each W_v is one with the corresponding eigenvector being $\vec{1}_v/\|\vec{1}_v\|_{\Pi_0^v} = \vec{1}_v$, where $\vec{1}_v$ is the localization of $\vec{1}$ into $X_v(0)$ as described in Definition 19.22. To bound $\langle f_v, W_v f_v\rangle_{\Pi_0^v}$, we use Equation 19.5 to decompose the vector f_v as

$$f_v = \langle f_v, \vec{1}_v\rangle_{\Pi_0^v}\cdot\vec{1}_v + f_v^\perp \quad \text{where} \quad \langle\vec{1}_v, f_v^\perp\rangle_{\Pi_0^v} = 0.$$

Expanding the quadratic form using $W_v\vec{1}_v = \vec{1}_v$, $\langle\vec{1}_v, f_v^\perp\rangle_{\Pi_0^v} = 0$ and the self-adjoint property in Equation 19.3, the cross terms are zero and we get

$$\langle f_v, W_v f_v\rangle_{\Pi_0^v} = \langle f_v, \vec{1}_v\rangle^2_{\Pi_0^v} + \langle f_v^\perp, W_v f_v^\perp\rangle_{\Pi_0^v} \le \langle f_v, \vec{1}_v\rangle^2_{\Pi_0^v} + \gamma\langle f_v^\perp, f_v^\perp\rangle_{\Pi_0^v} = (1-\gamma)\langle f_v, \vec{1}_v\rangle^2_{\Pi_0^v} + \gamma\langle f_v, f_v\rangle_{\Pi_0^v}$$

where the inequality is by the characterization of the second largest eigenvalue in Equation 19.4 and the assumption that each link has second largest eigenvalue at most γ, and the last equality is by $\langle f_v, f_v\rangle_{\Pi_0^v} = \langle f_v, \vec{1}_v\rangle^2_{\Pi_0^v} + \langle f_v^\perp, f_v^\perp\rangle_{\Pi_0^v}$ by the same orthonormality argument. Therefore, using the decomposition of the denominator in Lemma 19.23 and $\langle f, f\rangle_{\Pi_0} = 1$,

$$\mathbb{E}_{v\in\Pi_0}\left[\langle f_v, W_v f_v\rangle_{\Pi_0^v}\right] \le \mathbb{E}_{v\in\Pi_0}\left[(1-\gamma)\langle f_v, \vec{1}_v\rangle^2_{\Pi_0^v} + \gamma\langle f_v, f_v\rangle_{\Pi_0^v}\right] = \gamma + (1-\gamma)\cdot\mathbb{E}_{v\in\Pi_0}\left[\langle f_v, \vec{1}_v\rangle^2_{\Pi_0^v}\right].$$

Note that, by Equation 19.2 and Definition 19.12,

$$\langle f_v, \vec{1}_v\rangle_{\Pi_0^v} = \sum_{w\in X_v(0)}\Pi_0^v(w) f_v(w) = \sum_{w\in X_v(0)}\frac{\Pi_1(\{v,w\})}{2\Pi_0(v)}\cdot f_v(w) = \sum_{w\in X_v(0)} W(v,w)\cdot f_v(w) = (Wf)(v).$$

Hence, as f is an eigenvector of W with eigenvalue λ and $\langle f, f\rangle_{\Pi_0} = 1$,

$$\mathbb{E}_{v\in\Pi_0}\left[\langle f_v, \vec{1}_v\rangle^2_{\Pi_0^v}\right] = \mathbb{E}_{v\in\Pi_0}\left[(Wf)(v)^2\right] = \langle Wf, Wf\rangle_{\Pi_0} = \lambda^2\cdot\langle f, f\rangle_{\Pi_0} = \lambda^2.$$

To summarize,

$$\lambda = \mathbb{E}_{v\in\Pi_0}\left[\langle f_v, W_v f_v\rangle_{\Pi_0^v}\right] \le \gamma + (1-\gamma)\cdot\mathbb{E}_{v\in\Pi_0}\left[\langle f_v, \vec{1}_v\rangle^2_{\Pi_0^v}\right] = \gamma + (1-\gamma)\lambda^2$$

Solving this quadratic inequality gives either $\lambda \ge 1$ or $\lambda \le \gamma/(1-\gamma)$. Since G is connected and $\lambda < 1$, we conclude that $\lambda \le \gamma/(1-\gamma)$ as stated in Theorem 19.15.

19.5 Problems

The following are two interesting problems.

Problem 19.25 (Spanning Tree Complex without Oppenheim)**.** *Use the results in Chapter 16 to prove directly that the spanning tree complex is a* 0*-local-spectral expander, without using Oppenheim's trickling down theorem.*

Problem 19.26 (Approximate Negative Correlation of Matroids)**.** *In Chapter 16, we have seen that the variables in a random spanning tree are negatively correlated, such that for any two edges* $e \neq f$*,*

$$\Pr_T[e \in T \mid f \in T] \leq \Pr_T[e \in T].$$

This is known to be not necessarily true for general matroids, but not all is lost. Use the result that any matroid complex is a 0*-local-spectral expander in Theorem 19.17 to prove that for any two elements* $i \neq j$ *in a matroid,*

$$\Pr_B[i \in B \mid j \in B] \leq 2 \Pr_B[i \in B],$$

where B *is a uniform random basis of the matroid.*

Question 19.27. *It is an open question what is the best constant that one could prove for the approximate negative correlation property of matroids in Problem 19.26. There are examples showing that the constant is at least* $8/7$*, and some conjectured that this is tight.*

19.6 References

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1–12. ACM, 2019. 184, 198, 209, 210

[DDFH18] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean function analysis on high-dimensional expanders. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2018. 180, 196

[DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 974–985. IEEE Computer Society, 2017. 179, 183

[HL21] Max Hopkins and Shachar Lovett. Course notes of "Expander graphs and high-dimensional expanders", lecture 6. 2021. https://cseweb.ucsd.edu/classes/sp21/cse291-g/CSE_291_Expanders_and_HDX_Course_Notes.pdf. 179

[KM17] Tali Kaufman and David Mass. High dimensional random walks and colorful expansion. In *8th Innovations in Theoretical Computer Science Conference, ITCS*, volume 67, pages 4:1–4:27, 2017. 179, 183, 191

[KO20] Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Comb.*, 40(2):245–281, 2020. 179, 183, 194, 196

[Lub18] Alexander Lubotzky. High dimensional expanders. In *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pages 705–730, 2018. 179, 183

[Opp18] Izhar Oppenheim. Local spectral expansion approach to high dimensional expanders part I: descent of spectral gaps. *Discrete & Computational Geometry*, 59(2):293–330, 2018. 183, 184

[Ove20] Shayan Oveis Gharan. Course notes of "Polynomial paradigm in algorithm design", lecture 12. 2020. https://homes.cs.washington.edu/~shayan/courses/polynomials/. 128, 130, 153, 158, 171, 173, 175, 179, 187, 211

Chapter 20

Higher Order Random Walks

We study two related random walks on simplicial complexes, called the down-up walks and up-down walks. The main result is that they are fast mixing if the simplicial complex is a good local-spectral expander. A consequence is that the natural random walks on matroid bases is fast mixing, proving the long-standing matroid expansion conjecture.

20.1 Random Walks on Simplicial Complexes

Kaufman and Mass [KM17] defined two natrual random walks on faces of dimension k in a simplicial complex, the up-down walks that go through faces of dimension $k+1$ and the down-up walks that go through faces of dimension $k-1$. The most intuitive way to define these walks is to consider the following bipartite graphs.

Definition 20.1 (Bipartite Graph of a Layer)**.** *Let (X,Π) be a pure d-dimensional simplicial complex. For any $-1 \le k \le d-1$, the bipartite graph $H_k = (X(k), X(k+1); E)$ has one vertex for each face in $X(k) \cup X(k+1)$, with an edge between a face $\alpha \in X(k)$ and a face $\beta \in X(k+1)$ if and only if $\alpha \subset \beta$ and the weight of this edge is $\frac{1}{k+2} \cdot \Pi_{k+1}(\beta)$.*

Up and Down Operators

We consider the random walk matrix of these bipartite graphs and define the important up and down operators, which correspond to one-step random walks on the bipartite graphs in Definition 20.1.

Definition 20.2 (Up and Down Operators)**.** *Let (X,Π) be a pure d-dimensional simplicial complex. Let A_k be the adjacency matrix of H_k with $A_k(\alpha,\beta) = A_k(\beta,\alpha) = \frac{1}{k+2} \cdot \Pi_{k+1}(\beta)$ if $\alpha \subset \beta$ for any $\alpha \in X(k)$ and $\beta \in X(k+1)$ and zero otherwise. For each face $\alpha \in X(k)$, the weight degree of α is*

$$\deg(\alpha) := \sum_{\beta \in X(k+1): \beta \supset \alpha} A_k(\alpha,\beta) = \sum_{\beta \in X(k+1): \beta \supset \alpha} \frac{1}{k+2} \cdot \Pi_{k+1}(\beta) = \Pi_k(\alpha),$$

where the last equality is by Equation 19.1. For each face $\beta \in X(k+1)$, the weighted degree of β is

$$\deg(\beta) := \sum_{\alpha \in X(k): \alpha \subset \beta} A_k(\alpha,\beta) = \sum_{\alpha \in X(k): \alpha \subset \beta} \frac{1}{k+2} \cdot \Pi_{k+1}(\beta) = \Pi_{k+1}(\beta).$$

The random walk matrix W_k of H_k can thus be written as

$$W_k = \begin{pmatrix} 0 & D_{k+1} \\ U_k & 0 \end{pmatrix},$$

where D_{k+1} is a $X(k) \times X(k+1)$ matrix and U_k is a $X(k+1) \times X(k)$ matrix with

$$D_{k+1}(\alpha, \beta) = \frac{A_k(\alpha, \beta)}{\deg(\alpha)} = \frac{\Pi_{k+1}(\beta)}{(k+2)\Pi_k(\alpha)} \quad \text{and} \quad U_k(\beta, \alpha) = \frac{A_k(\alpha, \beta)}{\deg(\beta)} = \frac{1}{k+2}.$$

for $\alpha \in X(k)$ and $\beta \in X(k+1)$ satisfying $\alpha \subset \beta$. The matrix D_{k+1} is called the down operator from $X(k+1)$ to $X(k)$ and U_k is called the up operator from $X(k)$ to $X(k+1)$.

The following remark may clear up some potential confusion about the naming convention.

Remark 20.3 (Down Up Confusion)**.** *The name down operator comes from the perspective that D_{k+1} is an operator that maps a function $f : X(k+1) \to \mathbb{R}$ to a function $g = D_{k+1}f : X(k) \to \mathbb{R}$, and so the output is one dimension lower and it is called a down operator. In other words, the name comes from when we do right-multiplication on the matrix.*

When we do random walks, however, we do left-multiplication of the form $p^T W_k$. So D_{k+1} actually maps a distribution in $X(k)$ to a distribution in $X(k+1)$, and the output is one dimension higher. It is a bit confusing for us because we mostly think about random walks, but it won't be a big issue that we won't often talk about these down and up operators alone.

A useful property is the adjoint property of the up and down operators.

Exercise 20.4 (Adjoint Property)**.** *Let (X, Π) be a pure d-dimensional simplicial complex. Prove that for any $f : X(k) \to \mathbb{R}$ and $g : X(k+1) \to \mathbb{R}$,*

$$\langle U_k f, g \rangle_{\Pi_{k+1}} = \langle f, D_{k+1} g \rangle_{\Pi_k}.$$

Up-Down Walks and Down-Up Walks

The two random walks defined by Kaufman and Mass correspond to two-steps random walks on the bipartite graphs in Definition 20.1.

Definition 20.5 (Up-Down Walks and Down-Up Walks)**.** *Let (X, Π) be a pure d-dimensional simplicial complex. Let H_k be the bipartite graph in Definition 20.1 and W_k be the random walk matrix on H_k in Definition 20.2. Consider*

$$W_k^2 = \begin{pmatrix} D_{k+1}U_k & 0 \\ 0 & U_k D_{k+1} \end{pmatrix} =: \begin{pmatrix} P_k^{\triangle} & 0 \\ 0 & P_{k+1}^{\triangledown} \end{pmatrix},$$

where $P_k^{\triangle} \in \mathbb{R}^{X(k) \times X(k)}$ is called the up-down walk matrix and $P_{k+1}^{\triangledown} \in \mathbb{R}^{X(k+1) \times X(k+1)}$ is called the down-up walk matrix.

A simple but important property of $P_k^{\triangle}$ and $P_{k+1}^{\triangledown}$ is that they have the same spectrum. This will be used in an inductive proof to analyze the spectrum of $P_d^{\triangledown}$.

Exercise 20.6 (Same Spectrum of $P_k^{\triangle}$ and $P_{k+1}^{\triangledown}$). *Prove that there is a one-to-one correspondence between the non-zero eigenvalues of $P_k^{\triangle}$ and $P_{k+1}^{\triangledown}$.*

It will be helpful to write out the entries of $P_k^{\triangle}$ and $P_{k+1}^{\triangledown}$ explicitly.

Exercise 20.7 (Entries of $P_k^{\triangle}$ and $P_{k+1}^{\triangledown}$). *Let (X,Π) be a pure d-dimensional simplicial complex. For $\alpha,\alpha' \in X(k)$,*

$$P_k^{\triangle}(\alpha,\alpha') = \begin{cases} \frac{1}{k+2} & \text{if } \alpha = \alpha' \\ \frac{\Pi_{k+1}(\alpha\cup\alpha')}{(k+2)^2\cdot\Pi_k(\alpha)} & \text{if } \alpha\cup\alpha' \in X(k+1) \\ 0 & \text{otherwise.} \end{cases}$$

For $\beta,\beta' \in X(k+1)$.

$$P_{k+1}^{\triangledown}(\beta,\beta') = \begin{cases} \sum_{\alpha\in X(k):\alpha\subset\beta} \frac{\Pi_{k+1}(\beta')}{(k+2)^2\cdot\Pi_k(\alpha)} & \text{if } \beta = \beta' \\ \frac{\Pi_{k+1}(\beta')}{(k+2)^2\cdot\Pi_k(\beta\cap\beta')} & \text{if } \beta\cap\beta' \in X(k) \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $P_0^{\triangle}$ is just the standard lazy random walks on a graph. The non-lazy up-down walks turn out to be important in the analysis.

Definition 20.8 (Non-Lazy Up-Down Walks). *Let (X,Π) be a pure d-dimensional simplicial complex. For $-1 \le k \le d-1$, the non-lazy up-down walk matrix $P_k^{\wedge} \in \mathbb{R}^{X(k)\times X(k)}$ is defined as*

$$P_k^{\wedge} := \frac{k+2}{k+1}\Big(P_k^{\triangle} - \frac{I}{k+2}\Big).$$

Explicitly, for $\alpha,\alpha' \in X(k)$,

$$P_k^{\wedge}(\alpha,\alpha') = \begin{cases} \frac{\Pi_{k+1}(\alpha\cup\alpha')}{(k+1)(k+2)\cdot\Pi_k(\alpha)} & \text{if } \alpha\cup\alpha' \in X(k+1) \\ 0 & \text{otherwise.} \end{cases}$$

For the notations, remember that the $\triangle$ in $P^{\triangle}$ represents that there could be self-loops, while the $\wedge$ in $P^{\wedge}$ represents that the two endpoints are different.

The stationary distributions of $P_k^{\triangle}, P_k^{\wedge}, P_k^{\triangledown}$ are all the same. This can be checked by direct calculations or check that the time reversible condition (i.e. $\pi_i P(i,j) = \pi_j P(j,i)$ for all i,j) is satisfied.

Exercise 20.9 (Stationary Distributions). *The stationary distributions of $P_k^{\triangle}, P_k^{\wedge}, P_k^{\triangledown}$ are Π_k.*

This will allow us to use the inner product $\langle\cdot,\cdot\rangle_{\Pi_k}$ to bound the eigenvalues of $P_k^{\triangle}, P_k^{\wedge}, P_k^{\triangledown}$ using the Rayleigh quotients with Π_k in Definition 19.21.

Random Walks on Matroid Bases: To sample a uniform random basis of a matroid, we consider the matroid complex with the uniform distribution on the bases, and run the down-up walk $P_d^{\triangledown}$. Then, by Exercise 20.9, the stationary distribution is the uniform distribution. Note that the down-up walk $P_d^{\triangledown}$ is the natural algorithm that we start from an arbitrary basis B_0, and in each iteration $t \ge 0$ we drop a random element i of the current basis and then add a random element j so that $B_{t+1} := B_t - i + j$ is a basis, and repeat. Observe that the random spanning tree algorithm in

Chapter 6 is a special case. We know from Theorem 19.17 that a matroid complex is a 0-local-spectral expander. We will see in the next section that the up-down walks and the down-up walks of a good local-spectral expander mix quickly. Thus this provides a simple and efficient algorithm to sample a uniform matroid basis, answering a long-standing open question called the matroid expansion conjecture that we will explain at the end in the next section.

20.2 Kaufman-Oppenheim Theorem

Kaufman and Oppenheim [KO20] proved that if the simplicial complex is a good local-spectral expander, then the up-down walks and the down-up walks mix quickly.

Theorem 20.10 (Kaufman-Oppenheim Second Eigenvalue Bound [KO20])**.** *If (X,Π) is a γ-local-spectral expander, then for any $0 \le k \le d$, the second eigenvalue of $P_k^{\triangledown}$ is*

$$\lambda_2(P_k^{\triangledown}) \le 1 - \frac{1}{k+1} + k\gamma.$$

We will present the proof of this theorem in the rest of this section, and discuss the matroid expansion conjecture at the end.

Garland's Method

An important step in the proof is to use Garland's method to decompose the down-up walk matrix and the up-down walk matrix into matrices of the links. We need a definition similar to, but different from, Definition 19.22.

Definition 20.11 (Restriction to Link)**.** *Given a simplicial complex (X,Π), a function $f : X(k) \to \mathbb{R}$ and a face $\tau \in X(k-1)$, the restriction of f to $X_\tau(0)$ is defined as $f_\tau : X_\tau(0) \to \mathbb{R}$ such that $f_\tau(x) = f(\tau \cup \{x\})$ for all $x \in X_\tau(0)$.*

The following lemma shows that the non-lazy up-down walk matrix $P_k^{\wedge}$ can be decomposed into the random walk matrix W_τ of the links for $\tau \in X(k-1)$. The reason that we consider the non-lazy up-down walk is that there are no self-loops in the random walk matrices of the links.

Lemma 20.12 (Decomposition of Non-Lazy Up-Down Walk Matrix)**.** *For any pure d-dimensional simplicial complex (X,Π) and any function $f : X(k) \to \mathbb{R}$,*

$$\langle f, P_k^{\wedge} f\rangle_{\Pi_k} = \mathbb{E}_{\tau\sim\Pi_{k-1}} \langle f_\tau, W_\tau f_\tau\rangle_{\Pi_0^\tau},$$

where W_τ is the random walk matrix of the link τ in Definition 19.12.

Proof. The main idea is to decompose $P_k^{\wedge}$ into transition matrices where each is about the transitions involving a particular link $\tau \in X(k-1)$. Let $P_\tau^{\wedge}$ be the $X(k) \times X(k)$ matrix with

$$P_\tau^{\wedge}(\sigma,\sigma') = \frac{\Pi_{k+1}(\alpha\cup\alpha')}{(k+1)(k+2)\cdot\Pi_k(\alpha)} \quad \text{if } \alpha\cap\alpha' = \tau \quad \text{and} \quad P_\tau^{\wedge}(\sigma,\sigma') = 0 \quad \text{otherwise.}$$

Note that $P_k^{\wedge} = \sum_{\tau\in X(k-1)} P_\tau^{\wedge}$ by Definition 20.8, as the transition between any two faces $\alpha,\alpha' \in X(k)$ involves a unique link $\tau = \alpha\cap\alpha' \in X(k-1)$. The observation is that this matrix $P_\tau^{\wedge}$ is almost

the same as the random walk matrix W_τ of the link τ. Let $\alpha = \tau \cup \{x\}$ and $\alpha' = \tau \cup \{y\}$. Then, from Definition 19.12,

$$W_\tau(x,y) = \frac{\Pi(\tau \cup \{x,y\})}{(|\tau|+2)\cdot \Pi(\tau \cup \{x\})} = \frac{\Pi_{k+1}(\alpha \cup \alpha')}{(k+2)\cdot \Pi_k(\alpha)} \quad \Longrightarrow \quad P_\tau^\wedge(\sigma,\sigma') = \frac{1}{k+1} W_\tau(\sigma \setminus \tau, \sigma' \setminus \tau).$$

So, if we extend the small matrices W_τ appropriately to $\tilde{W}_\tau$ (i.e. put the (x,y)-entry of W_τ on the $(\tau \cup \{x\}, \tau \cup \{y\})$-entry of $\tilde{W}_\tau$ and set all other entries to be zero), then

$$P_k^\wedge = \frac{1}{k+1} \sum_{\tau \in X(k-1)} \tilde{W}_\tau,$$

and so it should be clear that a quadratic form involving $P_k^\wedge$ can be decomposed as a sum of quadratic forms involving W_τ as in the statement.

To write it concisely, we decompose the quadratic form directly (instead of decomposing the matrix $P_k^\wedge$). By writing the quadratic form as a sum of $|X(k)| \times |X(k)|$ terms,

$$\langle f, P_k^\wedge f \rangle_{\Pi_k} = \sum_{\sigma \in X(k), \sigma' \in X(k) : \sigma \cup \sigma' \in X(k+1)} \frac{\Pi_{k+1}(\sigma \cup \sigma')}{(k+1)(k+2)} \cdot f(\sigma) \cdot f(\sigma')$$

For each pair $\sigma, \sigma' \in X(k)$ with $\sigma \cup \sigma' \in X(k+1)$, their intersection $\tau := \sigma \cap \sigma' \in X(k-1)$. Let $x = \sigma \setminus \tau$ and $y = \sigma' \setminus \tau$. Then the corresponding term on the RHS is from the link X_τ with contribution

$$\begin{aligned} & \Pi_{k-1}(\tau) \cdot \Pi_0^\tau(x) \cdot W_\tau(x,y) \cdot f_\tau(x) \cdot f_\tau(y) \\ = \; & \Pi_{k-1}(\tau) \cdot \frac{\Pi_k(\tau \cup \{x\})}{(|\tau|+1)\cdot \Pi_{k-1}(\tau)} \cdot \frac{\Pi_{k+1}(\tau \cup \{x,y\})}{(|\tau|+2)\cdot \Pi_k(\tau \cup \{x\})} \cdot f(\tau \cup \{x\}) \cdot f(\tau \cup \{y\}) \\ = \; & \frac{\Pi_{k+1}(\sigma \cup \sigma')}{(k+1)(k+2)} \cdot f(\sigma) \cdot f(\sigma'). \end{aligned}$$

The statement follows by noting that there is a one-to-one correspondence because each transition in $P_k^\wedge$ involves a unique link $\tau \in X(k-1)$. □

The next lemma shows that the down-up walk matrix can be decomposed as the down-up walk matrices of the links which are simple rank-one matrices.

Lemma 20.13 (Decomposition of Down-Up Walk Matrix)**.** *For any pure d-dimensional simplicial complex (X,Π) and any function $f : X(k) \to \mathbb{R}$,*

$$\langle f, P_k^\vee f \rangle_{\Pi_k} = \mathbb{E}_{\tau \sim \Pi_{k-1}} \langle f_\tau, J_\tau f_\tau \rangle_{\Pi_0^\tau},$$

where $J_\tau := \vec{1}(\Pi_0^\tau)^T$ is a $X_\tau(0) \times X_\tau(0)$ rank-one matrix.

Proof. The proof is similar to that in Lemma 20.12. We decompose $P_k^\vee$ into transition matrices where each is about the transitions involving a particular link $\tau \in X(k-1)$. Let $P_\tau^\vee$ be the $X(k) \times X(k)$ matrix with

$$P_\tau^\vee(\sigma,\sigma') = \frac{\Pi_k(\sigma')}{(k+1)^2 \cdot \Pi_{k-1}(\tau)} \quad \text{if } \sigma \cap \sigma' \supseteq \tau \quad \text{and} \quad P_\tau^\vee(\sigma,\sigma') = 0 \quad \text{otherwise.}$$

Note that $P_k^{\triangledown} = \sum_{\tau \in X(k-1)} P_\tau^{\triangledown}$ by Exercise 20.7, where the summation in the self-loop probability for σ is split into the subsets $\tau \subset \sigma$ for $\tau \in X(k-1)$ where each takes a summand. Let $\sigma = \tau \cup \{x\}$ and $\sigma' = \tau \cup \{y\}$. By the definition of J_τ and Equation 19.2,

$$J_\tau(x,y) = \Pi_0^\tau(y) = \frac{\Pi(\tau \cup \{y\})}{(|\tau|+1)\cdot \Pi(\tau)} = \frac{\Pi_k(\alpha')}{(k+1)\cdot \Pi_{k-1}(\tau)} \quad \Longrightarrow \quad P_\tau^{\triangledown}(\sigma,\sigma') = \frac{1}{k+1} J_\tau(\sigma\backslash\tau, \sigma'\backslash\tau).$$

Check that the remaining calculations are similar to that in Lemma 20.12, with the contribution from $\sigma, \sigma' \in X(k)$ involving at link $\tau \in X(k-1)$ is the same from LHS and RHS, being equal to

$$\frac{\Pi_k(\sigma)\cdot\Pi_k(\sigma')}{(k+1)^2\cdot\Pi_{k-1}(\tau)} \cdot f(\sigma)\cdot f(\sigma').$$

□

Comparing Down-Up Walk and Non-Lazy Up-Down Walk

The decomposition of the down-up walk matrix in Lemma 20.13 shows that $P_k^{\triangledown}$ can be written as the sum of rank-one matrices in the links each with second largest eigenvalue 0, while the decomposition of the non-lazy up-down walk matrix in Lemma 20.12 shows that $P_k^{\wedge}$ can be wrttien as the sum of random walk matrices in the links each with second largest eigenvalue at most γ for a γ-local-spectral expander. The main step in Kaufman-Oppenheim's theorem is to compare the spectrum of the down-up walk matrix $P_k^{\triangledown}$ with the non-lazy up-down walk matrix $P_k^{\wedge}$, which intuitively can be understood as a term-by-term comparison between a complete graph and an expander graph.

Proposition 20.14 (Comparision of $P_k^{\triangledown}$ and $P_k^{\wedge}$ [KO20, DDFH18]). *If (X,Π) is a γ-local-spectral expander, then*

$$P_k^{\wedge} - P_k^{\triangledown} \preccurlyeq_{\Pi_k} \gamma I$$

for any $0 \le k \le d-1$, where $A \preccurlyeq_\Pi B$ denotes $\langle f, Af\rangle_\Pi \le \langle f, Bf\rangle_\Pi$ for all f.

Proof. Using Lemma 20.12 and Lemma 20.13,

$$\big\langle f, (P_k^{\wedge} - P_k^{\triangledown}) f\big\rangle_{\Pi_k} = \mathbb{E}_{\tau\sim\Pi_{k-1}} \big\langle f_\tau, (W_\tau - J_\tau) f_\tau\big\rangle_{\Pi_0^\tau}.$$

For each term, write $f_\tau = c\vec{1} + f_\tau^\perp$ where $\big\langle \vec{1}, f_\tau^\perp\big\rangle_{\Pi_0^\tau} = 0$ as in Equation 19.5. Then note that

$$\big\langle f_\tau, (W_\tau - J_\tau) f_\tau\big\rangle_{\Pi_0^\tau} = \big\langle c\vec{1} + f_\tau^\perp, (W_\tau - J_\tau)(c\vec{1} + f_\tau^\perp)\big\rangle_{\Pi_0^\tau} = \big\langle f_\tau^\perp, (W_\tau - J_\tau) f_\tau^\perp\big\rangle_{\Pi_0^\tau},$$

because $W_\tau \vec{1} = J_\tau \vec{1} = \vec{1}$ and also $(\Pi_0^\tau)^T W_\tau = (\Pi_0^\tau)^T J_\tau = (\Pi_0^\tau)^T$. Therefore,

$$\begin{aligned}
\langle f, (P_k^{\wedge} - P_k^{\triangledown}) f\rangle_{\Pi_k} &= \mathbb{E}_{\tau\sim\Pi_{k-1}} \big\langle f_\tau^\perp, (W_\tau - J_\tau) f_\tau^\perp\big\rangle_{\Pi_0^\tau} \\
&\le \mathbb{E}_{\tau\sim\Pi_{k-1}} \gamma \big\langle f_\tau^\perp, f_\tau^\perp\big\rangle_{\Pi_0^\tau} \\
&\le \mathbb{E}_{\tau\sim\Pi_{k-1}} \gamma \big\langle f_\tau, f_\tau\big\rangle_{\Pi_0^\tau} \\
&= \gamma \big\langle f, f\big\rangle_{\Pi_k},
\end{aligned}$$

where the first inequality is by the Rayleigh quotient characterization in Equation 19.4 and the assumption that (X,Π) is a γ-local-spectral expander, and the last equality is left as Exercise 20.15. □

Exercise 20.15 (Decomposition of Identity). *Prove that $\langle f, f\rangle_{\Pi_k} = \mathbb{E}_{\tau\sim\Pi_{k-1}} \langle f_\tau, f_\tau\rangle_{\Pi_0^\tau}$.*

Inductive Proof

Now we are ready to prove Kaufman-Oppenheim's Theorem 20.10. The proof is by an interesting induction, that we start from the spectrum of $P_0^{\triangledown}$ and use Exercise 20.6 and Proposition 20.14 to reason about the spectrums of $P_k^{\triangledown}$ and $P_k^{\triangle}$ and $P_k^{\wedge}$. We prepare with the following exercise which will be used in reasoning about the spectrums.

Exercise 20.16 (Bounding Spectrum by Quadratic Forms)**.** *Let $A, B \in \mathbb{R}^{n\times n}$ be two self-adjoint matrices with respect to the inner product Π (see Equation 19.3). If $A \preccurlyeq_\Pi B$ as described in Proposition 20.14, then $\lambda_i(A) \le \lambda_i(B)$ for all $1 \le i \le n$.*

Proof of Theorem 20.10. The proof is by induction on k. In the base case when $k = 0$, the matrix $P_0^{\triangledown}$ is of rank one (see Exercise 20.7), and thus the second largest eigenvalue is at most 0, and the statement holds.

Now, assume the statement holds for k, and we would like to prove the inductive step. By Proposition 20.14, $P_k^{\wedge} \preccurlyeq_{\Pi_k} P_k^{\triangledown} + \gamma I$. It follows from Exercise 20.16 and Exercise 20.9 that

$$\lambda_2(P_k^{\wedge}) \le \lambda_2(P_k^{\triangledown}) + \gamma \le 1 - \frac{1}{k+1} + (k+1)\gamma,$$

where the second inequality is by the induction hypothesis on $P_k^{\triangledown}$. Recall from Definition 20.8 that

$$P_k^{\wedge} = \frac{k+2}{k+1}\Big(P_k^{\triangle} - \frac{I}{k+2}\Big) \quad \Longrightarrow \quad P_k^{\triangle} = \frac{k+1}{k+2}P_k^{\wedge} + \frac{I}{k+2}.$$

Therefore, the second largest eigenvalue of $P_k^{\triangle}$ is

$$\lambda_2(P_k^{\triangle}) \le \frac{k+1}{k+2}\Big(1 - \frac{1}{k+1} + (k+1)\gamma\Big) + \frac{1}{k+2} = 1 - \frac{1}{k+2} + \frac{(k+1)^2\gamma}{k+2} \le 1 - \frac{1}{k+2} + (k+1)\gamma.$$

Finally, recall that $P_{k+1}^{\triangledown}$ and $P_k^{\triangle}$ have the same spectrum by Exercise 20.6, and this completes the induction step. □

Combinatorial Interpretation: To summarize, one could visualize the proof as having a stack of bipartite graphs, one for each layer as in Definition 20.1. To reason about the spectrum of the top layer, we start from the down-up walk of the bottom layer. The key step using Garland's method is to observe that the (non-lazy) up-down walk in the layer above has a similar structure to the down-up walk in the layer below, by replacing each clique J_τ in a link τ in the down-up walk in Lemma 20.13 by an expander graph W_τ in the up-down walk in Lemma 20.12, whose expansion comes from the assumption of the local-spectral expander. So, if the down-up walk is an expander then the up-down walk is still an expander but with slightly weaker expansion (as we just replace a complete graph by an expander graph), and this is essentially the term-by-term comparison step of Kaufman and Oppenheim in Proposition 20.14. Finally, within the same layer, we use the simple but important property that the up-down walk and the down-up walk having the same spectrum to carry out the induction.

Matroid Expansion Conjecture

Recall from Theorem 19.17 that the matroid complex is a 0-local-spectral expander (using Oppenheim's trickling down Theorem 19.15). Then, by Kaufman-Oppenheim's Theorem 20.10, $\lambda_2(P_d^{\nabla}) \leq 1 - \frac{1}{d+1}$ where $r := d+1$ is the rank of the matroid. By standard analysis of mixing time in Theorem 6.16, the ϵ-mixing time of the natural down-up walks is at most $O(r \log \frac{N}{\epsilon}) = O(r^2 \log \frac{n}{\epsilon})$ where N is the number of bases and n is the number of elements in the ground set.

Theorem 20.17 (Sampling Matroid Bases by Down-Up Walks [ALOV19]). *Given a matroid M with n elements and rank at most r, the mixing time of the down-up walk of the matroid complex is at most $O(r^2 \log \frac{n}{\epsilon})$.*

The matroid expansion conjecture by Mihail and Vazirani from 1989 states that the bases exchange graph has edge expansion at least one, which follows from Theorem 20.17 and Cheeger's inequality in Theorem 4.3.

Problem 20.18 (Matroid Expansion Conjecture). *The bases exchange graph $G = (V, E)$ is the underlying unweighted graph of the down-up walk matrix of the matroid complex. Prove that the edge expansion of G is at least one, that is, $|\delta(S)|/|S| \geq 1$ for all $S \subseteq V$ with $|S| \leq |V|/2$.*

One may understand the resolution of the matroid expansion conjecture as using the right induction for the problem which may not be easy to come up with without the perspective of a simplicial complex and the concepts such as links. It would be great if someone could write a completely combinatorial proof (without using any linear algebra) of the matroid expansion conjecture using the combinatorial interpretation above.

Question 20.19 (Combinatorial Proof of Matroid Expansion Conjecture). *Is there a purely combinatorial proof of the matroid expansion conjecture?*

20.3 References

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1–12. ACM, 2019. 184, 198, 209, 210

[DDFH18] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean function analysis on high-dimensional expanders. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2018. 180, 196

[KM17] Tali Kaufman and David Mass. High dimensional random walks and colorful expansion. In *8th Innovations in Theoretical Computer Science Conference, ITCS*, volume 67, pages 4:1–4:27, 2017. 179, 183, 191

[KO20] Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Comb.*, 40(2):245–281, 2020. 179, 183, 194, 196

Chapter 21

Spectral Independence

We will first show that Kaufman-Oppenheim Theorem 20.10 can be improved to a natural product form. Then we will see the notion of "spectral independence", a nice probabilistic formulation of this improved result without the language of high-dimensional expanders. Finally, we will mention some recent developments using this notion in analyzing random sampling algorithms for combinatorial objects.

21.1 Improved Analysis of Higher Order Random Walks

The proof of the matroid expansion conjecture shows that the techniques developed in higher order random walks provide a completely new approach to analyze mixing times of Markov chains. Unlike previous approaches such as couplings and multicommodity flows, this simplicial complex approach directly bounds the spectral gap of the random walk matrix. It is of great interest to investigate whether this approach can be extended to other problems such as independent sets and graph colorings.

First we discuss some limitations of the results in Chapter 20. Note that Theorem 20.10 can be used to establish a non-trivial spectral gap of $P_d^{\triangledown}$ only when $\lambda < \frac{1}{d(d+1)}$, which is a very strong spectral requirement of the simplicial complex. As discussed in Problem 19.19, the second eigenvalue is at most zero if and only if the graph is a complete multi-partite graph, and more generally a 0-local-spectral expander can be shown to be a weighted matroid complex. For most natural combinatorial simplicial complexes, it does not hold that $\lambda_2(G_\alpha) \leq O(\frac{1}{d^2})$ even when restricted to faces α of dimension $d-2$. This suggests that we need to sharpen the bound in Theorem 20.10 in order to apply this approach for other problems.

Small Improvement

It was observed that the comparison bound in Proposition 20.14 can be slightly improved.

Proposition 21.1 (Improved Comparison of $P_k^{\triangledown}$ and $P_k^{\wedge}$ [AL20])**.** *Let (X, Π) be a pure d-dimensional simplicial complex. For any $0 \leq k \leq d-1$,*

$$P_k^{\wedge} - P_k^{\triangledown} \preccurlyeq_{\Pi_k} \gamma_{k-1}(I - P_k^{\triangledown})$$

where $\gamma_j := \max_{\alpha \in X(j)} \lambda_2(W_\alpha)$ is the maximum second largest eigenvalue of the link graphs of dimension j as in Definition 19.13.

Proof. Following the proof in Proposition 20.14,

$$\langle f, (P_k^{\wedge} - P_k^{\triangledown}) f \rangle_{\Pi_k} = \mathbb{E}_{\tau \sim \Pi_{k-1}} \langle f_\tau^{\perp}, (W_\tau - J_\tau) f_\tau^{\perp} \rangle_{\Pi_0^\tau} \leq \mathbb{E}_{\tau \sim \Pi_{k-1}} \gamma_{k-1} \langle f_\tau^{\perp}, f_\tau^{\perp} \rangle_{\Pi_0^\tau} \tag{21.1}$$

Instead of bounding the right hand side simply by $\mathbb{E}_{\tau \sim \Pi_{k-1}} \gamma_{k-1} \langle f_\tau, f_\tau \rangle_{\Pi_0^\tau}$, we collect the dropped terms to prove the stated bound. As in Proposition 20.14, write $f_\tau = c\vec{1} + f_\tau^{\perp}$ where $\langle \vec{1}, f_\tau^{\perp} \rangle_{\Pi_0^\tau} = 0$ and $c = \langle f_\tau, \vec{1} \rangle_{\Pi_0^\tau}$ as in Equation 19.5. So, the dropped terms are

$$\mathbb{E}_{\tau \sim \Pi_{k-1}} \gamma_{k-1} \langle c\vec{1}, c\vec{1} \rangle_{\Pi_0^\tau} = \gamma_{k-1} \cdot \mathbb{E}_{\tau \sim \Pi_{k-1}} \langle f_\tau, \vec{1} \rangle_{\Pi_0^\tau}^2 = \gamma_{k-1} \cdot \mathbb{E}_{\tau \sim \Pi_{k-1}} \langle f_\tau, J_\tau f_\tau \rangle_{\Pi_0^\tau} = \gamma_{k-1} \cdot \langle f, P_k^{\triangledown} f \rangle_{\Pi_k},$$

where the second equality is because $\langle f_\tau, \vec{1} \rangle_{\Pi_0^\tau}^2 = \langle f_\tau, \Pi_0^\tau \rangle^2 = \langle f_\tau, (\Pi_0^\tau)(\Pi_0^\tau)^T f_\tau \rangle = \langle f_\tau, J_\tau f_\tau \rangle_{\Pi_0^\tau}$ since $J_\tau = \vec{1}(\Pi_0^\tau)^T$ as defined in Lemma 20.13, and the last equality is by the statement in Lemma 20.13. Therefore, we conclude that

$$\begin{aligned}
\langle f, (P_k^{\wedge} - P_k^{\triangledown}) f \rangle_{\Pi_k} &\leq \mathbb{E}_{\tau \sim \Pi_{k-1}} \gamma_{k-1} \langle f_\tau^{\perp}, f_\tau^{\perp} \rangle_{\Pi_0^\tau} \\
&= \mathbb{E}_{\tau \sim \Pi_{k-1}} \gamma_{k-1} \langle f_\tau, f_\tau \rangle_{\Pi_0^\tau} - \mathbb{E}_{\tau \sim \Pi_{k-1}} \gamma_{k-1} \langle c\vec{1}, c\vec{1} \rangle_{\Pi_0^\tau} \\
&= \gamma_{k-1} \langle f, f \rangle_{\Pi_k} - \gamma_{k-1} \cdot \langle f, P_k^{\triangledown} f \rangle_{\Pi_k} \\
&= \gamma_{k-1} \cdot \langle f, (I - P_k^{\triangledown}) f \rangle_{\Pi_k},
\end{aligned}$$

where the second line is by $\langle f_\tau, f_\tau \rangle_{\Pi_0^\tau} = \langle c\vec{1}, c\vec{1} \rangle_{\Pi_0^\tau} + \langle f_\tau^{\perp}, f_\tau^{\perp} \rangle_{\Pi_0^\tau}$ using orthonormality, and the third line is using Exercise 20.15 and the calculation above. This holds for any f and thus implies the statement. □

Product Form

The small improvement in Proposition 21.1 is very simple, but what is perhaps surprising is that this is all we needed to prove a much sharper bound on $\lambda_2(P_k^{\triangledown})$.

Theorem 21.2 (Improved Second Eigenvalue Bound on $P_k^{\triangledown}$ [AL20])**.** *Let (X, Π) be a pure d-dimensional simplicial complex. For any $0 \leq k \leq d$,*

$$\lambda_2(P_k^{\triangledown}) \leq 1 - \frac{1}{k+1} \prod_{j=-1}^{k-2} (1 - \gamma_j),$$

where $\gamma_j := \max_{\alpha \in X(j)} \lambda_2(W_\alpha)$ is as defined in Definition 19.13.

Proof. We prove by induction on k. The base case is when $k = 0$, where $P_0^{\triangledown}$ is a rank one matrix and so $\lambda_2(P_0^{\triangledown}) \leq 0$, and hence the statement trivially holds.

Now, assume the statement holds for k, and we would like to prove the induction step. By Proposition 21.1, $P_k^{\wedge} \preccurlyeq_{\Pi_k} \gamma_{k-1} \cdot I + (1 - \gamma_{k-1}) P_k^{\triangledown}$, which implies by Exercise 20.16 and Exercise 20.9 that

$$\lambda_2(P_k^{\wedge}) \leq \gamma_{k-1} + (1 - \gamma_{k-1}) \cdot \lambda_2(P_k^{\triangledown}) \leq 1 - \frac{1}{k+1} \prod_{i=-1}^{k-1} (1 - \gamma_i),$$

where the last inequality is by the induction hypothesis. Recall from Definition 20.8 that

$$P_k^{\wedge} = \frac{k+2}{k+1}\Big(P_k^{\triangle} - \frac{I}{k+2}\Big) \quad \Longrightarrow \quad P_k^{\triangle} = \frac{k+1}{k+2}P_k^{\wedge} + \frac{I}{k+2}.$$

Therefore, the second largest eigenvalue of $P_k^{\triangle}$ is

$$\lambda_2(P_k^{\triangle}) \leq \frac{k+1}{k+2}\Big(1 - \frac{1}{k+1}\prod_{i=-1}^{k-1}(1-\gamma_i)\Big) + \frac{1}{k+2} = 1 - \frac{1}{k+2}\prod_{i=-1}^{k-1}(1-\gamma_i).$$

The induction step follows from Exercise 20.6 that $P_{k+1}^{\triangledown}$ and $P_k^{\triangle}$ have the same spectrum. □

Implications

We discuss some implications of the product form in Theorem 21.2. A basic result is that a simplicial complex X has $\lambda_2(P_d^{\triangledown}) < 1$ if and only if $\lambda_2(G_\alpha) < 1$ for every face α of dimension up to $d-2$. Theorem 21.2 provides a quantitative generalization of this result. The product form matches the combinatorial intuition that we replace the complete graphs in the links of $P_k^{\triangledown}$ by expander graphs in the links of $P_k^{\wedge}$ as described in Chapter 20, and so we expect that the spectral gap decreases by a multiplicative factor but is always non-zero.

Combining with Oppenheim's trickling down Theorem 19.15, Theorem 21.2 provides the following convenient bound for the second eigenvalue of higher order random walks in a black box fashion.

Exercise 21.3. *Let (X,Π) be a pure d-dimensional simplicial complex. For any $0 \leq k \leq d$, suppose $\gamma_{k-2} \leq \frac{1}{k+1}$ and G_α is connected for every face α up to dimension $k-2$, then*

$$\lambda_2(P_k^{\triangledown}) \leq 1 - \frac{1}{(k+1)^2}.$$

In particular, this implies that the down-up walk $P_d^{\triangledown}$ is fast mixing for any $O\big(\frac{1}{d}\big)$-local-spectral expander, which is an improvement of Theorem 20.10 where it requires the simplicial complex to be a $O\big(\frac{1}{d^2}\big)$-local-spectral expander. See [AL20] for an application of Exercise 21.3 in sampling a random independent set of size up to $n/(2\Delta)$ where n is the number of vertices and Δ is the maximum degree of the input graph.

Another consequence is that the following type of eigenvalue profile is enough to guarantee polynomial mixing time.

Exercise 21.4 (Improving Profile)**.** *Let (X,Π) be a pure d-dimensional simplicial complex. If there is a constant $0 < c < 1$ such that*

$$(\gamma_{-1},\gamma_0,\ldots,\gamma_{d-2}) = \Big(\frac{c}{d},\frac{c}{d-1},\ldots,\frac{c}{1}\Big),$$

then

$$\lambda_2(P_d^{\triangledown}) \leq 1 - \frac{1}{d^{1+c}}.$$

21.2 Spectral Independence

Anari, Liu and Oveis Gharan [ALO20] defined a notion called spectral independence, which is a nice probabilistic formulation of Theorem 21.2 without using the language of high-dimensional expanders.

The following correlation matrix is a natural matrix that records the pairwise correlation of the elements. As we will see, this matrix is closely related to the random walk matrix of the empty link of a corresponding simplicial complex of the probability distribution.

Definition 21.5 (Correlation Matrix). *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution on subsets of $[n]$. The correlation matrix of μ is a $2n \times 2n$ matrix Ψ, whose rows and columns are indexed by $[n] \times \{0,1\}$, with*

$$\Psi\big((i,a_i),(j,a_j)\big) = \Pr_{Z\sim\mu}\big[Z(j) = a_j \mid Z(i) = a_i\big] - \Pr_{Z\sim\mu}\big[Z(j) = a_j\big]$$

for $i \neq j$ and $a_i, a_j \in \{0,1\}$, and $\Psi\big((i,a_i),(j,a_j)\big) = 0$ if $i = j$.

Remark 21.6. *The correlation matrix in [ALO20] is defined slightly differently, with*

$$\Psi\big((i,a_i),(j,a_j)\big) = \Pr_{Z\sim\mu}\big[Z(j) = a_j \mid Z(i) = a_i\big] - \Pr_{Z\sim\mu}\big[Z(j) = a_j \mid Z(i) = 1-a_i\big].$$

The above definition of the correlation matrices is from [AASV21, CGSV21].

The following conditional correlation matrices are the correlation matrices given a partial assignment. As we will see, they are closely related to the random walk matrices of the links of a corresponding simplicial complex of the probability distribution.

Definition 21.7 (Conditional Correlation Matrices). *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution on subsets of $[n]$. Let $S \subseteq [n]$ be a subset of size k and let $a_S \in \{0,1\}^k$ be a binary string of length k with an entry for each element $i \in S$. Let $Z(S) = a_S$ be the event that $Z(i) = a_S(i)$ for all $i \in S$ when $Z \sim \mu$. The conditional correlation matrix Ψ_{a_S} is a $2(n-k) \times 2(n-k)$ matrix, whose rows and columns are indexed by $([n] \setminus S) \times \{0,1\}$, with*

$$\Psi_{a_S}\big((i,a_i),(j,a_j)\big) = \Pr_{Z\sim\mu}\big[Z(j) = a_j \mid Z(i) = a_i, Z(S) = a_S\big] - \Pr_{Z\sim\mu}\big[Z(j) = a_j \mid Z(S) = a_S\big]$$

for $i \neq j$ and $a_i, a_j \in \{0,1\}$, and $\Psi\big((i,a_i),(j,a_j)\big) = 0$ if $i = j$.

The following definition of spectral independence is closely related to local-spectral expansion of a corresponding simplicial complex of the probability distribution.

Definition 21.8 (Spectral Independence). *A probability distribution $\mu : \{0,1\}^n \to \mathbb{R}$ on subsets of $[n]$ is called η-spectrally independent if for any $S \subseteq [n]$ with $|S| \leq n-2$ and partial assignment $a_S \in \{0,1\}^{|S|}$,*

$$\lambda_{\max}\big(\Psi_{a_S}\big) \leq \eta.$$

Let's see some examples before we go on. First, it is easy to see that if μ is an independent product distribution (i.e. there exist $\lambda_1, \ldots, \lambda_n$ such that $\mu(S) \propto \sum_{i\in S} \lambda_i$), then μ is 0-spectrally independent. This suggests that spectral independence is an algebraic way to quantity the independence of a probability distribution.

A more interesting example is the class of negatively correlated distributions that we studied in Chapter 16.

Problem 21.9 (Spectral Independence of Strongly Rayleigh Distributions). *Let $\mu : \{0,1\}^n$ be a homogeneous distribution such that for all $i \neq j$,*

$$\Pr_{Z\sim\mu}\left[Z(i) = 1 \mid Z(j) = 1\right] \leq \Pr_{Z\sim\mu}\left[Z(i) = 1\right].$$

Prove that $\lambda_{\max}(\Psi) \leq 1$ where Ψ is the correlation matrix of μ. Conclude that a homogeneous strongly Rayleigh distribution as defined in Definition 16.1 is 1-spectrally independent.

Glauber Dynamics

A natural random walk on a probability distribution $\mu : \{0,1\}^n \to \mathbb{R}$ is called Glauber dynamics.

Definition 21.10 (Glauber Dynamics). *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution on subsets of $[n]$. Start with an arbitrary subset $S_0 \in \mathrm{supp}(\mu)$. At each iteration $t \geq 1$, we choose a uniformly random element $i \in [n]$ and set*

$$S_t := \begin{cases} S_{t-1} \setminus \{i\} & \text{with probability } \frac{\mu\left(S_{t-1}\setminus\{i\}\right)}{\mu\left(S_{t-1}\setminus\{i\}\right)+\mu\left(S_{t-1}\cup\{i\}\right)} \\ S_{t-1} \cup \{i\} & \text{otherwise.} \end{cases}$$

Check that this Markov chain has stationary distribution μ.

The main result of this formulation is to bound the spectral gap of the transition matrix of the Glauber dynamics by the spectral independence of the probability distribution.

Theorem 21.11 (Spectral Gap via Spectral Independence [ALO20]). *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution that is η-spectrally independent. The random walk matrix of the Glauber dynamics of μ has spectral gap at least*

$$\frac{1}{n}\prod_{i=0}^{n-2}\left(1 - \frac{\eta}{n-i-1}\right).$$

Simplicial Complex for Glauber Dynamics

The proof of Theorem 21.11 is by (1) defining a simplicial complex X^μ for μ, (2) showing that the down-up walk P^∇_{n-1} of X^μ is exactly the Glauber dynamics in Definition 21.10, (3) seeing that the conditional correlation matrices of μ are basically the matrices $W_\tau - J_\tau$ of the links of X^μ in the proofs of Proposition 20.14 and Proposition 21.1, and (4) seeing that the spectral gap bound in Theorem 21.11 for Glauber dynamics follows from that in Theorem 21.2 for down-up walks.

Definition 21.12 (Simplicial Complex of Assignments). *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution on subsets of $[n]$. The simplicial complex (X^μ, Π) is defined with ground set $[n]\times\{0,1\}$, with a maximal face $\zeta := \big((1,Z(1)),(2,Z(2)),\ldots,(n,Z(n))\big)$ of dimension $n-1$ with $\Pi(\zeta) := \mu(Z)$ for each $Z \in \mathrm{supp}(\mu)$. In words, each maximal face of X^μ corresponds to an assignment of the n binary variables with non-zero probability in μ.*

Note that there is a one-to-one correspondence between a face ζ_{a_S} of X^μ and a partial assignment $a_S \in \{0,1\}^{|S|}$ on a subset $S \subseteq [n]$ of binary variables. Hence we denote the links of X^μ by $X^\mu_{a_S}$ for $S \subseteq [n]$ and for $a_S \in \{0,1\}^{|S|}$.

Step (2) is left as an exercise.

Exercise 21.13 (Glauber Dynamics and Down-Up Walks)**.** *Verify that the down-up walk matrix $P^{\triangledown}_{n-1}$ on X^{μ} is exactly the transition matrix of Glauber dynamics on μ in Definition 21.10.*

Step (3) is to see that the correspondence between conditional correlation matrices of μ and random walk matrices of links of X^{μ}.

Lemma 21.14 (Correlation Matrices and Random Walk Matrices of Links)**.** *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution on subsets of $[n]$ and X^{μ} be the simplicial complex in Definition 21.12. For a partial assignment $a_S \in \{0,1\}^{|S|}$ on a subset $S \subseteq [n]$,*

$$W_{a_S} - J_{a_S} \preccurlyeq_{\Pi_0^{a_S}} \frac{1}{n-|S|-1}\big(J_{a_S} + \Psi_{a_S}\big),$$

where Ψ_{a_S} is the conditional correlation matrix in Definition 21.7, W_{a_S} is the random walk matrix of the link $X^{\mu}_{a_S}$ in Definition 19.12, and $J_{a_S} := \vec{1}(\Pi_0^{a_S})^T$ is defined as in Lemma 20.13.

Proof. We first check that each off-diagonal entry of LHS and RHS matches. Let a_S be a partial assignment and a_i, a_j be two bits for $i, j \notin S$. Then $a_{S\cup\{i\}}$ is used to denote the partial assignment on $S\cup\{i\}$ which extends a_S with the i-th variable being assigned a_i, and $a_{S\cup\{i,j\}}$ is defined analogously. By Definition 19.12 and Equation 19.1,

$$\begin{aligned} W_{a_S}\big((i,a_i),(j,a_j)\big) &= \frac{\Pi(a_{S\cup\{i,j\}})}{(|S|+2)\cdot\Pi(a_{S\cup\{i\}})} \\ &= \frac{\binom{n}{|S|+2}^{-1}\cdot \Pr_{Z\sim\mu}\left[Z\big(S\cup\{i,j\}\big) = a_{S\cup\{i,j\}}\right]}{(|S|+2)\cdot\binom{n}{|S|+1}^{-1}\cdot \Pr_{Z\sim\mu}\left[Z\big(S\cup\{i\}\big) = a_{S\cup\{i\}}\right]} \\ &= \frac{1}{n-|S|-1}\cdot \Pr_{Z\sim\mu}\left[Z(j) = a_j \mid Z(i) = a_i, Z(S) = a_S\right]. \end{aligned}$$

Similarly, by Equation 19.2,

$$J_{a_S}\big((i,a_i),(j,a_j)\big) = \Pi_0^{a_S}((j,a_j)) = \frac{\Pi(a_{S\cup\{j\}})}{(|S|+1)\cdot\Pi(a_S)} = \frac{1}{n-|S|}\cdot \Pr_{Z\sim\mu}\left[Z(j) = a_j \mid Z(S) = a_S\right].$$

This shows that the non-diagonal entries of Ψ_{a_S} and $(n-|S|-1)\cdot W_{a_S} - (n-|S|)\cdot J_{a_S}$ are the same. Rearranging and noting that the diagonal entries on RHS are bigger than that on LHS proves the statement. □

We are ready to prove step (4) and thus Theorem 21.11.

Proof of Theorem 21.11. The plan is to use the assumption that μ is η-spectrally independent (instead of local-spectral expansion of X^{μ}) to prove Proposition 21.1 with γ_{k-1} replaced by $\frac{\eta}{n-k-1}$, and then the theorem follows by plugging in $\gamma_{k-1} = \frac{\eta}{n-k-1}$ into Theorem 21.2 to obtain

$$\lambda_2(P^{\triangledown}_{n-1}) \le 1 - \frac{1}{n}\prod_{j=-1}^{n-3}(1-\gamma_j) = 1 - \frac{1}{n}\prod_{j=0}^{n-2}\Big(1 - \frac{\eta}{n-k-1}\Big).$$

To see Proposition 21.1 holds with $\gamma_{k-1} = \frac{\eta}{n-k-1}$, we use Lemma 21.14 in Equation 21.1 with $X_\tau = X_{a_S}$ so that

$$\begin{aligned}\langle f, (P_k^{\wedge} - P_k^{\triangledown}) f\rangle_{\Pi_k} &= \mathbb{E}_{\tau\sim\Pi_{k-1}}\langle f_\tau^{\perp}, (W_\tau - J_\tau) f_\tau^{\perp}\rangle_{\Pi_0^\tau} \\ &\le \mathbb{E}_{\tau\sim\Pi_{k-1}} \frac{1}{n-k-1}\langle f_\tau^{\perp}, (J_\tau + \Psi_\tau) f_\tau^{\perp}\rangle_{\Pi_0^\tau} \\ &\le \mathbb{E}_{\tau\sim\Pi_{k-1}} \frac{\eta}{n-k-1}\langle f_\tau^{\perp}, f_\tau^{\perp}\rangle_{\Pi_0^\tau},\end{aligned}$$

where the last inequality uses the assumption that $\lambda_{\max}(\Psi_\tau) \le \eta$ and $\langle f_\tau^{\perp}, J_\tau f_\tau\rangle = 0$. Then the rest of the proof of Proposition 21.1 is the same with γ_{k-1} replaced by $\frac{\eta}{n-k-1}$. □

To summarize, Theorem 21.11 can be seen as finding the corresponding simplicial complex so that Glauber dynamics is the same as down-up walks, and then interpreting the matrix $W_\tau - J_\tau$ in the proof of Proposition 21.1 as correlation matrices to define spectral independence. Spectral independence is a nice formulation so that probabilists do not need to know about high-dimensional expanders to use the result, and indeed this notion has led to many recent developments and we will discuss some in the next section.

21.3 Applications

In this section, we just briefly discuss some of the recent developments and point to the relevant references.

Sampling Independent Sets from Hardcore Distributions

The first major application of the spectral independence formulation is to prove fast mixing for sampling independent sets from the hardcore distribution.

Definition 21.15 (Hardcore Distributions)**.** *Given a graph $G = (V, E)$ and a parameter $\lambda > 0$, define the hardcore distribution $\mu_\lambda : \{0,1\}^{|V|} \to \mathbb{R}$ as $\mu_\lambda(S) = \lambda^{|S|}/Z_G(\lambda)$ for each independent set $S \subseteq V$, where*

$$Z_G(\lambda) := \sum_{S\subseteq V: S \text{ is an independent set}} \lambda^{|S|}$$

is the normalization constant called the partition function.

Estimating the partition function is a well-studied problem in statistical physics. Given a graph of maximum degree Δ, there is a critical threshold $\lambda(\Delta) = (\Delta-1)^{\Delta-1}/(\Delta-2)^{\Delta}$ called the "tree uniqueness threshold", where $\lambda < \lambda(\Delta)$ corresponds to the regime where the "influence" of a vertex u on another vertex v in the infinite Δ-regular tree decays exponentially fast in the distance between u and v.

The tree uniqueness threshold is about a mathematical property, but very interestingly this is also about computational complexity. A seminal work of Weitz showed that for any $\lambda < \lambda(\Delta)$, there is a deterministic fully polynomial time approximate scheme to estimate $Z_G(\lambda)$. Another seminal work of Sly proved that for any $\lambda > \lambda(\Delta)$, there is no such scheme to estimate $Z_G(\lambda)$ unless NP = RP. Both proofs connect explicitly the mathematical property to the computational complexity.

It was conjectured that the simple Glauber dynamics in Definition 21.10 for the hardcore distributions mixes in polynomial time whenever $\lambda < \lambda(\Delta)$. Anari, Liu and Oveis Gharan [ALO20] introduced spectral independence and used this notion to resolve the conjecture positively. Their proof uses the self-avoiding walk tree defined by Weitz to write a recurrence to bound the maximum row sum of the correlation matrices Ψ in Definition 21.5 to bound their maximum eigenvalue $\lambda_{\max}(\Psi)$ to apply Theorem 21.11 to conclude fast mixing. The proof is interesting and nontrivial which extends previous techniques in "correlation decay". One advantage of this randomized algorithm is that the dependency on Δ in the running time is much better than that of Weitz.

Sampling Graph Coloring from Glauber Dynamics

The work on spectral independence [ALO20] inspired many recent developments. One natural generalization is to sample from distributions $\mu : [k]^n \to \mathbb{R}$ where the variables are of larger arity. This class of distributions includes the problem of sampling a random graph coloring of a graph. The long standing major open problem for sampling graph coloring is that the simple Glauber dynamics as in Definition 21.10 mixes rapidly as long as the number of colors k is at least $\Delta + 2$ where Δ is the maximum degree of the input graph. Note that the Glauber dynamics may not be irreducible when $k \leq \Delta + 1$. The best known result is by Vigoda that the Glauber dynamic mixes in polynomial time as long as $k \geq 11\Delta/6$, so there is a very large gap between the upper bound and the lower bound.

The random graph coloring problem is very well-studied, where previous results are mostly based on the coupling techniques to prove fast mixing. Using spectral independence with "correlation decay" arguments, the previous results can be recovered [CGSV21, CLV21] with improved running time. Some of these results also rely on log-Sobolev inequalities and entropy techniques that we will study in later chapters. The main goal in this line of work is to use these new ideas originally from high-dimensional expanders to make progress on the long standing open problem about mixing time of Glauber dynamics for graph coloring.

Problem 21.16 (Simplicial Complex for Graph Coloring)**.** *Define a simplicial complex for graph coloring so that the down-up walk matrix corresponds exactly to the Glauber dynamics. Define the corresponding notion of spectral independence and compare to those defined in [CGSV21, CLV21].*

Coupling and Spectral Independence

A general question is how does the spectral independence method relate to other methods for proving fast mixing such as the most popular coupling techniques. Recent work in [Liu21, BCC+22] show that certain types of coupling proofs imply spectral independence as well, suggesting the spectral independence method could be a unifying method in analyzing mixing times of Markov chains.

21.4 References

[AASV21] Yeganeh Alimohammadi, Nima Anari, Kirankumar Shiragur, and Thuy-Duong Vuong. Fractionally log-concave and sector-stable polynomials: counting planar matchings and more. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 433–446. ACM, 2021. 202, 211, 212

[AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 1198–1211. ACM, 2020. 199, 200, 201

[ALO20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 1319–1330. IEEE, 2020. 202, 203, 206

[BCC+22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Stefankovic, and Eric Vigoda. On mixing of markov chains: Coupling, spectral independence, and entropy factorization. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692, 2022. 206

[CGSV21] Zongchen Chen, Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1548–1557. SIAM, 2021. 202, 206

[CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of glauber dynamics: entropy factorization via high-dimensional expansion. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1537–1550. ACM, 2021. 206

[Liu21] Kuikui Liu. From coupling to spectral independence and blackbox comparison with the down-up walk. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2021, August 16-18, 2021, University of Washington, Seattle, Washington, USA (Virtual Conference)*, volume 207 of *LIPIcs*, pages 32:1–32:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. 206

Chapter 22

Log-Concave Polynomials

In this chapter, we take a different perspective to view a 0-local-spectral expander as a strongly log-concave polynomial. Then we see two related notions of polynomials, sector-stable polynomials and fractionally log-concave polynomials, and their connections to spectral independence.

22.1 Log-Concave Polynomials

The polynomial approach is actually the original approach that was used to solve the matroid expansion conjecture [ALOV19] that we saw in Chapter 20.

Definition 22.1 (Strongly Log-Concave Distribution)**.** *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution and $g_\mu(x) = \sum_{S\subseteq[n]} \mu(S) \cdot \prod_{i\in S} x_i$ be its generating polynomial as defined in Definition 16.1. We say μ is a log-concave distribution if $\log g_\mu$ is a concave function at the point $\vec{1}$.*

We say μ is a strongly log-concave distribution if for any $k \geq 0$ and any sequence of integers $1 \leq i_1, \ldots, i_k \leq n$,

$$(\partial_{x_{i_1}} \cdots \partial_{x_{i_k}} g_\mu)(x_1, \ldots, x_n)$$

is log-concave at the point $\vec{1}$.

In Chapter 19, given a d-homogeneous probability distribution μ, we use μ to define a pure $(d-1)$-dimensional weighted simplicial complex (X_μ, Π) as in Definition 19.5, with $\Pi_{d-1} := \mu$ being the distribution on the maximal faces of dimension $d-1$. In this chapter, give a d-homogeneous probability distribution, we use μ to define a d-homogeneous generating polynomial g_μ as in Definition 22.1. The connection between the weighted simplicial complex (X_μ, Π) and the generating polynomial g_μ is through the Hessian matrix of $\log g_\mu$ at the point $\vec{1}$.

Exercise 22.2 (Hessian Matrix of a Polynomial)**.** *The Hessian matrix of $\log p$ is*

$$\nabla^2 \log p = \frac{p \cdot (\nabla^2 p) - (\nabla p)(\nabla p)^T}{p^2}.$$

A basic result in convex analysis is that $\log p$ is concave at a point x if and only if $\nabla^2 \log p$ is negative semidefinite at x.

The main observation in [ALOV19] is that the Hessian matrix of g_μ is closely related to the random walk matrix of the empty link of X_μ.

Theorem 22.3 (Strongly Log-Concave Polynomial and 0-Local-Spectral Expander [ALOV19])**.** *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution, g_μ be its generating polynomial as in Definition 22.1, and X_μ be its weighted simplicial complex as in Definition 19.5. Then g_μ is strongly log-concave if and only if X_μ is a 0-local-spectral expander.*

Proof. The main step is to show that the "random walk matrix" of the Hessian matrix of g_μ at point $\vec{1}$ and the random walk matrix of the empty link of X_μ are the same. On one hand, let $H := \nabla^2 g_\mu|_{x=\vec{1}}$ be the Hessian matrix of g_μ at the point $\vec{1}$. Note that $g_\mu|_{x=\vec{1}} = 1$,

$$\big(\nabla g_\mu\big)\big|_{x=\vec{1}}(i) = \partial_{x_i} g_\mu\big|_{x=\vec{1}} = \Pr_{S\sim\mu}\big[i \in S\big] \quad \text{and} \quad \big(\nabla^2 g_\mu\big)\big|_{x=\vec{1}}(i,j) = \partial_{x_i}\partial_{x_j} g_\mu\big|_{x=\vec{1}} = \Pr_{S\sim\mu}\big[\{i,j\} \subseteq S\big].$$

Let D_H be the degree matrix of H. Then, as μ is d-homogeneous, for any $i \in [n]$,

$$D_H(i,i) = \sum_{j=1}^n H(i,j) = \sum_{j=1}^n \Pr_{S\sim\mu}\big[\{i,j\} \subseteq S\big] = (d-1)\Pr_{S\sim\mu}\big[i \in S\big].$$

On the other hand, let the random walk matrix of the empty link be W. For $i \neq j \in [n]$, by Definition 19.12 and Definition 19.6,

$$W(i,j) = \frac{\Pi\big(\{i,j\}\big)}{2\Pi\big(\{i\}\big)} = \frac{\binom{d}{2}^{-1} \cdot \Pr_{S\sim\mu}\big[\{i,j\} \subseteq S\big]}{2\binom{d}{1}^{-1} \cdot \Pr_{S\sim\mu}\big[i \in S\big]} = \frac{\Pr_{S\sim\mu}\big[\{i,j\} \subseteq S\big]}{(d-1)\cdot \Pr_{S\sim\mu}\big[i \in S\big]}.$$

Therefore,

$$W = D_H^{-1} H.$$

With this identity, we next show that g_μ is log-concave if and only if $\lambda_2(W) \leq 0$. In one direction, if g_μ is log-concave, then $\nabla^2 \log g_\mu$ is negative semidefinite by Definition 22.1. Note that this implies that $H = \nabla^2 g_\mu$ has at most one positive eigenvalue by the identity in Exercise 22.2. Check that it follows that $W = D_H^{-1} H$ also has at most one positive eigenvalue, and thus $\lambda_2(W) \leq 0$.

In the other direction, if W has at most one positive eigenvalue, then note that W has exactly one positive eigenvalue, as $\vec{1}$ is an eigenvector of W with eigenvalue 1. Check that this implies that $\lambda_1(W - \vec{1}\Pi_0^T) \leq 0$, and then it follows that $D_H(W - \vec{1}\Pi_0^T) = H - \frac{d-1}{d}(\nabla g_\mu)(\nabla g_\mu)^T$ is negative semidefinite. By the identity in Exercise 22.2, this implies that $\nabla^2 \log p$ is negative semidefinite.

Finally, observe that there is a one-to-one correspondence between the differentiated polynomials $\big(\partial_{x_{i_1}} \cdots \partial_{x_{i_k}} g_\mu\big)(x_1, \ldots, x_n)$ and the links $X_{\{i_1,\ldots,i_k\}}$ of the simplicial complex. Thus, by the arguments above, the differentiated polynomial at point $\vec{1}$ is log-concave if and only if the second eigenvalue of the random walk matrix of the corresponding link is at most 0. Therefore, we conclude that g_μ is strongly log-concave if and only if X_μ is a 0-local-spectral expander. □

Corollary 22.4 (Matroid Polynomial is Strongly Log-Concave)**.** *The generating polynomial of the uniform distribution on matroid bases is strongly log-concave.*

This result was proved earlier in [AOV18] using advanced techniques from Hodge theory for matroids [AHK18], so the techniques from high-dimensional expanders provide a more elementary and simpler proof. It is very interesting to see a correspondence between the concepts in polynomials and the concepts in high-dimensional expanders (see [ALOV19] for more).

Mason's Ultra Log-Concavity Conjecture

An important consequence of the polynomial perspective is a proof of the conjecture that the rank sequence of a matroid is ultra log-concave.

Theorem 22.5 (Mason's Conjecture [ALOV18, BH20]). *For a matroid M on n elements with m_k independent sets of size k, the sequence $m_0, m_1, \ldots, m_n$ is ultra log-concave such that for $1 < k < n$,*

$$\left(\frac{m_k}{\binom{n}{k}}\right)^2 \geq \frac{m_{k-1}}{\binom{n}{k-1}} \cdot \frac{m_{k+1}}{\binom{n}{k+1}}.$$

The question that whether the sequence $m_0, m_1, \ldots, m_n$ is log-concave was a long standing open problem in combinatorics from the 70s, and was first proved in [AHK18] using Hodge theory for matroids. The proof of the stronger ultra log-concavity in [ALOV18, Ove20] is short and elementary, and should be readily understandable for readers who followed the course thus far. There is another proof of ultra log-concavity in [BH20] using a closely related notion called Lorentzian polynomials. We remark that Gurvits also studied log-concave polynomials earlier in his work on generalizations of permanent problems.

22.2 Sector-Stable Polynomials

Given that the matroid expansion conjecture can be solved from both the high-dimensional expander perspective and the strongly log-concave polynomial perspective, and that the high-dimensional expander approach can be extended further as in Chapter 21, one may wonder whether the polynomial approach can also be extended further and possibly in different directions.

A very interesting recent paper by Alimohammadi, Anari, Shirajur, Vuong [AASV21] proposed two notions for polynomials called sector-stability and fractional log-concavity. We discuss sector-stable polynomials in this section, and fractional log-concave polynomials in the next section.

Definition 22.6 (Sector-Stable Polynomials). *The open sector of aperature $\alpha\pi$ centered around the positive real axix is denoted by*

$$\Gamma_\alpha := \left\{ \exp(x + \imath y) \mid x \in \mathbb{R}, y \in \left(-\frac{\alpha\pi}{2}, \frac{\alpha\pi}{2} \right) \right\}$$

A polynomial $g(z_1, \ldots, z_n)$ is Γ_α-stable if

$$z_1, \ldots, z_n \in \Gamma_\alpha \quad \Longrightarrow \quad g(z_1, \ldots, z_n) \neq 0.$$

Note that Γ_1 is the right half-plane, and Γ_1-stability is called Hurwitz-stability. The following exercise shows that it is a generalization of real-stable polynomials for homogeneous polynomials.

Exercise 22.7. *Show that a homogeneous polynomial is Hurwitz-stable if and only if it is $\mathcal{H}$-stable in Definition 13.8.*

A very interesting theorem in [AASV21] is a connection between sector-stability and spectral independence. The proof is very nice and elegant, using some elementary complex analysis.

Theorem 22.8 (Sector-Stability Implies Spectral Independence). *Suppose that $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is a probability distribution whose generating polynomial is Γ_α-stable. Then the ℓ_1-norm of any row in the correlation matrix Ψ in Definition 21.5 is bounded by*

$$\sum_j \left| \Pr_{S\sim\mu}[j \in S \mid i \in S] - \Pr_{S\sim\mu}[j \in S] \right| \leq \frac{2}{\alpha}.$$

A consequence is that $\lambda_{\max}(\Psi) \leq \frac{2}{\alpha}$.

Sector Stability Preserving Operations

Before we see some examples of sector-stable polynomials, we first record some sector-stability preserving operations from [AASV21].

Exercise 22.9 (Sector-Stability Preserving Operations). *Show that the following operations preserve Γ_α-sector stability.*

1. *Specialization: $g(z_1, \ldots, z_n) \to g(a, z_2, \ldots, z_n)$, where $a \in \bar{\Gamma}_\alpha$.*
2. *Scaling: $g(z_1, \ldots, z_n) \to g(\lambda_1 z_1, \lambda_2 z_2, \ldots, \lambda_n z_n)$, where $\lambda_i \geq 0$ for $1 \leq i \leq n$.*
3. *Dual: $g \to g^*$, where $g(z) = \sum_{S\subseteq[n]} c_S \cdot z^S$ and $g^*(z_1, \ldots, z_n) := \sum_{S\subseteq[n]} c_S \cdot z^{[n]\setminus S}$.*

Exercise 22.10 (Homogenization). *If multi-affine polynomial $g(z_1, \ldots, z_n) := \sum_{S\subseteq[n]} c_S \cdot z^S$ is Γ_α-stable, then its homogenization*

$$g^{\text{hom}}(z_1, \ldots, z_n, w_1, \ldots, w_n) := \sum_{S\subseteq[n]} c_S \cdot z^S \cdot w^{[n]\setminus S}$$

is multi-affine, homogeneous of degree n, and $\Gamma_{\alpha/2}$-stable.

Proposition 22.11 (Partial Derivative). *If $g(z_1, \ldots, z_n)$ is a multiaffine polynomial, then $\partial_{z_i} g$ is sector stable for $1 \leq i \leq n$.*

Theorem 22.12 (Truncation). *If $g(z_1, \ldots, z_n)$ is Γ_1-stable, then g_k is either identically zero or $\Gamma_{1/2}$-stable, where g_k is the truncation of g that keeps only the degree k terms.*

Applications

One application in [AASV21] is to sample matchings of a given size in planar graphs. The starting point is a theorem by Heilman and Lieb.

Theorem 22.13 (Monomer-Dimer Polynomial). *Given a graph $G = (V, E)$ with edge weight w_e for $e \in E$, the polynomial*

$$\sum_{M\subseteq E: M \text{matching}} \prod_{e\in M} w(e) \prod_{v: v\notin M} z_v$$

is Γ_1-stable, where $\{z_v\}_{v\in V}$ are the variables in this polynomial.

This polynomial is not homogeneous, and homogenization does not preserve Hurwitz-stabiliy, nor truncation to matchings of a given size. Applying Theorem 22.12, however, they can say that the truncation to matchings of a given size is still $\Gamma_{1/2}$-stable, so that they can apply Theorem 22.8 to prove that the Markov chain on the set of "monomers" is fast mixing. For any class of graphs that counting matchings is polynomial time solvable, including planar graphs and bounded genus graphs, their results can be used to approximate sample and count matchings of a given size.

Another application is in non-asymmetric determinantal point process.

Theorem 22.14 (Non-Symmetric k-DPPs). *For any matrix $L \in \mathbb{R}^{n\times n}$ satisfying $L + L^T \succcurlyeq 0$ and a number k, the polynomial*

$$g(z_1, \ldots, z_n) = \sum_{S\in\binom{[n]}{k}} \det(L_{S,S}) \prod_{i\in S} z_i$$

is $\Gamma_{1/2}$-stable

The following result shows some limitation on the class of sector-stable polynomials for combinatorial problems.

Lemma 22.15 (Bounded Length of Sector-Stable Distributions). *If $\mu : \{0,1\}^n \to \mathbb{R}$ is a $\Gamma_{1/k}$-sector-stable distribution, then the length of edges of its Newton polytope* $\mathrm{newt}(\mu)$ *is at most $2k$, where*

$$\mathrm{newt}(\mu) := \mathrm{conv}\big(\{S : \mu(S) > 0\}\big).$$

This result shows that if the polytope of the combinatorial problem has unbounded edge length (such as the matching polytope and the arborescence polytope), then the corresponding generating polynomial cannot be sector-stable.

22.3 Fractionally Log-Concave Polynomials

The class of fractionally log-concave polynomials is a generalization of the class of log-concave polynomials.

Definition 22.16 (Fractionally Log-Concave Polynomials). *A polynomial $g_\mu(z_1, \ldots, z_n)$ is called α-fractionally log-concave for $\alpha \in [0,1]$ if $\log g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is concave when viewed as a function over $\mathbb{R}^n_{\geq 0}$.*

The key observation is that the Hessian matrix of $\log g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ at the point $\vec{1}$ is closely related to the correlation matrix in Definition 21.5. The proof is similar to that in Exercise 22.2 and Theorem 22.3.

Proposition 22.17 (α-Fractionally Log-Concavity and Spectral Independence). *Let $\mu : \{0,1\}^n \to \mathbb{R}$ be a probability distribution and g_μ be its generating polynomial. Then g_μ is α-fractionally log-concave at the point $\vec{1}$ if and only if $\lambda_{\max}(\Psi) \leq \frac{1}{\alpha}$ where Ψ is the correlation matrix of μ as defined in Definition 21.5.*

Proof. Let $H := \nabla^2 \log g_\mu(z_1^\alpha, \ldots, z_n^\alpha)|_{z=\vec{1}}$. Let $\Pr[i] := \Pr_{S\sim\mu}[i \in S]$ and similarly $\Pr[i \wedge j] := \Pr_{S\sim\mu}[i \in S \wedge j \in S]$. Check that a similar calculation as in Exercise 22.2 gives

$$H_{i,j} = \begin{cases} \alpha(\alpha-1)\Pr[i] - \alpha^2 \Pr[i]^2 & \text{if } i = j \\ \alpha^2\big(\Pr[i \wedge j] - \Pr[i]\cdot\Pr[j]\big) & \text{if } i \neq j \end{cases}$$

Let $D := \mathrm{diag}(\Pr[i])$ be the diagonal matrix of marginal probability. It follows from Definition 21.5 that

$$\Psi = \frac{1}{\alpha^2} D^{-1} H + \frac{1}{\alpha} I.$$

This implies that $\lambda_{\max}(\Psi) \leq \frac{1}{\alpha}$ if and only if $\lambda_{\max}(D^{-1}H) \leq 0$ if and only if $H \preccurlyeq 0$ if and only if g_μ is α-fractionally log-concave at the point $\vec{1}$. □

Recall from Theorem 18.11 that a homogeneous real-stable polynomial is log-concave, and thus log-concavity is a generalization of real-stability that does not involve root locations. Using Proposition 22.17 and Theorem 22.8, we see that fractionally log-concavity is a generalization of sector-stability that does not involve root locations.

Theorem 22.18 (Sector-Stability Implies Fractionally Log-Concavity)**.** *For $\alpha \in [0, \frac{1}{2}]$, if g_μ is $\Gamma_{2\alpha}$-stable, then g_μ is α-fractionally log-concave.*

Proof. Theorem 22.8 proves that $\Gamma_{2\alpha}$-stability of g_μ implies that $\lambda_{\max}(\Psi) \leq \frac{1}{\alpha}$, and thus implies that g_μ is α-fractionally log-concave at the point $\vec{1}$ by Proposition 22.17.

Note that sector-stability is preserved under the change of variables $z_i \to \lambda_i z_i$ when $\lambda_1, \ldots, \lambda_n$ are positive reals by Exercise 22.9. This allows us to map any point in $\mathbb{R}^n_{\geq 0}$ to the point $\vec{1}$, and to use the above argument to show that g_μ is α-fractionally log-concave at any point in $\mathbb{R}^n_{\geq 0}$. □

While fractional log-concavity at the point $\vec{1}$ is equivalent to a bound on the eigenvalues of the correlation matrix Ψ, it does not imply a bound for the conditioned distributions μ_S. However, fractional log-concavity at all points in $\mathbb{R}^n_{\geq 0}$ does, because the polynomial for conditional distributions μ_S can be obtained as the following limit:

$$g_{\mu_S} \propto \lim_{\lambda\to\infty} \frac{g_\mu\left(\lambda z_1, \ldots, \lambda z_{|S|}, \ldots, z_n\right)}{\lambda^{|S|}}.$$

Scaling the variables or the polynomial, and taking limits all preserve fractional log-concavity.

Theorem 22.19 (Fractional Log-Concavity Implies Spectral Independence)**.** *If $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ has a α-fractionally log-concave generating polynomial, then the correlation matrix of every conditioned distribution μ_S has maximum eigenvalue $\frac{1}{\alpha}$. It follows that μ is $\frac{1}{\alpha}$-spectrally independent as defined in Definition 21.8.*

22.4 References

[AASV21] Yeganeh Alimohammadi, Nima Anari, Kirankumar Shiragur, and Thuy-Duong Vuong. Fractionally log-concave and sector-stable polynomials: counting planar matchings and more. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 433–446. ACM, 2021. 202, 211, 212

[AHK18] Karim Adiprasito, June Huh, and Eric Katz. Hodge theory for combinatorial geometries. *Annals of Mathematics*, 188(2):381–452, 2018. 210, 211

[ALOV18] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials III: mason's ultra-log-concavity conjecture for independent sets of matroids. *CoRR*, abs/1811.01600, 2018. 211

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1–12. ACM, 2019. 184, 198, 209, 210

[AOV18] Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 35–46. IEEE Computer Society, 2018. 210

[BH20] Petter Brändén and June Huh. Lorentzian polynomials. *Annals of Mathematics*, 192(3):821–891, 2020. 211

[Ove20] Shayan Oveis Gharan. Course notes of "Polynomial paradigm in algorithm design", lecture 10. 2020. https://homes.cs.washington.edu/~shayan/courses/polynomials/. 128, 130, 153, 158, 171, 173, 175, 179, 187, 211

Chapter 23

Log-Sobolev Inequalities

We introduce log-Sobolev inequalities for analyzing mixing time of random walks, and see that they provide the optimal bound for the down-up walks on strongly log-concave distributions.

23.1 Analyzing Mixing Time Using Log-Sobolev Inequalities

In this section, we first define variance and relative entropy, and then define spectral gap and log-Sobolev constants, and then see their uses in bounding mixing time, and finally some intuition about these definitions. The presentation in this section is based on [CGM21, AJK+21, BT06, MT06].

Variance and Entropy

In Chapter 6, when we analyze the mixing time of random walks, we upper bound the total variation distance $d_{\mathrm{TV}}(p_t, \pi)$ by upper bounding $\|D^{-1/2}(p_t - \pi)\|_2$ (see Problem 6.20). This can be understood as bounding the variance of $f := p_t/\pi$, the density of p_t with respect to π at time $t \geq 0$.

Definition 23.1 (π-Variance)**.** *Let $f : [n] \to \mathbb{R}$ be a function and π be a probability distribution on $[n]$. The variance of f with respect to π is defined as*

$$\mathrm{Var}_\pi[f] := \mathbb{E}_\pi[f^2] - \big(\mathbb{E}_\pi[f]\big)^2,$$

where $\mathbb{E}_\pi[f] = \sum_{i \in [n]} \pi(i) f(i)$.

There are other ways to measure the closeness of two probability distributions. A well-known measure is the relative entropy between the two distributions.

Definition 23.2 (KL-Divergence)**.** *Let p and q be probability distributions on $[n]$ such that $q(i) = 0$ implies $p(i) = 0$ for $1 \leq i \leq n$. The Kullback-Liebler divergence, or relative entropy, between p and q is defined as*

$$\mathcal{D}_{\mathrm{KL}}(p \parallel q) = \sum_{i=1}^{n} p(i) \log \frac{p(i)}{q(i)},$$

where we follow the convention that $0 \log 0 = 0$. Check that $\mathcal{D}_{\mathrm{KL}}(p, q) \geq 0$ by Jensen's inequality.

Pinsker's inequality shows that KL-divergence can be used to upper bound the total variation distance.

Theorem 23.3 (Pinsker's Inequality). *For any two probability distributions p, q on $[n]$,*

$$d_{\mathrm{TV}}(p,q)^2 \le 2\mathcal{D}_{\mathrm{KL}}(p \parallel q).$$

So, to analyze mixing time, it is also natural to consider the relative entropy betwen p_t and π.

Definition 23.4 (π-Entropy). *Let $f : [n] \to \mathbb{R}$ be a function and π be a probability distributions on $[n]$. Define*

$$\mathrm{Ent}_\pi[f] := \mathbb{E}_\pi[f \log f] - \mathbb{E}_\pi\big[f \log(\mathbb{E}_\pi[f])\big].$$

Check that $\mathrm{Ent}_\pi[\frac{p}{\pi}] = \mathcal{D}_{\mathrm{KL}}(p \parallel \pi)$ for a probability distribution p.

Spectral Gap and Log-Sobolev Constants

We only consider reversible Markov chains in this course, which include transition matrices of random walks on undirected graphs.

Definition 23.5 (Reversible Markov Chain). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a Markov chain whose stationary distribution is π. We say P is reversible if for all $i, j \in [n]$,*

$$\pi(i) \cdot P(i,j) = \pi(j) \cdot P(j,i)$$

Let $\Pi := \mathrm{diag}(\pi)$. Then the reversible condition can be stated as ΠP being a symmetric matrix.

The following definition should be understood as the quadratic form of the Laplacian matrix when we consider random walks on undirected graphs.

Definition 23.6 (Dirichlet Form). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. For two vectors $f, g \in \mathbb{R}^n$, the Dirichlet form is defined as*

$$\mathcal{E}_P(f,g) := \langle (I-P)f, g\rangle_\pi = g^T(\Pi - \Pi P)f = \frac{1}{2}\sum_{1\le i,j\le n} \pi(i)\cdot P(i,j)\cdot \big(g(i)-g(j)\big)\cdot\big(f(i)-f(j)\big),$$

where $\Pi := \mathrm{diag}(\pi)$. The last equality can be seen by thinking of $\Pi - \Pi P$ as the Laplacian matrix of the underlying undirected graph with an edge weight $\Pi(i)\cdot P(i,j)$ for each pair $i, j \in [n]$.

Remark 23.7 (Laplacian?). *The matrix $I - P$ is often called the Laplacian in the literature, but we resist not to do so as it is not consistent with our convention (e.g. $I - P$ may not be symmetric).*

For random walks on an undirected graph G, the adjacency matrix of G is ΠP, and note that $I - P$ has the same spectrum as the normalized *Laplacian matrix of G.*

The Dirichlet form is sometimes called the energy of the function f, which can be thought of as a measure of the *local* variation of f along the edges of the underlying graph. On the other hand, the variance in Definition 23.1 can be thought of as a measure of the *global* variation of f. Then the spectral gap can be interpreted as a lower bound on the local variance by the global variance, and this perspective is useful in designing approximation algorithms on expander graphs (e.g. [BRS11]).

Definition 23.8 (Variational Characterization of Spectral Gap). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Define*

$$\lambda(P) := \inf\left\{ \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)} \;\middle|\; f : [n] \to \mathbb{R},\ \mathrm{Var}_\pi[f] \neq 0 \right\}.$$

Exercise 23.9. *Show that $\lambda(P) = 1 - \alpha_2(P)$ where $\alpha_2(P)$ is the second largest eigenvalue of P.*

We note that the spectral gap is sometimes called the Poincaré constant. The log-Sobolev constant replaces the variance of f in the denominator of the spectral gap by the π-entropy of f in Definition 23.4.

Definition 23.10 (Log-Sobolev Constant [DSC96])**.** *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Define the log-Sobolev constant of P as*

$$\alpha(P) := \inf\left\{ \frac{\mathcal{E}_P(f,f)}{\mathrm{Ent}_\pi(f^2)} \;\middle|\; f : [n] \to \mathbb{R}_{\geq 0},\ \mathrm{Ent}_\pi[f^2] \neq 0 \right\}.$$

The modified log-Sobolev constant is introduced by Bobkov and Tetali [BT06].

Definition 23.11 (Modified Log-Sobolev Constant [BT06])**.** *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Define the modified log-Sobolev constant of P as*

$$\rho(P) := \inf\left\{ \frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}_\pi(f)} \;\middle|\; f : [n] \to \mathbb{R}_{\geq 0},\ \mathrm{Ent}_\pi[f] \neq 0 \right\}.$$

These definitions may not look intuitive. In the following, we will first state the results of using spectral gap and log-Sobolev constants to bounding mixing time, and then we will provide some intuitions about these definitions.

Bounding Mixing Time by Log-Sobolev Constants

We have already seen in Chapter 6 that we can upper bound the mixing time by lower bounding the spectral gap. The following is a generalization of Theorem 6.16.

Theorem 23.12 (Mixing Time by Spectral Gap)**.** *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. The ϵ-mixing time in Definition 6.15 is*

$$\tau_\epsilon(P) \lesssim \frac{1}{\lambda(P)}\left(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\epsilon}\right),$$

where $\pi_{\min} := \min_{i\in[n]} \pi(i)$ (which is $\frac{1}{n}$ when π is the uniform distribution).

The significance of the log-Sobolov constant is a much better dependence on $1/\pi_{\min}$. The following result is proved by Diaconis and Saloff-Coste [DSC96].

Theorem 23.13 (Mixing Time by Log-Sobolev Constant [DSC96])**.** *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Then*

$$\tau_\epsilon(P) \lesssim \frac{1}{\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{\epsilon}\right).$$

Bobkov and Tetali [BT06] proved a similar result for modified log-Sobolev constant.

Theorem 23.14 (Mixing Time by Modified Log-Sobolev Constant [BT06]). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Then*

$$\tau_\epsilon(P) \lesssim \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{\epsilon}\right).$$

Bobkov and Tetali also proved that

$$2\lambda(P) \geq \rho(P) \geq 4\alpha(P),$$

and so the lower bounds on these constants are increasingly difficult to obtain. The modified log-Sobolev constant has the advantage that it provides the same upper bound on the mixing time, while it is always at least as large as the log-Sobolev constant.

Intuition from Continuous Time Random Walks

The definitions of the spectral gap and the modified log-Sobolev constant come quite naturally from continuous time random walks. We would not be able to introduce continuous time random walks properly, so we just state the definition.

Definition 23.15 (Continuous Time Random Walks). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. For any $t \geq 0$, the transition matrix, or the heat kernel, is defined as*

$$H_t = e^{-t(I-P)} = \sum_{k=0}^{\infty}\frac{t^k(P-I)^k}{k!}.$$

Let $p_0 \in \mathbb{R}^n$ be an initial distribution. Then $p_t^T = p_0^T H_t$ is the distribution at time t.

As discussed earlier, we will consider $f_t := p_t/\pi$ and keep track of how fast it converges to $\vec{1}$.

Exercise 23.16 (Change of Density). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Let $f_t(i) = p_t(i)/\pi(i)$ for all $i \in [n]$ be the density of p_t with respect to π at time $t \geq 0$. For any initial distribution p_0 and all $t \geq 0$, show that*

$$f_t = H_t f_0.$$

Furthermore, for any $i \in [n]$, show that

$$\frac{df_t(i)}{dt} = \big((P-I)f_t\big)(i).$$

It turns out that the change of variance is exactly the Dirichlet form.

Lemma 23.17 (Change of Variance). *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Let $f_t = p_t/\pi$. Then*

$$\frac{d}{dt}\mathrm{Var}_\pi(f_t) = -2\mathcal{E}_P(f_t, f_t).$$

Proof. Note that $\mathrm{Var}_\pi(f_t) = \mathbb{E}_\pi[f_t^2] - 1$ by Definition 23.1, and so

$$\frac{d}{dt}\mathrm{Var}_\pi(f_t) = \sum_{i=1}^n \pi(i) \cdot \frac{d}{dt} f_t(i)^2 = 2\sum_{i=1}^n \pi(i) \cdot f_t(i) \cdot \big((P-I)f_t\big)(i) = -2\mathcal{E}_P(f_t, f_t).$$

□

So we can understand the spectral gap in Definition 23.8 is *defined* to ensure that

$$\frac{d}{dt}\mathrm{Var}_\pi(f_t) = -2\mathcal{E}_P(f_t, f_t) \le -2\lambda(P) \cdot \mathrm{Var}_\pi(f_t) \quad \Longrightarrow \quad \frac{d}{dt}\log\big(\mathrm{Var}_\pi(f_t)\big) \le -2\lambda(P).$$

By integrating on both sides, we see that the variance is exponentially decreasing as

$$\log\big(\mathrm{Var}_\pi(f_t)\big) - \log\big(\mathrm{Var}_\pi(f_0)\big) \le -2\lambda(P) \cdot t \quad \Longrightarrow \quad \mathrm{Var}_\pi(f_t) \le \mathrm{Var}_\pi(f_0) \cdot e^{-2\lambda(P)\cdot t}.$$

Note that the initial variance $\mathrm{Var}_\pi(f_0) \le 1/\pi_{\min}$, and this implies Theorem 23.12 for continuous time random walks. This argument can be adapted for discrete time random walks. This is a good exercise to work out; see [MT06] for a solution.

Bobkov and Tetali used the same logic to define the modified log-Sobolev constant, by using relative entropy in place of variance.

Lemma 23.18 (Change of Relative Entropy)**.** *Let $P \in \mathbb{R}^{n\times n}$ be the transition matrix of a reversible Markov chain whose stationary distribution is π. Let $f_t = p_t/\pi$. Then*

$$\frac{d}{dt}\mathrm{Ent}_\pi(f_t) = -\mathcal{E}_P(f_t, \log f_t).$$

Proof. Note that $\mathbb{E}_\pi[f] = 1$ and thus $\mathrm{Ent}_\pi[f] = \mathbb{E}_\pi[f \log f]$ by Definition 23.4, and hence

$$\frac{d}{dt}\mathrm{Ent}_\pi(f_t) = \sum_{i=1}^n \pi(i) \cdot \frac{d}{dt} f_t(i) \log f_t(i) = \sum_{i=1}^n \pi(i) \cdot \big(1 + \log f_t(i)\big) \cdot \big((P-I)f_t\big)(i) = -\mathcal{E}_P(f_t, \log f_t),$$

where in the last equality we use that $\sum_{i=1}^n \pi(i)((P-I)f_t)(i) = \langle \pi, (P-I)f_t\rangle = 0$. □

So the modified log-Sobolev constant is *defined* to ensure that

$$\frac{d}{dt}\mathrm{Ent}_\pi(f_t) = -\mathcal{E}_P(f_t, \log f_t) \le -\rho(P) \cdot \mathrm{Ent}_\pi(f_t) \quad \Longrightarrow \quad \mathrm{Ent}_\pi(f_t) \le \mathrm{Ent}_\pi(f_0) \cdot e^{-\rho(P)\cdot t}.$$

Crucially, the initial relative entropy is

$$\mathrm{Ent}_\pi[f_0] = \sum_{i=1}^n p_0(i) \log \frac{p_0(i)}{\pi(i)} \le \log \frac{1}{\pi_{\min}},$$

and this implies Theorem 23.14 for continuous time random walks by Pinsker's inequality in Theorem 23.3.

I have not found a general proof of Theorem 23.14 for discrete time random walks, and it was mentioned in [MT06] that "there seems to be no discrete-time analog" of it.

In the combinatorial applications that we will see, however, there are direct proofs of the exponential decreasing of the relative entropy, and thus the mixing time bound in Theorem 23.14 holds.

23.2 Log-Sobolev Constant for Strongly Log-Concave Distribution

It is already difficult to prove a lower bound on the spectral gap, and so there are very few known result on proving a lower bound on the log-Sobolev constants. This is starting to change after the resolution of the matroid expansion conjecture. Not only do the techniques from high dimensional expanders provide a direct way to establish a lower bound on the spectral gap, recent developments extend the techniques further to establish a lower bound on the modified log-Sobolev constant. The first result in this direction is by Cryan, Guo and Mousa [CGM21].

Theorem 23.19 (Modified Log-Sobolev Constant for Strongly Log-Concave Distribution [CGM21])**.** *Let μ be a d-homogeneous strongly log-concave distribution. Then the modified log-Sobolev constant of the down-up walk $P_d^{\triangledown}$ in Definition 20.5 is*

$$\rho(P_d^{\triangledown}) \geq \frac{1}{d}.$$

The proof in [CGM21] is very nice, but we will not present it here. Rather, we will present a recent generalization for fractionally log-concave distributions in the next chapter. We just note here that the proof in [CGM21] shows that the relative entropy is exponentially decreasing after one step of the down-up walk such that

$$\mathcal{D}_{\mathrm{KL}}(P_d^{\triangledown} p \parallel \pi) \leq \left(1 - \frac{1}{d}\right) \cdot \mathcal{D}_{\mathrm{KL}}(p \parallel \pi),$$

and using Pinsker's inequality as in the previous section gives the optimal mixing time analysis for the down-up walk on matroid bases.

Corollary 23.20 (Optimal Mixing Time for Sampling Matroid Bases [CGM21])**.** *The mixing time of the down-up walk in Chapter 20 for sampling uniform random matroid bases of size d is*

$$\tau_\epsilon(P_d^{\triangledown}) \lesssim d\Big(\log d + \log\log n + \log\frac{1}{\epsilon}\Big).$$

Near-Linear Time Algorithm for Random Spanning Trees

One immediate consequence of Corollary 23.20 is that the mixing time of the down-up walk for sampling uniform random spanning trees is at most $O(n \log n)$. To design a near-linear time algorithm, one needs to implement each iteration in the down-up walk efficiently, but it is not known how to do so.

Fortunately, the trick in [ALO+21] is to consider the down-up walk on the *dual* matroid. Given a graph $G = (V, E)$, the rank of the dual matroid is $|E| - |V| + 1 \leq |E|$, so the mixing time of the down-up walk on the dual matroid $O(|E| \log \frac{|E|}{\epsilon})$ by Corollary 23.20. The resulting algorithm is as follows. Let T_0 be an arbitrary spanning tree. In iteration $t \geq 0$, sample a uniform random edge $e \in E - T_t$, and then sample a uniform random edge f in the unique cycle in $T_t + e$ and set $T_{t+1} := T_t + e - f$, and repeat. This algorithm has been studied by Russo, Teixeira and Francisco, and they show that each iteration can be implemented in amortized $O(\log |E|)$ time using the cut-link trees data structures.

Theorem 23.21 (Near-Linear Time Algorithm for Sampling Random Spanning Trees [ALO+21])**.** *Given a graph $G = (V, E)$, there is an algorithm to sample a random spanning tree in G with distribution ϵ-close to the uniform distribution and running time $O(|E| \log |E| \log \frac{|E|}{\epsilon})$.*

The problem of designing a fast algorithm for sampling a uniform random spanning tree is well-studied. The previous best known algorithm is by Schild with almost-linear running time $O(m^{1+o(1)})$. This is based on a line of work that simulated another Markov chain for generating random spanning trees, using techniques from Laplacian solvers and electrical flows. The algorithm by Schild is very sophisicated and complicated, and so Theorem 23.21 is a dramatic simplification based on better analysis of mixing time.

Concentration Inequality for Strongly Log-Concave Distribution

One main application of log-Sobolev inequalities is to prove concentration inequalities [BLM13, VH14]. The following result is a consequence of Theorem 23.19 for strongly log-concave distributions.

Theorem 23.22 (Concentration of Strongly Log-Concave Distributions [CGM21]). *Let μ be a d-homogeneous strongly log-concave distribution with support $\Omega \subseteq \{0,1\}^n$. For any observable function $f : \Omega \to \mathbb{R}$ and $a \geq 0$,*

$$\Pr_{x\sim\mu}\left[|f(x) - \mathbb{E}_\pi f| \geq a\right] \leq 2\exp\Big(-\frac{a^2}{2d\cdot\nu(f)}\Big),$$

where $\nu(f)$ is the maximum of one-step variances

$$\nu(f) := \max_{x\in\Omega}\left\{\sum_{y\in\Omega} P_d^{\triangledown}(x,y)\cdot\big(f(x)-f(y)\big)^2\right\}.$$

The proof of Theorem 23.22 follows from the Herbst argument (see [BLM13]). The reader is referred to [CGM21] for the proof. For a c-Lipschitz function under the graph distance in the bases exchange graph, $\nu(f) \leq c^2$ and thus Theorem 23.22 generalizes the concentration result for strongly Rayleigh distributions in Theorem 16.18.

23.3 References

[AJK+21] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence in high-dimensional expanders: Modified log-sobolev inequalities for fractionally log-concave polynomials and the ising model. *CoRR*, abs/2106.04105, 2021. 217, 225, 226, 227

[ALO+21] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, Cynthia Vinzant, and Thuy-Duong Vuong. Log-concave polynomials IV: approximate exchange, tight mixing times, and near-optimal sampling of forests. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 408–420. ACM, 2021. 222

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013. 223

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *2011 ieee 52nd annual symposium on foundations of computer science*, pages 472–481. IEEE, 2011. 218

[BT06] Sergey G Bobkov and Prasad Tetali. Modified logarithmic sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006. 217, 219, 220

[CGM21] Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-sobolev inequalities for strongly log-concave distributions. *Annals of Probability*, 49(1):506–525, 2021. 217, 222, 223

[DSC96] Persi Diaconis and Laurent Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996. 219

[MT06] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006. 217, 221

[VH14] Ramon Van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014. 223

Chapter 24

Entropic Independence

We end with a very recent paper by Anari, Jain, Koehler, Pham, Vuong [AJK+21a], who introduced the notion of entropic independence and used it to lower bound the modified log-Sobolev constants of fractionally log-concave distributions. The presentation is taken directly from [AJK+21a].

24.1 Larger-Step Down-Up Walks

We will consider the following natural generalization of Definition 20.5 for down-up walks.

Definition 24.1 (Larger-Step Down-Up Walks). *For a probability distribution $\mu : \binom{[n]}{k} \to \mathbb{R}$ and an integer $l \leq k$, the $k \leftrightarrow l$ down-up random walk $P^{\nabla}_{k \leftrightarrow l}$ is the sequence of random sets $S_0, S_1, \ldots$ generated by the following algorithm:*

for *$t = 0, 1, \ldots$* ***do***
  Select T_t uniformly at random from subsets of size l of S_t.
  Select S_{t+1} with probability $\propto \mu(S_{t+1})$ from supersets of size k of T_t.
end for

We also take this opportunity to redefine the down and up operators, but use different notations, that are more natural for the analysis of random walks. Compare with Definition 20.2 and see Remark 20.3. We will think of these as matrices that act on probability distributions on the left.

Definition 24.2 (Down Operator). *For a ground set $[n]$ and $n \geq k \geq l$, define the row-stochastic down operator $D_{k \to l} \in \mathbb{R}^{\binom{[n]}{k} \times \binom{[n]}{l}}$ as*

$$D_{k \to l}(S, T) = \begin{cases} \frac{1}{\binom{k}{l}} & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 24.3 (Up Operator). *For a ground set $[n]$ and $n \geq k \geq l$, and probability distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, define the up operator $U_{l \to k} \in \mathbb{R}^{\binom{[n]}{l} \times \binom{[n]}{k}}$ as*

$$U_{l \to k}(T, S) = \begin{cases} \frac{\mu(S)}{\sum_{S' \supseteq T} \mu(S')} & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

The following properties should be familiar.

Exercise 24.4 (Stationary Distributions of $P^{\nabla}_{k\leftrightarrow l}$ and $P^{\triangle}_{k\leftrightarrow l}$). *Define $\mu_k := \mu$ and more generally define $\mu_l := \mu_k D_{k\to l}$. The operators $P^{\nabla}_{k\leftrightarrow l} := D_{k\to l}U_{l\to k}$ and $P^{\triangle}_{l\leftrightarrow k} := U_{l\to k}D_{k\to l}$ both define Markov chains that are time-reversible and have nonnegative eigenvalues. Moreover μ_k and μ_l are respectively their stationary distributions.*

24.2 Entropic Independence

The notion of entropic independence is about the entropy contraction of the down operator.

Definition 24.5 (Entropic Independence). *A probability distribution μ on $\binom{[n]}{k}$ is said to be $(1/\alpha)$-entropically independent, for $\alpha \in (0,1]$, if for all probability distributions ν on $\binom{[n]}{k}$,*

$$\mathcal{D}_{\mathrm{KL}}\big(\nu D_{k\to 1} \parallel \mu D_{k\to 1}\big) \le \frac{1}{\alpha k}\cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

We remark that spectral independence can be used to prove the variance contraction of the down operator (see [Ove22]), which can be used to prove a lower bound on the spectral gap as in Chapter 23. So Definition 24.5 can be seen as a natural analog of spectral independence for proving a lower bound on the modified log-Sobolev constant as in Chapter 23.

We study two main results from [AJK+21a]. The first one characterizes entropic independence and fractionally log-concavity. In particular, it proves that an α-fractionally log-concave distribution is $(1/\alpha)$-entropically independent for all conditional distributions.

Theorem 24.6 (Entropic Independence and Fractionally Log-Concavity [AJK+21a]). *Let μ be a probability distribution on $\binom{[n]}{k}$ and let $\mu_1 = \mu D_{k\to 1} \in \mathbb{R}^n$. Then, for any $\alpha \in (0,1]$,*

$$\mu \text{ is } \frac{1}{\alpha}\text{-entropically independent} \iff \forall (z_1,\dots,z_n)\in\mathbb{R}^n_{\ge 0},\quad g_\mu(z_1^\alpha,\dots,z_n^\alpha)^{\frac{1}{k\alpha}} \le \sum_{i=1}^n \mu_1(i)\cdot z_i.$$

Consequently, if μ is α-fractionally log-concave, then μ is $(1/\alpha)$-entropically independent. Moreover,

$$\mu \text{ is } \alpha\text{-fractionally log-concave} \iff \lambda{*}\mu \text{ is } (1/\alpha)\text{-entropically independent } \forall \lambda=(\lambda_1,\dots,\lambda_n)\in\mathbb{R}^n_{>0},$$

where $\lambda{*}\mu$ is the distribution scaled by external field λ such that $g_{\lambda*\mu}(z_1,\dots,z_n) \propto g_\mu(\lambda_1 z_1,\dots,\lambda_n z_n)$.

The second one proves that if μ is $(1/\alpha)$-entropically independent for all conditional distributions, then the $k \leftrightarrow k-\frac{1}{\alpha}$ down-up walk is fast mixing.

Theorem 24.7 (Local to Global Entropy Contraction [AJK+21a]). *Suppose $\mu : \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ is α-fractionally log-concave, or more generally $(1/\alpha)$-entropically independent for all conditional distributions. Let $l \le k - \lceil 1/\alpha \rceil$. Then, for all probability distributions ν on $\binom{[n]}{k}$,*

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l}) \le (1-\kappa)\cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

Consequently, the $k \leftrightarrow l$ down-up walk with respect to μ has modified log-Sobolev constant $\Omega(\kappa)$ where Ω hides an absolute constant and when $1/\alpha$ is an integer then

$$\kappa = \binom{k-l}{1/\alpha} \Big/ \binom{k}{1/\alpha}.$$

Combining Theorem 24.6 and Theorem 24.7, the $k \leftrightarrow k-\frac{1}{\alpha}$ down-up is fast mixing for α-fractionally log-concave distributions. Note that this is a generalization of the result of Cryan, Guo, and Mousa in Theorem 23.19 for strongly log-concave distributions, which are 1-fractionally log-concave. This also gives the optimal mixing time analysis for the monomer-dimer systems and the non-symmetric determinantal point process that we discussed in Chapter 22 (see [AJK+21a]).

24.3 Fractional Log-Concavity Implies Local Entropic Contraction

In this section, we prove one direction of Theorem 24.6, that fractionally log-concavity implies entropic independence. We refer the reader to [AJK+21a] for the other direction, which is not needed for the conclusion of fast mixing.

The following statement will be used to replace log-concavity by $\frac{1}{d}$-th root concavity.

Problem 24.8. *Let $\mathcal{C} \subseteq \mathbb{R}^n_{\geq 0}$ denote a convex cone. For a d-homogeneous function $f : C \to \mathbb{R}_{\geq 0}$, f is log-concave if and only if $f^{1/d}$ is concave.*

The first step is to show that fractional log-concavity implies that the transformed generating polynomial is upper bounded by its linear tangent.

Lemma 24.9 (Linear Tangent Upper Bound)**.** *If μ is a α-fractionally log-concave distribution on $\binom{[n]}{k}$, then*

$$g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \leq \Big(\sum_{i=1}^n \mu_1(i) \cdot z_i\Big)^{\alpha k}.$$

Proof. Let $f(z_1, \ldots, z_n) := g_\mu(z_1^\alpha, \ldots, z_n^\alpha)^{\frac{1}{\alpha k}}$ be the transformed generating polynomial. As the polynomial $g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is αk-homogeneous and log-concave, it follows from Problem 24.8 that f is concave. Therefore, by concavity, for all $z_1, \ldots, z_n > 0$,

$$f(z_1, \ldots, z_n) \;\leq\; f(\vec{1}) + \langle \nabla f(\vec{1}), \vec{z} - \vec{1} \rangle \;=\; f(\vec{1}) + \sum_{i=1}^n \partial_i f(\vec{1}) \cdot (z_i - 1) \;=\; \sum_{i=1}^n \partial_i f(\vec{1}) \cdot z_i,$$

where the last equality is because f is 1-homogeneous and so $\sum_{i=1}^n \partial_i f(\vec{1}) = f(\vec{1})$. By the chain rule,

$$\partial_i f(\vec{1}) \;=\; \big(\alpha \cdot \partial_i g_\mu(\vec{1})\big)\Big(\frac{1}{\alpha k} \cdot g_\mu(\vec{1})^{\frac{1}{\alpha k}-1}\Big) \;=\; \frac{1}{k} \Pr_{S \sim \mu}[i \in S] \;=\; \mu_1(i),$$

where the last equality is by the definition that $\mu_1 = \mu D_{k \to 1}$. Therefore,

$$f(z_1, \ldots, z_n) \leq \sum_{i=1}^n \mu_1(i) \cdot z_i \quad \Longrightarrow \quad g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \leq \Big(\sum_{i=1}^n \mu_1(i) \cdot z_i\Big)^{\alpha k}.$$

□

The second step is to show that the linear tangent upper bound implies entropic independence. A key idea in the proof is to fix the marginal probability and to use the following Gurvits' capacity-type bound proved by Singh and Vishnoi, which is obtained by convex duality.

Lemma 24.10 (Capacity Bound of Relative Entropy). *Consider a homogeneous distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ and let $g_\mu(z_1, \ldots, z_n)$ be its multivariate generating polynomial. Then, for any $q \in \mathbb{R}^n_{\geq 0}$ with $\sum_{i=1}^n q_i = 1$,*

$$\inf \left\{ \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \mid \nu D_{k\to 1} = q \right\} = -\log \left(\inf_{z_1,\ldots,z_n>0} \frac{g_\mu(z_1, \ldots, z_n)}{z_1^{kq_1} \cdots z_n^{kq_n}} \right).$$

Lemma 24.11 (Linear Tangent Upper Bound Implies Entropic Independence). *If μ is a homogeneous distribution whose generating polynomial g_μ satisfies*

$$g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \leq \left(\sum_{i=1}^n \mu_1(i) \cdot z_i \right)^{\alpha k}.$$

for all $z_1, \ldots, z_n \in \mathbb{R}^n_{\geq 0}$, then μ is $(1/\alpha)$-entropically independent.

Proof. Let ν be an arbitrary probability distribution on $\binom{[n]}{k}$ and let $\nu_1 := \nu D_{k\to 1}$, so that $\nu_1 \in \mathbb{R}^n_{\geq 0}$ and $\sum_{i=1}^n \nu_1(i) = 1$. By Lemma 24.10,

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \geq \inf \left\{ \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \mid \nu D_{k\to 1} = \nu_1 \right\} = -\log \left(\inf_{z_1,\ldots,z_n>0} \frac{g_\mu(z_1, \ldots, z_n)}{z_1^{k\cdot\nu_1(1)} \cdots z_n^{k\cdot\nu_1(n)}} \right).$$

By the linear tangent upper bound,

$$\inf_{z_1,\ldots,z_n>0} \frac{g_\mu(z_1, \ldots, z_n)}{z_1^{k\cdot\nu_1(1)} \cdots z_n^{k\cdot\nu_1(n)}} \leq \inf_{z_1,\ldots,z_n>0} \frac{\left(\sum_{i=1}^n \mu_1(i) \cdot z_i^{\frac{1}{\alpha}} \right)^{\alpha k}}{z_1^{k\cdot\nu_1(1)} \cdots z_n^{k\cdot\nu_1(n)}} \leq \prod_{i=1}^n \left(\frac{\mu_1(i)}{\nu_1(i)} \right)^{\alpha k \cdot \nu_1(i)}$$

where the last inequality is by plugging in $z_i = (\nu_1(i)/\mu_1(i))^\alpha$. Taking log and negating gives

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \geq -\log \prod_{i=1}^n \left(\frac{\mu_1(i)}{\nu_1(i)} \right)^{\alpha k \cdot \nu_1(i)} = \alpha k \sum_{i=1}^n \nu_1(i) \log \left(\frac{\nu_1(i)}{\mu_1(i)} \right) = \alpha k \cdot \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1} \parallel \mu D_{k\to 1}),$$

which establishes $(1/\alpha)$-entropic independence by Definition 24.5. □

Finally, we note that α-fractional log-concavity is preserved by scaling, which implies that the generating polynomial of any conditional distribution

$$g_{\mu_S} \propto \lim_{\lambda\to\infty} \frac{g_\mu\left(\lambda z_1, \ldots, \lambda z_{|S|}, \ldots, z_n\right)}{\lambda^{|S|}}.$$

is also α-fractionally log-concave, and thus μ_S is also $(1/\alpha)$-entropically independent by Lemma 24.9 and Lemma 24.11.

Therefore, we have proved the direction of Theorem 24.6 that we need, that an α-fractionally log-concave distribution is $(1/\alpha)$-entropic independent for all conditional distributions.

24.4 Local Entropy Contraction to Global Entropy Contraction

In this section, we prove Theorem 24.7, which also proves Theorem 23.19 for strongly log-concave distributions.

Proof of Theorem 24.7. The plan is to write both $\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$ and $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l})$ as a telescoping sum of terms of the form $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to i} \parallel \mu D_{k\to i}) - \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to(i-1)} \parallel \mu D_{k\to(i-1)})$.

Consider the following random process. Sample a set $S \sim \mu$ and uniformly at random permute its elements to obtain $X_1, \ldots, X_k$. Notice that any prefix $X_1, \ldots, X_i$ is distributed according to $\mu D_{k\to i}$. Consider the random variable

$$\tau_i := \frac{\nu D_{k\to i}(\{X_1, \ldots, X_i\})}{\mu D_{k\to i}(\{X_1, \ldots, X_i\})} \log \frac{\nu D_{k\to i}(\{X_1, \ldots, X_i\})}{\mu D_{k\to i}(\{X_1, \ldots, X_i\})}$$

Then

$$\mathbb{E}_{S\sim\mu}[\tau_i] = \mathbb{E}_{\{X_1,\ldots,X_i\}\sim\mu D_{k\to i}}[\tau_i] = \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to i} \parallel \mu D_{k\to i}).$$

Therefore, we can write both $\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$ and $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l})$ as telescoping sums

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) = \mathbb{E}[\tau_k] = \sum_{i=0}^{k-1} \big(\mathbb{E}[\tau_{i+1}] - \mathbb{E}[\tau_i]\big) \text{ and } \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l}) = \mathbb{E}[\tau_l] = \sum_{i=0}^{l-1} \big(\mathbb{E}[\tau_{i+1}] - \mathbb{E}[\tau_i]\big).$$

To prove entropy contraction, it is equivalent to proving that the last $k-l$ terms in the telescoping sum are sufficiently large compared to the rest.

Let $\Delta_i := \mathbb{E}[\tau_{i+1}] - \mathbb{E}[\tau_i]$ and $\beta_i := \frac{1}{\alpha(k-i)-1}$. As μ is α-fractionally log-concave, by Theorem 24.6, μ is $(1/\alpha)$-entropically independent, and thus it follows from Definition 24.5 that

$$\Delta_0 = \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1} \parallel \mu D_{k\to 1}) \le \frac{1}{\alpha k} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \quad \Longrightarrow \quad \Delta_0 \le \beta_0(\Delta_1 + \cdots + \Delta_{k-1}).$$

As conditioning preserves α-fractionally log-concavity, we can apply the same argument to each conditional distribution $\mu(\cdot \mid X_1, \ldots, X_i)$ and then take the expectation over $X_1, \ldots, X_i$ to get

$$\Delta_i \le \beta_i(\Delta_{i+1} + \ldots + \Delta_{k-1})$$

for each $1 \le i \le l-1$. Combining these inequalities, it follows inductively that for all $0 \le i \le l \le k - \frac{1}{\alpha}$,

$$\Delta_i \le \beta_i \cdot (\Delta_l + \ldots + \Delta_{k-1}) \cdot \prod_{j=i+1}^{l-1} (\beta_j + 1).$$

Hence,

$$\frac{\Delta_0 + \cdots + \Delta_{k-1}}{\Delta_l + \cdots + \Delta_{k-1}} = 1 + \frac{\Delta_0 + \cdots + \Delta_{l-1}}{\Delta_l + \cdots + \Delta_{k-1}} \le 1 + \sum_{i=0}^{l-1} \beta_i \cdot \prod_{j=i+1}^{l-1} (\beta_j + 1) = \prod_{i=0}^{l-1} (1 + \beta_i) = \prod_{j=k-l+1}^{k} \frac{j}{j - \frac{1}{\alpha}}.$$

Let $\Gamma(\cdot)$ be the Gamma function. If $1/\alpha$ is an integer, then the RHS is

$$\frac{\Gamma(k+1)/\Gamma(k+1-l)}{\Gamma(k+1-1/\alpha/\Gamma(k+1-l-1/\alpha)} = \frac{k!/(k-l)!}{(k-1/\alpha)!(k-l-1/\alpha)!} = \binom{k}{1/\alpha} \Big/ \binom{k-l}{1/\alpha}.$$

This implies that

$$\frac{\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)}{\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) - \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l})} = \frac{\Delta_0 + \cdots + \Delta_{k-1}}{\Delta_l + \cdots + \Delta_{k-1}} \le \binom{k}{1/\alpha} \Big/ \binom{k-l}{1/\alpha},$$

and rearranging gives the entropy contraction statement $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l}) \le (1-\kappa) \cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$.

Finally, by the data processing inequality in information theory, apply the up operator would not increase the relative entropy, and so

$$\mathcal{D}_{\mathrm{KL}}(\nu P^{\nabla}_{k\leftrightarrow l} \parallel \mu P^{\nabla}_{k\leftrightarrow l}) = \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l}U_{l\to k} \parallel \mu D_{k\to l}U_{l\to k}) \leq \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l}) \leq (1-\kappa)\cdot\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

This implies that the relative entropy is exponentially decreasing and thus the we can upper bound the mixing time as in Theorem 23.14 with $\rho(P) \approx \kappa$. This proves the consequence of the modified log-Sobolev constant in bounding mixing time. It is left as in Problem 24.12 to prove that entropy contraction indeed implies a lower bound on the modified log-Sobolev constant. □

Problem 24.12 (Entropy Contraction Implies Modified Log-Sobolev Constant)**.** *Let μ be a probability distribution on $[n]$. Let P denote the transition matrix of an irreducible, reversible Markov chain on $[n]$ with stationary distribution μ. Suppose there exists some $\alpha \in (0,1]$ such that for all probability measures ν on $[n]$ which are absolutely continuous with respect to μ, we have*

$$\mathcal{D}_{\mathrm{KL}}(\nu P \parallel \mu P) \leq (1-\alpha)\cdot\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

Then the modified log-Sobolev constant of P is

$$\rho(P) \geq 2\alpha.$$

24.5 Summary

Since the resolution of the matroid expansion conjecture using the connection to high-dimensional expanders, there are many recent developments in analyzing mixing time of Markov chains. It is quite amazing to see that the techniques can be extended to bounding the (notorious) modified log-Sobolev constants, with applications in proving optimal mixing times, in bounding correlations, and in proving concentration inequalities. In these most recent developements, the concepts from high-dimensional expanders have been bypassed and one could understand the results directly using probabilistic and analytical concepts (see the newer papers [AJK+21b, CE22]). This area is progressing and evolving very quickly, and the notes in the next offering may be entirely different.

24.6 References

[AJK+21a] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence in high-dimensional expanders: Modified log-sobolev inequalities for fractionally log-concave polynomials and the ising model. *CoRR*, abs/2106.04105, 2021. 217, 225, 226, 227

[AJK+21b] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence ii: Optimal sampling and concentration via restricted modified log-sobolev inequalities. *arXiv preprint arXiv:2111.03247*, 2021. 230

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. *arXiv preprint arXiv:2203.04163*, 2022. 230

[Ove22] Shayan Oveis Gharan. Course notes of "Modern spectral graph theory", lecture 12. 2022. https://courses.cs.washington.edu/courses/cse599s/22wi/. 153, 159, 226