New Approaches to the Asymmetric Traveling Salesman and Related Problems

Nima AhmadiPourAnari

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2016-2
http://www.eecs.berkeley.edu/Pubs/TechRpts/2016/EECS-2016-2.html

January 4, 2016

New Approaches to the Asymmetric Traveling Salesman and Related Problems

by

Nima Ahmadipouranari

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Satish Rao, Chair
Professor Prasad Raghavendra
Professor Nikhil Srivastava
Professor Luca Trevisan

Fall 2015

New Approaches to the Asymmetric Traveling Salesman and Related Problems

Abstract

New Approaches to the Asymmetric Traveling Salesman and Related Problems

by

Nima Ahmadipouranari

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Satish Rao, Chair

The Asymmetric Traveling Salesman Problem and its variants are optimization problems that are widely studied from the viewpoint of approximation algorithms as well as hardness of approximation. The natural LP relaxation for ATSP has been conjectured to have an $O(1)$ integrality gap. Recently the best known approximation factor for this problem was improved from the decades-old $O(\log(n))$ to $O(\log(n)/\log\log(n))$ using the connection between ATSP and Goddyn's Thin Tree conjecture.

In this work we show that the integrality gap of the famous Held-Karp LP relaxation for ATSP is bounded by $\log\log(n)^{O(1)}$ which entails a polynomial time $\log\log(n)^{O(1)}$-estimation algorithm; that is we provide a polynomial time algorithm that finds the cost of the best possible solution within a $\log\log(n)^{O(1)}$ factor, but does not provide a solution with that cost. This is one of the very few instances of natural problems studied in approximation algorithms where the state of the art approximation and estimation algorithms do not match.

We prove this by making progress on Goddyn's Thin Tree conjecture; we show that every k-edge-connected graph contains a $\log\log(n)^{O(1)}/k$-thin tree.

To tackle the Thin Tree conjecture, we build upon the recent resolution of the Kadison-Singer problem by Marcus, Spielman, and Srivastava. We answer the following question by providing sufficient conditions: Given a set of rank 1 quadratic forms, can we select a subset of them from a given collection of subsets, whose total sum is bounded by a fraction of the sum of all rank 1 quadratic forms?

Finally we address the problem of designing polynomial time approximation algorithms, algorithms that also output a solution, matching the guarantee of the estimation algorithm. We prove that this entirely relies on finding a polynomial time algorithm for our extension of the Kadison-Singer problem. Namely we prove that ATSP can be $\log(n)^{\epsilon}$-approximated in polynomial time for any $\epsilon > 0$ and that it can be $\log\log(n)^{O(1)}$-approximated in quasi-polynomial time, assuming access to an oracle which solves our extension of Kadison-Singer.

To Maman, Baba, Parima,
and Dorsa.

Contents

List of Figures

List of Tables

List of Algorithms

Acknowledgments

Most of this dissertation is based on joint work with Shayan Oveis Gharan. Over the course of two years that Shayan visited Berkeley, I have been fortunate to receive significant advice and assistance from him, not only on the topic of this dissertation, but rather on all aspects of academic life. Shayan's patience is extraordinary to say the least, and that has been perhaps the biggest lesson I have learned from him. Reaching dead ends in research is often discouraging and many, admittedly including me, choose to give up and move on to other problems. If it was not for Shayan's perseverance and encouragement, this dissertation would not exist today. My advisor, Satish Rao, described the work behind this dissertation as a marathon; I cannot choose better words to describe it, and I certainly benefited from a drafting effect because of running with Shayan.

I was fortunate enough to be advised by Satish Rao, who supported me throughout the years, and has always been kind enough to be available for discussion about research and guidance. I would also like to thank Umesh Vazirani. He has been a source of encouragement and invaluable advice throughout my Ph.D. I also thank Vahab Mirrokni, Mohit Singh, and Yuval Peres who have mentored me during my internships.

I thank all of the graduate students and postdocs in the Theory group for their friendship and the creation of an encouraging atmosphere: Sam Wong, Anand Bhaskar, Anindya De, Di Wang, Aviad Rubinstein, Tselil Schramm, Ben Weitz, Jarett Schwartz, Piyush Srivastava, Seung Woo Shin, Alex Psomas, Antonio Blanca, Jonah Brown-Cohen, James Cook, Omid Etesami, Lorenzo Orecchia, Chris Wilkens, Siu Man Chan, Siu On Chan, Tom Watson, Thomas Vidick, Ilias Diakonikolas, Rafael Frongillo, George Pierrakos, Paul Christiano, Fotis Iliopoulos, Urmila Mahadev, Manuel Sabin, Jingcheng Liu, Peihan Miao, Aaron Schild, Chris Wilkens, Gregory Valiant, Virginia Vassilevska Williams, Isabelle Stanton, and Rishi Gupta.

I spent the last years of graduate school having access to the Simons institute for the theory of computing and visitors to the institute. For that I thank all of the faculty who put a great effort into making sure that the institute gets created at Berkeley and ensured such great quality for its programs.

I also thank all of the people I have not mentioned already with whom I have had stimulating discussions that have affected my research even to the slightest degree: Lap Chi Lau, Afshin Nikzad, Morteza Zadimoghaddam, Amin Saberi, Nikhil Srivastava, Prasad Raghavendra, Dan Spielman, Nikhil Bansal, Yin Tat Lee, Nika Haghtalab, Elchanan Mossel, Luca Trevisan, and Russell Lyons.

My experience at Berkeley was made pleasant because of the support I received from my friends. I would like to acknowledge those I have not mentioned already: Ramtin Pedarsani, Zhaleh Amini, Payam Delgosha, Negar Zahedimehr, Shayan Ehsani, Hadi Pouransari, Mohammad Bavarian, Sara Magliacane, Sadra Yazdanbod.

Finally I want to thank Dorsa Sadigh without whom my days would not be as colorful.

Chapter 1

Introduction

The Traveling Salesman Problem (TSP) and its variants are optimization problems widely studied from the viewpoint of approximation algorithms as well as hardness of approximation. They find applications in scheduling, manufacturing of microchips, genome sequencing, etc; see [App+11] for details. Besides having many applications, they are of particular interest to theoretical computer scientists because they have resisted many well-known and widely applied techniques developed in the field of approximation algorithms. Major developments in the field have been initiated to tackle these problems.

In the traveling salesman problem, we are given a set V of n vertices, and nonnegative costs for traveling between pairs of vertices. The goal is to find the shortest tour that visits each vertex *at least* once. Two main variants of TSP, the Symmetric Traveling Salesman Problem (STSP) and the Asymmetric Traveling Salesman Problem (ATSP) have both enjoyed wide interest from the approximation algorithms community. In STSP, the cost for traveling from u to v is the same as the cost for traveling from v to u, and in ATSP this does not necessarily hold.

One of the first approximation algorithms ever developed was for the symmetric traveling salesman problem [Chr76]; STSP was also one of the original NP-complete problems introduced by Karp [Kar72]. For STSP the best approximation algorithm and the best known approximation hardness are both constants. In the case of ATSP, the gap is more dramatic.

The natural linear programming relaxation for ATSP, proposed by Held and Karp [HK70], has been conjectured to have a constant integrality gap. This would mean that the optimal tour's cost is no more than a constant multiple of the linear program's solution. This would provide an algorithm for estimating the cost of the optimum solution in ATSP, since the linear program can be solved in polynomial time. The method used to prove upper bounds on the integrality gap of linear programs might sometimes also yield an algorithm which finds a solution. Most known rounding techniques fall into this category. In this dissertation we use a rounding method for which no polynomial-time algorithm is known. As such, we provide an algorithm for estimating the optimum solution's cost, but not one for finding a solution. As such we create one of the few instances of natural optimization problems that have this property; see [FJ15] for more instances.

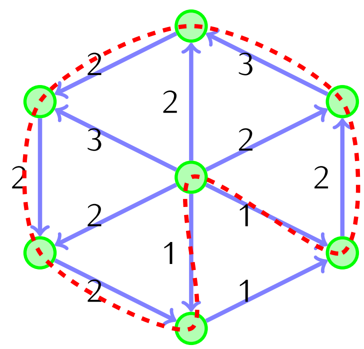

Figure 1.1: An example ATSP instance and its optimum solution

Recently the best known approximation algorithm for ATSP was improved from the decades-old $O(\log(n))$ to $O(\log(n)/\log\log(n))$ [Asa+10]. The same work introduced the connection between ATSP and Goddyn's thin tree conjecture. In this dissertation, we explore this relationship further; we improve the upper bound on the integrality gap of the linear programming relaxation to $(\log\log n)^{O(1)}$ by making progress on Goddyn's thin tree conjecture.

As our main tool, we build upon the recent resolution of the Kadison-Singer problem and the method of interlacing polynomials [MSS13b]. We extend Kadison-Singer to work with the so-called "strongly Rayleigh" measures. Strongly Rayleigh measures are a class of negatively correlated point processes that include random spanning tree distributions on graphs. Their properties have been already used in other works on TSP; see [OSS11] for an example.

1.1 Asymmetric Traveling Salesman Problem

In the asymmetric traveling salesman problem one is given a directed graph $G = (V, E)$, together with a cost function $c : E \to \mathbb{R}_{\geq 0}$ and the goal is to find the shortest tour that visits every vertex at least once; see fig. 1.1 for an example.

There is a natural LP relaxation for ATSP proposed by Held and Karp [HK70]:

$$\begin{aligned}
\min \sum_{u,v \in V} & c(u,v) x_{u,v} & \\
\text{s.t.} \sum_{u \in S, v \notin S} & x_{u,v} \geq 1 & \forall S \subseteq V, \\
\sum_{v \in V} x_{u,v} &= \sum_{v \in V} x_{v,u} = 1 & \forall u \in V, \\
x_{u,v} &\geq 0 & \forall u, v \in V.
\end{aligned} \tag{1.1}$$

In (1.1) the variable $x_{u,v}$ indicates whether the edge (u, v) is part of the tour. The constraints ensure that the solution is connected and Eulerian, i.e., that the tour enters and exits each vertex the same number of times.

Work	Type	Factor
[CGK06]	Integrality gap lower bound	$2 - \epsilon$
[FGM82]	Approximation algorithm	$\log_2 n$
[Blä02]		$0.999 \log_2 n$
[Kap+05]		$\frac{4}{3} \log_3 n$
[FS07]		$\frac{2}{3} \log_2 n$
[Asa+10]		$O(\frac{\log n}{\log \log n})$
[OS11]	Approximation algorithm for bounded genus graphs	$O(1)$
[ES14]		$O(1)$
[Sve15]	Approximation algorithm for node-weighted graphs	$O(1)$

Table 1.1: Previous Works on ATSP

It is conjectured that the integrality gap of the Held-Karp LP relaxation is a constant, i.e., the optimum value of the above LP relaxation is within a constant factor of the length of the optimum ATSP tour. Until very recently, we had a very limited understanding of the solutions of the LP relaxation. To this date, the best known lower bound on the integrality gap is 2 [CGK06].

Despite many efforts, no constant factor approximation algorithm is known for ATSP. The first nontrivial attempt was the work of [FGM82] who designed a $\log_2 n$-approximation algorithm for ATSP. Subsequently in a series of works by [Blä02; Kap+05; FS07], $\log_2 n$ was improved by constant factors, the end result being a $\frac{2}{3} \log_2 n$-approximation algorithm. Finally in their seminal work [Asa+10] broke the $\log n$ barrier and provided a $O(\log n / \log \log n)$-approximation algorithm for ATSP. All aforementioned results are based on the Held-Karp LP, and as such these approximation algorithms also prove upper bounds on the integrality gap of the LP.

For special cases of ATSP, better approximation algorithms have been developed. For planar and bounded genus graph [OS11; ES14] designed $O(1)$-approximation algorithms. More recently [Sve15] designed an $O(1)$-approximation algorithm for the class of node-weighted graphs, i.e., graphs where the weights of edges coming out of each vertex are the same. A summary of these results can be seen in table 1.1.

There has been generally two approaches for attacking ATSP:

1. Start with a connected subgraph, i.e., a spanning tree, and make it Eulerian by adding edges. This approach was successfully used by [Asa+10] to get the $O(\log n / \log \log n)$-approximation algorithm. This approach was also successfully used for the symmetric variant of TSP [Chr76].

2. Start with an Eulerian subgraph, i.e., a union of cycles covering the graph, and make it connected by adding edges. This has been the approach taken by the earlier works [Blä02; Kap+05; FS07]. This approach has made a recent comeback in [Sve15].

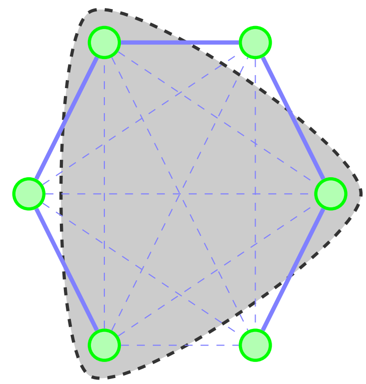

Figure 1.2: In the complete graph K_n, a Hamiltonian path is $O(1/n)$-thin. The Hamiltonian path takes $\simeq 4/n$ fraction of the edges from the depicted cut, which separates consecutive vertices of the Hamiltonian path.

In this dissertation we will be using the first approach. As shown by [Asa+10], Goddyn's thin tree conjecture has strong implications for this approach. We will formally prove that the integrality gap of the Held-Karp LP relaxation is bounded by $(\log \log n)^{O(1)}$.

1.2 Goddyn's Thin Tree Conjecture

First proposed by Goddyn [God04], the thin tree conjecture was devised as a tool for proving results about nowhere-zero flows. The conjecture states that every k-edge-connected undirected graph contains a $f(k)$-thin spanning tree, i.e., a spanning tree that contains at most $f(k)$ fraction of the edges in every cut. The only stipulation about the function f is that $\lim_{k\to\infty} f(k) = 0$, but the strongest form of the conjecture states that $f(k)$ can be taken to be $O(1/k)$. For an example of thin trees in complete graphs see fig. 1.2.

The assumption about k-edge-connectivity is necessary. An $O(1/k)$-thin tree contains at least one edge in every cut. Therefore the graph must contain at least $\Omega(k)$ edges in every cut. Goddyn's conjecture in its strongest form implies that this constraint is also sufficient, up to constant factors.

The main difficulty about Goddyn's thin tree conjecture is that thinness should not depend on n, the number of vertices in the graph. If one is allowed dependency on n, by independently sampling edges one can obtain a $O(\log n/k)$-thin tree in k-edge-connected graphs. By using dependent sampling from spanning tree distributions, [Asa+10] improved this to $O(\log n/k \log \log n)$.

Thin trees can be used to obtain solutions to ATSP. Very roughly, a tree that is α-thin with respect to the solution of the Held-Karp LP relaxation can be completed into an ATSP tour without incurring more than α times the cost of the LP solution. This idea, first introduced in [Asa+10], can be strengthened to get the following formal implication: If every k-edge-connected graph contains a $f(n)/k$-thin tree, then the integrality gap of the Held-Karp LP relaxation is bounded by $O(f(n))$.

In this dissertation, we will formally prove that every k-edge-connected graph contains a $(\log\log n)^{O(1)}/k$-thin tree.

1.3 The Kadison-Singer Problem and Preconditioning

Sampling techniques have been successful to some degree for obtaining thin trees [Asa+10]. But they do not seem to provide better than $O(\log n/\log\log n)$-approximation for ATSP or better than $O(\log n/k \log\log n)$-thin trees in k-edge-connected graphs. This barrier seems to be because these techniques need to work with high probability. One way to get around this barrier is to use techniques that are designed to show the occurrence of low-probability events. One widely applied technique that falls into this category is the Lovász Local Lemma [EL75]; for some examples see [Sze13].

Another notable technique which was introduced very recently to show the existence of Ramanjuan graphs [MSS13a; MSS15] and solve the Kadison-Singer problem [MSS13b] is the method of interlacing polynomials. Very roughly, this method consists of showing that the roots of a certain group of polynomials interlace the roots of their average, i.e., the roots of the average polynomial fall in between the roots of the individual polynomials. Thus any bound on the roots of the average polynomial translate to bounds on the roots of *at least one* of the individual polynomials.

Interlacing polynomials have been successfully used to solve the Kadison-Singer problem. The technique was used by [MSS13b] to prove the paving conjecture due to [And79a; And79b; And81; Cas+07], and Weaver's conjecture due to [Wea04], both of which imply the original Kadison-Singer problem proposed in [KS59]. Weaver's formulation is in particular relevant to the thin tree problem [HO14]. By standard techniques, one can reduce Weaver's formulation to the following: One is given a set of rank 1 positive semidefinite quadratic forms (i.e., matrices) $A_1, \ldots, A_n$, and a random subset S of $\{1, \ldots, n\}$. The goal is to provide sufficient conditions on $A_1, \ldots, A_n$ and the law of S to have the following event happen with positive probability:

$$\sum_{i\in S} A_i \preceq \epsilon \sum_{i=1}^{n} A_i. \tag{1.2}$$

When the quadratic forms A_i are Laplacians of edges in a graph and the set S is a random spanning tree, (1.2) implies that S is a thin tree with positive probability. Equation (1.2) actually implies something stronger than thinness; a tree satisfying (1.2) is called a *spectrally thin tree*.

We will give sufficient conditions for the occurrence of the event described by (1.2), when the random set S has a law which is *strongly Rayleigh*. Strongly Rayleigh distributions will be discussed in depth, but the random spanning tree distribution is an important example of them which will be used to construct spectrally thin trees. In [MSS13b] the case of S having an independent distribution, a special case of strongly Rayleigh distributions, was considered and resolved.

Not every k-edge-connected graph has an $O(1/k)$-spectrally thin tree. Therefore we will have to transform the graph to meet the sufficient conditions of Kadison-Singer. This transformation on a very high level leaves the cut structure of the graph intact while changing its spectral properties. In other words we will be *preconditioning* the input graph for the application of Kadison-Singer.

1.4 Organization

Chapter 2 will provide the notation used throughout the dissertation as well as well-known facts, lemmas, and theorems that will be used. For readers who are unfamiliar with real stable polynomials, reading the corresponding section in this chapter before reading the rest of the dissertation is recommended.

Chapter 3 will provide a high-level overview of all the pieces in this dissertation: ATSP, thin trees, spectrally thin trees, the Kadison-Singer problem. It also has an overview of the main proofs without going into much detail.

Chapter 4 will introduce the Kadison-Singer problem, strongly Rayleigh measures, and our extension of Kadison-Singer to such measures. This chapter can be read by itself, but it is recommended to read about real stable polynomials in the preliminaries beforehand.

Chapter 5 will provide an abstract framework that this dissertation fits into, and provides an example problem, finding bounded degree spanning trees, for which this framework provides nontrivial results.

Chapter 6 goes over the construction of hierarchical decompositions. It is proved in this section that any graph has induced subgraphs that weakly expand. The results of this chapter might be of independent interest.

Chapter 7 is the main technical chapter. This chapter goes over the analysis of the convex programs used for preconditioning in the Kadison-Singer problem. In this chapter it is proved that k-edge-connected graphs can always be "massaged" into a form suitable for the application of Kadison-Singer.

Chapter 8 addresses the issue of finding the guaranteed ATSP tour in polynomial time.

1.5 Bibliographic Notes

The results in this dissertation were derived in collaboration with Shayan Oveis Gharan. Some of these results have already been published [AO14; AO15]. I thank my collaborator, Shayan Oveis Gharan, for allowing the inclusion of coauthored work in this dissertation.

Chapter 2

Preliminaries

2.1 Basic Notations

For an integer $n \geq 1$ we use $[n]$ to denote the set $\{1, \dots, n\}$. We also use $\binom{[n]}{k}$ to denote the set of subsets of size k from $\{1, \dots, n\}$. We write $2^{[m]}$ to denote the family of all subsets of the set $[m]$.

We write ∂_{z_i} to denote the operator that performs partial differentiation with respect to z_i.

We use $\mathbf{1}$ to denote the all 1 vector.

For a matrix $A \in \mathbb{R}^{m \times n}$ we write A_i to denote the i-th column of A, A^i to denote the i-th row of A, and $A_{i,j}$ to denote the i, j-th entry of A.

Given a graph $G = (V, E)$, for a set $S \subseteq V$, we use $G[S]$ to denote the induced subgraph of G on S. Besides the ATSP graphs, all graphs that we work with are unweighted with no loops; but they may have an arbitrary number of parallel edges between every pair of vertices. Throughout the dissertation we assume that there is an arbitrary but fixed ordering on the edges of G.

For an edge $e = \{u, v\}$ we define the vector $\chi_e = \mathbf{1}_u - \mathbf{1}_v$, where $\mathbf{1}_u$ is the u-th element of the standard basis. We also write the Laplacian of the edge $e = \{u, v\}$ as

$$L_e = L_{u,v} = \chi_{u,v}\chi_{u,v}^\intercal.$$

We use $\chi \in \mathbb{R}^{V \times E}$ to denote the matrix where the e-th column is χ_e.

For disjoint sets $S, T \subseteq V$ we write

$$E(S, T) := \{\{u, v\} : u \in S, v \in T\}.$$

We say two sets $S, T \subseteq V$ cross if $S \cap T, S \setminus T, T \setminus \neq \emptyset$.

For a set S of elements we write $\mathbb{E}_{e \sim S}[.]$ to denote the expectation under the uniform distribution over the elements of S.

We think of a permutation of a set S as a bijection mapping the elements of S to $1, 2, \dots, |S|$.

For a vector $x \in \mathbb{R}^d$, we write

$$||x|| = \sqrt{\sum_{i=1}^{d} x_i^2},$$

$$||x||_1 = \sum_{i=1}^{d} |x_i|.$$

We will use the following inequality in many places: For any sequence of nonnegative numbers $a_1, \dots, a_m$ and $b_1, \dots, b_m$

$$\min_{1 \le i \le m} \frac{a_i}{b_i} \le \frac{a_1 + a_2 + \cdots + a_m}{b_1 + b_2 + \cdots + b_m} \le \max_{1 \le i \le m} \frac{a_i}{b_i}. \tag{2.1}$$

2.2 High-Dimensional Geometry

For $x \in \mathbb{R}^d$ and $r \in \mathbb{R}$, an L_1 ball is the set of points at L_1 distance less than r of x,

$$B(x, r) := \{y \in \mathbb{R}^d : 0 < \left\|x - y\right\|_1 < r\}.$$

Unless otherwise specified, any ball that we consider in this dissertation is an L_1 ball. We may also work with L_2 or L_2^2 balls and by that we are referring to a set of points whose L_2 or L_2^2 distance from a center is bounded by r.

An L_1 *hollowed* ball is a ball with part of it removed; for $0 \le r_1 < r_2$, we define the hollowed ball $B(x, r_1||r_2)$ as follows:

$$B(x, r_1||r_2) := \{y \in \mathbb{R}^d : r_1 < \left\|x - y\right\|_1 < r_2\}.$$

Observe that $B(x, r) = B_1(x, 0||r)$. The *width* of $B(x, r_1||r_2)$ is $r_2 - r_1$.

We say a point $y \in \mathbb{R}^d$ is inside a hollowed ball $B = B(x, r_1||r_2)$ if

$$r_1 < \left\|x - y\right\|_1 < r_2,$$

and we say it is outside of B otherwise. We also say a (hollowed) ball B_1 is inside a (hollowed) ball B_2 if every point $x \in B_1$ is also in B_2.

For a (finite) set of points $S \subseteq \mathbb{R}^d$, the L_1 diameters of S, $\mathrm{diam}(S)$ is defined as the maximum L_1 distance between points in S,

$$\mathrm{diam}(S) = \max_{x,y \in S} \left\|x - y\right\|_1.$$

For a set S of elements we say $X : S \to \mathbb{R}^h$ is an L_2^2 metric if for any three elements $u, v, w \in S$,

$$\left\|X_u - X_w\right\|^2 \le \left\|X_u - X_v\right\|^2 + \left\|X_v - X_w\right\|^2.$$

A cut metric of S is a mapping $X : S \to \{0,1\}^h$ equipped with the L_1 metric. Note that any cut metric of S is also a L_2^2 metric because for any two elements $u, v \in S$,

$$\left\|X_u - X_v\right\|_1 = \left\|X_u - X_v\right\|^2.$$

Similarly, we define a weighted cut metric, $X : S \to \{0,1\}^h$ together with nonnegative weights $w_1, \ldots, w_h$, to be the be the set of points $\{X_v\}_{v \in S}$ equipped with the weighted L_1 norm:

$$\|x\|_1 = \sum_{i=1}^{h} w_i \cdot |x_i|, \text{ for all } x \in \mathbb{R}^h.$$

If all the weights are 1 we simply get an (unweighted) cut metric. It is easy to see that any weighted cut metric can be embedded, with arbitrarily small loss, (up to scaling) in an unweighted cut metric of a (possibly) higher dimension.

We can look at an embedding X as a matrix where there is a column X_u for any vertex u. We also write

$$\mathsf{X} = X\chi.$$

Therefore, for any edge $e = \{u, v\} \in E$ (oriented from u to v),

$$\mathsf{X}_e = X\chi_e = X_u - X_v.$$

2.3 Linear Algebra

We use I to denote the identity matrix and J to denote the all 1's matrix.

A matrix $U \in \mathbb{R}^{n \times n}$ is called orthogonal/unitary if $UU^\intercal = U^\intercal U = I$. An orthogonal matrix is a nonsingular square matrix whose singular values are all 1. It follows by definition that orthogonal operators preserve L_2 norms of vectors, i.e., for any vector $x \in \mathbb{R}^n$,

$$\left\|Ux\right\| = \sqrt{(Ux)^\intercal Ux} = \sqrt{x^\intercal U^\intercal Ux} = \sqrt{x^\intercal x} = \|x\|.$$

A (not necessarily square) matrix U is called semiorthogonal if $UU^\intercal = I$, i.e. the rows are orthonormal, and the number of rows is less than the number of columns. For any semiorthogonal $U \in \mathbb{R}^{m \times n}$, we can extend U to an actual orthogonal matrix by adding $n - m$ rows.

For two matrices A, B of the same dimension we define the matrix inner product $A \bullet B := \mathrm{Tr}(AB^\intercal)$. For any matrix $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$,

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

For any two matrices $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times m}$, the nonzero eigenvalues of AB and BA are the same with the same multiplicities.

Lemma 2.1. *If A, B are positive semidefinite matrices of the same dimension, then*

$$\mathrm{Tr}(AB) \geq 0.$$

Proof.

$$\mathrm{Tr}(AB) = \mathrm{Tr}(AB^{1/2}B^{1/2}) = \mathrm{Tr}(B^{1/2}AB^{1/2}) \geq 0.$$

□

Also, we use the fact that for any positive semidefinite matrix A and any Hermitian matrix B, BAB is positive semidefinite.

Lemma 2.2. *If $A, B \in \mathbb{R}^{n\times n}$ are PD matrices and $A \preceq B$, then $B^{-1} \preceq A^{-1}$.*

Proof. Since $A \preceq B$,

$$B^{-1/2}AB^{-1/2} \preceq B^{-1/2}BB^{-1/2} = I.$$

So,

$$B^{1/2}A^{-1}B^{1/2} = (B^{-1/2}AB^{-1/2})^{-1} \succeq I$$

Multiplying both sides of the above by $B^{-1/2}$, we get

$$A^{-1} = B^{-1/2}B^{1/2}A^{-1}B^{1/2}B^{-1/2} \succeq B^{-1/2}IB^{-1/2} = B^{-1}.$$

□

Fact 2.3 (Schur's Complement [BV06, Section A.5])**.** *For any symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ a (column) vector $x \in \mathbb{R}^n$ and $c \geq 0$, we have $x^\intercal A^{-1}x \leq c$ if and only if*

$$\begin{bmatrix} c & x^\intercal \\ x & A \end{bmatrix} \succeq 0.$$

The following lemma proving the operator-convexity of the inverse of PD matrices is well-known.

Lemma 2.4. *For any two symmetric $n \times n$ matrices $A, B \succ 0$,*

$$\left(\frac{1}{2}A + \frac{1}{2}B\right)^{-1} \preceq \frac{1}{2}A^{-1} + \frac{1}{2}B^{-1}.$$

Proof. For any vector $x \in \mathbb{R}^n$,

$$\frac{1}{2}\begin{bmatrix} x^\intercal A^{-1}x & x^\intercal \\ x & A \end{bmatrix} + \frac{1}{2}\begin{bmatrix} x^\intercal B^{-1}x & x^\intercal \\ x & B \end{bmatrix} = \begin{bmatrix} \frac{1}{2}x^\intercal A^{-1}x + \frac{1}{2}x^\intercal B^{-1}x & x^\intercal \\ x & \frac{1}{2}A + \frac{1}{2}B \end{bmatrix}.$$

By the Schur's complement lemma both of the matrices on the LHS of above equality are PSD. Therefore, by convexity of PSD matrices, the matrix in RHS is also PSD. By another application of Schur complement to the matrix in RHS we obtain the lemma. □

For a matrix M, we write $\|M\| = \max_{\|x\|=1} \|Mx\|$ to denote the operator norm of M. There are other matrix norms that will be used throughout the dissertation.

Definition 2.5 (Matrix Norms). *The trace norm (or nuclear norm) of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as follows:*

$$\|A\|_* := \mathrm{Tr}((A^\intercal A)^{1/2}) = \sum_{i=1}^{\min\{m,n\}} \sigma_i,$$

where σ_i's are the singular values of A. The Frobenius norm of A is defined as follows:

$$\|A\|_F := \sqrt{\sum_{1\le i\le m, 1\le j\le n} A_{i,j}^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}.$$

The following lemma is a well-known fact about the trace norm.

Lemma 2.6. *For any matrix $A \in \mathbb{R}^{n\times m}$ such that $n \ge m$,*

$$\|A\|_* = \max_{\textit{Semiorthogonal } U} \mathrm{Tr}(UA),$$

where the maximum is over all semiorthogonal matrices $U \in \mathbb{R}^{m\times n}$. In particular, $\mathrm{Tr}(A) \le \|A\|_*$.

Proof. Let the singular value decomposition of A be the following

$$A = \sum_{i=1}^{m} \sigma_i u_i v_i^\intercal,$$

where $s_1, \ldots, s_m$ are the singular values and $u_1, \ldots, u_m \in \mathbb{R}^n$ are the left singular vectors and $v_1, \ldots, v_m \in \mathbb{R}^m$ are the right singular vectors. Now let

$$U = \sum_{i=1}^{m} v_i u_i^\intercal.$$

It is easy to observe that $U \in \mathbb{R}^{m\times n}$ is semiorthogonal, i.e. $UU^\intercal = I$. Now observe that

$$UA = \sum_{i=1}^{m} \sigma_i v_i \langle u_i, u_i\rangle v_i^\intercal = \sum_{i=1}^{m} \sigma_i v_i v_i^\intercal.$$

It is easy to see that $\mathrm{Tr}(UA) = \sum_{i=1}^m \sigma_i = \|A\|_*$.

It remains to prove the other side of the equation. By von Neumann's trace inequality [Mir75], for any semiorthogonal matrix $U \in \mathbb{R}^{m\times n}$ we can write

$$\mathrm{Tr}(UA) \le \sum_i 1 \cdot \sigma_i = \|A\|_*,$$

where $\sigma_1, \ldots, \sigma_m$ are the singular values of A. $\square$

Theorem 2.7 (Hoffman-Wielandt Inequality). *Let $A, B \in \mathbb{R}^{n\times n}$ have singular values $\sigma_1 \leq \sigma_2 \leq \ldots \sigma_n$ and $\sigma_1' \leq \sigma_2' \leq \ldots \leq \sigma_n'$. Then,*

$$\sum_{i=1}^{n} (\sigma_i - \sigma_i')^2 \leq \left\|A - B\right\|_F^2 .$$

For a Hermitian matrix $M \in \mathbb{C}^{d\times d}$, we write the characteristic polynomial of M in terms of a variable x as

$$\chi[M](x) = \det(xI - M).$$

We also write the characteristic polynomial in terms of the square of x as

$$\chi[M](x^2) = \det(x^2 I - M).$$

For $1 \leq k \leq n$, we write $\sigma_k(M)$ to denote the sum of all principal $k \times k$ minors of M, in particular,

$$\chi[M](x) = \sum_{k=0}^{d} x^{d-k}(-1)^k \sigma_k(M).$$

The following lemma follows from the Cauchy-Binet identity. See [MSS13b] for the proof.

Lemma 2.8. *For vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ and scalars $z_1, \ldots, z_m$,*

$$\det\left(xI + \sum_{i=1}^{m} z_i v_i v_i^{\mathsf{T}}\right) = \sum_{k=0}^{d} x^{d-k} \sum_{S \subseteq \binom{[m]}{k}} z^S \sigma_k\Big(\sum_{i\in S} v_i v_i^{\mathsf{T}}\Big).$$

In particular, for $z_1 = \ldots = z_m = -1$,

$$\det\left(xI - \sum_{i=1}^{m} v_i v_i^{\mathsf{T}}\right) = \sum_{k=0}^{d} x^{d-k}(-1)^k \sum_{S \subseteq \binom{[m]}{k}} \sigma_k\Big(\sum_{i\in S} v_i v_i^{\mathsf{T}}\Big).$$

The following is Jacboi's formula for the derivative of the determinant of a matrix.

Theorem 2.9. *For an invertible matrix A which is a differentiable function of t,*

$$\partial_t \det(A) = \det(A) \cdot \mathrm{Tr}(A^{-1}\partial_t A).$$

Lemma 2.10. *For an invertible matrix A which is a differentiable function of t,*

$$\frac{\partial A^{-1}}{\partial t} = -A^{-1}(\partial_t A)A^{-1}.$$

Proof. Differentiating both sides of the identity $A^{-1}A = I$ with respect to t, we get

$$A^{-1}\frac{\partial A}{\partial t} + \frac{\partial A^{-1}}{\partial t}A = 0.$$

Rearranging the terms and multiplying with A^{-1} gives the lemma's conclusion. □

2.4 Graph Theory

For a graph $G = (V, E)$, and a set $S \subseteq V$, we define

$$\phi_G(S) := \frac{\partial_G(S)}{d_G(S)}$$

where $\partial_G(S) := |E(S, V \setminus S)|$ is the number of edges that leave S, and $d_G(S)$ is the sum of the degrees (in G) of vertices of S. Note that, by definition, $d_G(v) = \partial_G(\{v\})$ for any vertex. If the graph is clear in the context we drop the subscript G. The expansion of G is defined as follows:

$$\phi(G) := \min_{S \subset V} \frac{\partial_G(S)}{\min\{d_G(S), d_G(V \setminus S)\}} = \min_{S \subset V} \max\{\phi_G(S), \phi_G(V \setminus S)\},$$

We say a graph G is an ϵ-expander, if $\phi(G) \geq \epsilon$. Recall that in an expander graph, $\phi(G) = \Omega(1)$.

An (unweighted) graph $G = (V, E)$ is k-edge-connected if and only if for any pair of vertices $u, v \in V$, there are at least k edge-disjoint paths between u, v in G. Equivalently, G is k-edge-connected if for any set $\emptyset \subsetneq S \subsetneq V$, $\partial(S) \geq k$.

There is a well-known theorem by Nash-Williams that gives an almost (up to a factor of 2) necessary and sufficient condition for k-connectivity.

Theorem 2.11 ([Nas61]). *For any k-edge-connected graph, $G = (V, E)$, there are at least $k/2$ disjoint spanning trees in G.*

Note that any union of $k/2$ edge-disjoint spanning trees is a $k/2$-edge-connected graph. So, the above theorem does not give a necessary and sufficient condition for k-connectivity. A cycle gives a tight example for the loss of 2 in the above theorem.

Given a graph $G = (V, E)$, and a set $S \subseteq V$, we write G/S to denote the graph where the set S is *contracted*, i.e., we remove all vertices $v \in S$ and add a new vertex u instead, and for any vertex $w \notin S$, we let $|E(S, \{w\})|$ be the number of (parallel) edges between u and w. We also remove any self-loops that result from this operation. The following fact will be used throughout the dissertation.

Fact 2.12. *For any k-edge-connected graph $G = (V, E)$ and any set $S \subseteq V$, G/S is k-edge-connected.*

Throughout the dissertation we may use a natural decomposition of a graph G (that is not necessarily k-edge-connected) into k-edge-connected subgraphs as defined below.

Definition 2.13. *For a graph $G = (V, E)$ a natural decomposition into k-edge-connected subgraphs is defined as follows: Start with a partition $S_1 = V$. While there is a nonempty set S_i in the partition such that $G[S_i]$ is not k-edge-connected, find an induced cut $(S_{i,1}, S_{i,2})$ in $G[S_i]$ of size less than k, remove S_i and add $S_{i,1}, S_{i,2}$ as new sets in the partition.*

The following fact follows directly from the above definition.

Lemma 2.14. *For any natural decomposition of a graph $G = (V, E)$ into k-edge-connected subgraphs $S_1, \dots, S_\ell$ and any $I \subseteq [\ell]$,*

$$\sum_{i_1, i_2 \in I : i_1 < i_2} |E(S_{i_1}, S_{i_2})| \leq (k-1)(|I| - 1).$$

Consequently,

$$\sum_{i=1}^{\ell} \partial(S_i) = 2 \sum_{i_1, i_2 \in [\ell] : i_1 < i_2} |E(S_{i_1}, S_{i_2})| \leq 2(k-1)(\ell - 1).$$

Proof. Let $S = \cup_{i \in I} S_i$.

A natural decomposition of the induced subgraph, $G[S]$ into k-edge-connected subgraphs gives exactly all set S_i where $i \in I$. This decomposition partitions $G[S]$ exactly $|I| - 1$ times and each time adds at most $k - 1$ new edges between the sets in the partition. □

2.5 Polynomials and Real Stability

Stable polynomials are natural multivariate generalizations of real-rooted univariate polynomials. For a complex number z, let $\mathrm{Im}(z)$ denote the imaginary part of z. We say a polynomial $p(z_1, \dots, z_m) \in \mathbb{C}[z_1, \dots, z_m]$ is *stable* if whenever $\mathrm{Im}(z_i) > 0$ for all $1 \leq i \leq m$, $p(z_1, \dots, z_m) \neq 0$. We say $p(.)$ is real stable, if it is stable and all of its coefficients are real. It is easy to see that any univariate polynomial is real stable if and only if it is real rooted.

One of the most interesting classes of real stable polynomials is the class of determinant polynomials as observed by Borcea and Brändén [BB08].

Theorem 2.15. *For any set of positive semidefinite matrices $A_1, \dots, A_m$, the following polynomial is real stable:*

$$\det\left(\sum_{i=1}^{m} z_i A_i\right).$$

Perhaps the most important property of stable polynomials is that they are closed under several elementary operations like multiplication, differentiation, and substitution. We will use these operations to generate new stable polynomials from the determinant polynomial. The following is proved in [MSS13b].

Lemma 2.16. *If $p \in \mathbb{R}[z_1, \dots, z_m]$ is real stable, then so are the polynomials $(1 - \partial_{z_1}) p(z_1, \dots, z_m)$ and $(1 + \partial_{z_1}) p(z_1, \dots, z_m)$.*

The following corollary simply follows from the above lemma.

Corollary 2.17. *If $p \in \mathbb{R}[z_1, \dots, z_m]$ is real stable, then so is*

$$(1 - \partial_{z_1}^2) p(z_1, \dots, z_m).$$

Proof. First, observe that

$$(1 - \partial_{z_1}^2) p(z_1, \dots, z_m) = (1 - \partial_{z_1})(1 + \partial_{z_1}) p(z_1, \dots, z_m).$$

The conclusion follows from two applications of lemma 2.16. □

The following closure properties are elementary.

Lemma 2.18. *If $p \in \mathbb{R}[z_1, \dots, z_m]$ is real stable, then so is $p(\lambda \cdot z_1, \dots, \lambda_m \cdot z_m)$ for real-valued $\lambda_1, \dots, \lambda_m > 0$.*

Proof. Say $(z_1, \dots, z_m) \in \mathbb{C}^m$ is a root of $p(\lambda \cdot z_1, \dots, \lambda_m \cdot z_m)$. Then $(\lambda_1 \cdot z_1, \dots, \lambda_m \cdot z_m)$ is a root of $p(z_1, \dots, z_m)$. Since p is real stable, there is an i such that $\operatorname{Im}(\lambda_i \cdot z_i) \leq 0$. But, since $\lambda_i > 0$, we get $\operatorname{Im}(z_i) \leq 0$, as desired. □

Lemma 2.19. *If $p \in \mathbb{R}[z_1, \dots, z_m]$ is real stable, then so is $p(z_1 + x, \dots, z_m + x)$ for a new variable x.*

Proof. Say $(z_1, \dots, z_m, x) \in \mathbb{C}^m$ is a root of $p(z_1 + x, \dots, z_m + x)$. Then $(z_1 + x, \dots, z_m + x)$ is a root of $p(z_1, \dots, z_m)$. Since p is real stable, there is an i such that $\operatorname{Im}(z_i + x) \leq 0$. But, then either $\operatorname{Im}(x) \leq 0$ or $\operatorname{Im}(z_i) \leq 0$, as desired. □

Chapter 3

ATSP and Goddyn's Thin Tree Conjecture

3.1 Connections Between ATSP and Thin Trees

In the Asymmetric Traveling Salesman Problem (ATSP) we are given a set V of $n := |V|$ vertices and a nonnegative cost function $c : V \times V \to \mathbb{R}_+$. The goal is to find the shortest tour that visits every vertex *at least* once.

There is a natural LP relaxation for ATSP proposed by Held and Karp [HK70],

$$\begin{aligned}
\min \sum_{u,v \in V} & c(u,v) x_{u,v} & & \\
\text{s.t.} \sum_{u \in S, v \notin S} & x_{u,v} \geq 1 & & \forall S \subseteq V, \\
\sum_{v \in V} x_{u,v} &= \sum_{v \in V} x_{v,u} = 1 & & \forall u \in V, \\
x_{u,v} &\geq 0 & & \forall u,v \in V.
\end{aligned} \tag{3.1}$$

Thin Trees. The main ingredient of the recent developments on ATSP is the construction of a "thin" tree. Let $G = (V, E)$ be an unweighted undirected k-edge-connected graph with n vertices. Recall that G is k-edge-connected if there are at least k edges in every cut of G, see section 2.4 for properties of k-edge-connected graphs. We allow G to have an arbitrary number of parallel edges, so we think of E as a multiset of edges. Roughly speaking, a spanning tree $T \subseteq E$ is α-thin with respect to G if it does not contain more than α-fraction of the edges of any cut in G.

Definition 3.1. *A spanning tree $T \subseteq E$ is α-thin with respect to a (unweighted) graph $G = (V, E)$, if for each set $S \subseteq V$,*

$$|T(S, \overline{S})| \leq \alpha \cdot |E(S, \overline{S})|,$$

where $T(S, \overline{S})$ and $E(S, \overline{S})$ are the set of edges of T and G in the cut $(S, \overline{S})$ respectively.

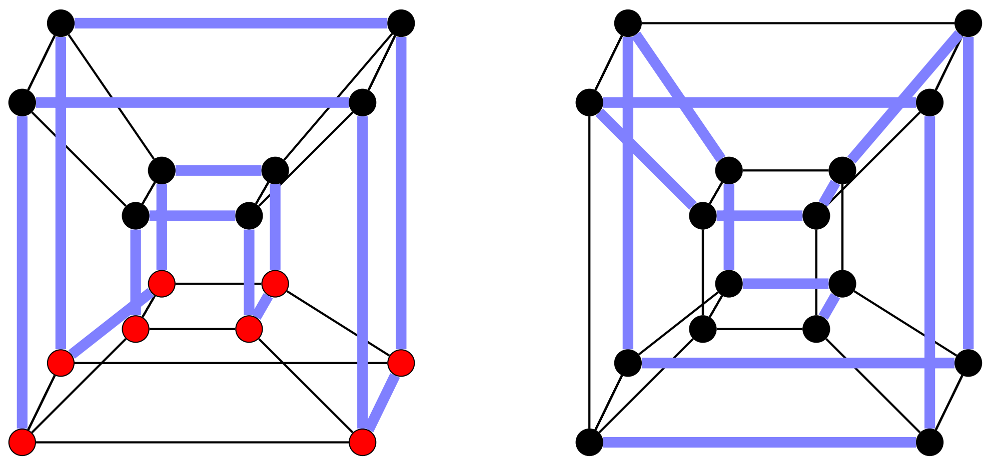

Figure 3.1: Two spanning trees of 4-dimensional hypercube that is 4-edge-connected. Although both of the trees are Hamiltonian paths, the left spanning tree is 1-thin because all of the edges of the cut separating red vertices from the black ones are in the tree while the right spanning tree is 0.667-thin.

One can analogously define α-thin edge covers, α-thin paths, etc. Note that thinness is a downward closed property, that is any subgraph of an α-thin subgraph of G is also α-thin. In particular, any spanning tree of an α-thin connected subgraph of G is an α-thin spanning tree of G. See fig. 3.1 for two examples of thin trees.

A key lemma in [Asa+10] shows that one can obtain an approximation algorithm for ATSP by finding a thin tree of small cost with respect to the graph defined by the fractional solution of the LP relaxation. In addition, proving the existence of a thin tree provides a bound on the integrality gap of the Held-Karp LP relaxation for ATSP.

Later, in [OS11] this connection is made more concrete. Namely, to break the $\Theta(\frac{\log(n)}{\log\log(n)})$ barrier, it suffices to ignore the costs of the edges and construct a thin tree in every k-edge-connected graph for $k = \Theta(\log(n))$.

Theorem 3.2. *For any $\alpha > 0$ (which can be a function of n), and $k \geq \log n$, a polynomial-time construction of an α/k-thin tree in any k-edge-connected graph gives an $O(\alpha)$-approximation algorithm for ATSP. In addition, even an existential proof gives an $O(\alpha)$ upper bound on the integrality gap of the LP relaxation.*

See the end of this chapter for the proof of the above theorem. The above theorem shows that to understand the solutions of LP (3.1) it is enough to understand the thin tree problem in graphs with low connectivity.

It is easy to show that any k-edge-connected graph has an $O(\log(n)/k)$-thin tree [Goe+09] using the independent randomized rounding method of Raghavan and Thompson [RT87]. It is enough to sample each edge of G independently with probability $\Theta(\log(n)/k)$ and then choose an arbitrary spanning tree of the sampled graph. [Asa+10] employ a more sophisticated randomized rounding algorithm and show that any k-edge-connected graph has a $\frac{\log(n)}{k \cdot \log\log(n)}$-thin tree. The basic idea of their algorithm is to use a correlated distribution, that is to sample edges almost independently while preserving the connectivity of the sampled set. More precisely, they sample a random spanning tree from a distribution where the edges are negatively correlated, so they get connectivity for free, and they only use the upper tail of the Chernoff types of bounds. The $1/\log\log(n)$ gain comes from the fact that the upper tail of the Chernoff bound is slightly stronger than the lower tail,

Independently of the above applications of thin trees, Goddyn formulated the thin tree conjecture because of the close connections to several long-standing open problems regarding nowhere-zero flows.

Conjecture 3.3 (Goddyn [God04]). *There exists a function $f(\alpha)$ such that, for any $0 < \alpha < 1$, every $f(\alpha)$-edge-connected graph (of arbitrary size) has an α-thin spanning tree.*

Goddyn's conjecture in the strongest form postulates that for a sufficiently large k that is independent of the size of G, every k-edge-connected graph has an $O(1/k)$-thin tree. Goddyn proved that if the above conjecture holds for an arbitrary function $f(.)$, it implies a weaker version of Jaeger's conjecture on the existence of circular nowhere-zero flows [Jae84]. Very recently, Thomassen proved a weaker version of Jaeger's conjecture [Tho12; Lov+13], but his proof has not yet shed any light on the resolution of the thin tree conjecture.

To this date, Conjecture 3.3 is only proved for planar and bounded genus graphs [OS11; ES14] and edge-transitive graphs[1] [MSS13b; HO14] for $f(\alpha) = O(1/\alpha)$. We remark that if Goddyn's thin tree conjecture holds for an arbitrary function $f(.)$, we get an upper bound of $O(\log^{1-\Omega(1)}(n))$ on the integrality gap of the LP relaxation of ATSP.

Summary of our Contribution. In this dissertation, we show that any k-edge-connected graph has a $\text{poly}\log\log(n)/k$-thin tree. Using theorem 3.2 for $\alpha = \text{poly}\log\log(n)$ and $k = \log(n)$ this implies that the integrality gap of the LP relaxation is $\text{poly}\log\log(n)$. Note that this does not resolve Goddyn's conjecture. Perhaps, one of the main consequences of our work is that we can round (not necessarily in polynomial time) the solutions of the LP relaxation exponentially better than the randomized rounding in the worst case.

The key to our proof is to rigorously relate the thin tree problem to a seemingly related spectral question that is known as the Kadison-Singer problem in operator theory [Wea04] and then to use tools in spectral (graph) theory to solve the new problem. Until very recently, the best solution to the Kadison-Singer problem and the Weaver conjecture was based on the randomized rounding technique and matrix Chernoff bounds and incurred a loss of $\log(n)$ [Rud99; AW02]. Marcus, Spielman, and Srivastava [MSS13b] in a breakthrough managed to resolve the conjecture using spectral techniques with no cost that is dependent on n. As we will elaborate in the next section, the Kadison-Singer problem can be seen as an "L_2"

[1] A graph $G = (V, E)$ is edge-transitive, if for any pair of edges $e, f \in E$ there is an automorphism of G that maps e to f.

version of the thin tree question, or thin tree question can be seen as an L_1 version of the Kadison-Singer problem. So, we can summarize our contribution as an L_1 to L_2 reduction.

We construct this L_1 to L_2 reduction using a convex program that symmetrizes the L_2 structure of a given graph while preserving its L_1 structure. More precisely, a convex program that equalizes the *effective resistance* of the edges while preserving the cut structure of G. We expect to see several other applications of this convex program in combinatorial optimization and approximation algorithms. In addition to that, we extend the result of Marcus, Spielman, and Srivastava to a larger family of distributions known as *strongly Rayleigh* distributions. We will discuss this in more details in chapter 4

The rest of this section is organized as follows: In section 3.2 we overview the connections of the thin tree problem and graph sparsifiers and in particular the Kadison-Singer problem. Then, in section 3.3 we present our main theorems. Finally, in section 3.4 we highlight the main ideas of the proof.

3.2 Spectrally Thin Trees

As mentioned before, thin trees are the basis for the best-known approximation algorithms for ATSP on planar, bounded genus, or general graphs. This follows from their intuitive definition and the fact that they eliminate the difficulty arising from the underlying asymmetry and the cost function. On the other hand, the major challenge in constructing thin trees or proving their existence is that we are not aware of any efficient algorithm for measuring or certifying the thinness of a given tree exactly. In order to verify the thinness of a given tree, it seems that one has to look at exponentially many cuts.

One possible way to avoid this difficulty is to study a stronger definition of thinness, namely the *spectral* thinness. First, we define some notation. For a set $S \subseteq V$ we use $\mathbf{1}_S \in \mathbb{R}^V$ to denote the indicator (column) vector of the set S. For a vertex $v \in V$, we abuse notation and write $\mathbf{1}_v$ instead of $\mathbf{1}_{\{v\}}$. For any edge $e = \{u, v\} \in E$ we fix an arbitrary orientation, say $u \to v$, and we define $\chi_e := \mathbf{1}_u - \mathbf{1}_v$. The Laplacian of G, L_G, is defined as follows:

$$L_G := \sum_{e \in E} \chi_e \chi_e^{\intercal}.$$

If G is weighted, then we scale up each term $\chi_e \chi_e^{\intercal}$ according to the weight of the edge e. Also, for a set $T \subseteq E$ of edges, we write

$$L_T := \sum_{e \in T} \chi_e \chi_e^{\intercal}.$$

We say a spanning tree, T, is α-spectrally thin with respect to G if

$$L_T \preceq \alpha \cdot L_G, \text{ i.e., for all } x \in \mathbb{R}^n,\ x^{\intercal} L_T x \leq \alpha \cdot x^{\intercal} L_G x. \tag{3.2}$$

We also say G has a spectrally thin tree if it has an α-spectrally thin tree for some $\alpha < 1/2$. Observe that if T is α-spectrally thin, then it is also α-(combinatorially) thin. To see that, note that for any set $S \subseteq V$, $\mathbf{1}_S^{\intercal} L_T \mathbf{1}_S = |T(S, \overline{S})|$ and $\mathbf{1}_S^{\intercal} L_G \mathbf{1}_S = |E(S, \overline{S})|$.

One can verify spectral thinness of T (in polynomial time) by finding the smallest $\alpha \in \mathbb{R}$ such that

$$L_G^{\dagger/2} L_T L_G^{\dagger/2} \preceq \alpha \cdot I,$$

i.e., by computing the largest eigenvalue of $L_G^{\dagger/2} L_T L_G^{\dagger/2}$. Recall that $L_G^\dagger$ is the pseudoinverse of L_G, and $L_G^{\dagger/2}$ is the square root of the pseudoinverse of L_G; $L_G^{\dagger/2}$ is well-defined because $L_G^\dagger \succeq 0$. So, unlike the combinatorial thinness, spectral thinness can be computed *exactly* in polynomial time.

The notion of spectral thinness is closely related to spectral sparsifiers of graphs, which have been studied extensively in the past few years [ST04; SS11; BSS14; Fun+11]. Roughly speaking, a spectrally thin tree is a one-sided spectral sparsifier. A spectrally thin tree T would be a true spectral sparsifier if in addition to (3.2), it satisfies $\alpha \cdot (1-\epsilon) x^\intercal L_G x \preceq L_T$ for some constant ϵ. Until the recent breakthrough of Batson, Spielman, and Srivastava, all constructions of spectral sparsifiers used at least $\Omega(n\log(n))$ edges of the graph [ST04; SS11; Fun+11]. Because of this they are of no use for the particular application of ATSP. Batson, Spielman, and Srivastava [BSS14] managed to construct a spectral sparsifier that uses only $O(n)$ edges of G. But in their construction, they assign different weights to the edges of the sparsifier which again makes their contribution not helpful for ATSP.

Indeed, it was observed by several people that there is an underlying barrier for the construction of spectrally thin trees and *unweighted* spectral sparsifiers. Many families of k-edge-connected graphs do not admit spectrally thin trees (see [HO14, Thm 4.9]). Let us elaborate on this observation. The *effective resistance* of an edge $e = \{u, v\}$ in G, $\mathcal{R}\mathrm{eff}_{L_G}(e)$, is the *energy* of the electrical flow that sends 1 unit of current from u to v when the network represents an electrical circuit with each edge being a resistor of resistance 1 (and if G is weighted, the resistance is the inverse of the weight of e). See [LP13, Ch. 2] for background on electrical flows and effective resistance. Mathematically, the effective resistance can be computed using $L_G^\dagger$,

$$\mathcal{R}\mathrm{eff}_{L_G}(e) := \chi_e^\intercal L_G^\dagger \chi_e.$$

It is not hard to see that the spectral thinness of any spanning tree T of G is at least the maximum effective resistance of the edges of T in G.

Lemma 3.4. *For any graph $G = (V, E)$, the spectral thinness of any spanning tree $T \subseteq E$ is at least* $\max_{e\in T} \mathcal{R}\mathrm{eff}_{L_G}(e)$.

Proof. Say the spectral thinness of T is α. Obviously, by the downward closedness of spectral thinness, the spectral thinness of any subset of edges of T is at most α, i.e., for any edge $e \in T$,

$$L_{\{e\}} \preceq L_T \preceq \alpha \cdot L_G.$$

But, the spectral thinness of an edge is indeed its effective resistance. More precisely, multiplying $L_G^{\dagger/2}$ on both sides of the above inequality we have

$$L_G^{\dagger/2} \chi_e \chi_e^\intercal L_G^{\dagger/2} = L_G^{\dagger/2} L_{\{e\}} L_G^{\dagger/2} \preceq \alpha \cdot L_G^{\dagger/2} L_G L_G^{\dagger/2} \preceq \alpha \cdot I.$$

$$\begin{tikzcd}
G \text{ is } k\text{-edge-connected} \arrow[r, dashed, "\text{Thin tree conjecture}"] & G \text{ has } O(1/k)\text{-thin tree} \\
\max_{e\in E} \mathcal{R}\mathrm{eff}_{L_G}(e) \le 1/k \arrow[u, Rightarrow] \arrow[r, Rightarrow, "{[\text{MSS13b}]}"'] & G \text{ has } O(1/k)\text{-spectrally thin tree} \arrow[u, Rightarrow]
\end{tikzcd}$$

Figure 3.2: A summary of the relationship between spectrally thin trees and combinatorially thin trees before this work.

Since the matrix on the LHS has rank one, its only eigenvalue is equal to its trace; therefore,

$$\mathrm{Tr}(\chi_e^\intercal L_G^\dagger \chi_e) = \mathrm{Tr}(L_G^{\dagger/2}\chi_e\chi_e^\intercal L_G^{\dagger/2}) \le \alpha.$$

The lemma follows by the fact that $\mathcal{R}\mathrm{eff}_{L_G}(e) = \mathrm{Tr}(\chi_e^\intercal L_G^\dagger \chi_e)$. □

In light of the above lemma, a necessary condition for G to have a spanning tree with spectral thinness bounded away from 1 is that every cut of G must have at least one edge with effective resistance bounded away from 1. In other words, any graph G with at least one cut where the effective resistance of every edge is very close to 1 has no spectrally thin tree (see fig. 3.3 for an example of a graph where the effective resistance of every edge in a cut is very close to 1).

In a very recent breakthrough, Marcus, Spielman, and Srivastava [MSS13b] proved the Kadison-Singer conjecture. As a byproduct of their result, it was shown in [HO14] that a stronger version of the above condition is sufficient for the existence of spectrally thin trees.

Theorem 3.5 ([MSS13b]). *Any connected graph $G = (V, E)$ has a spanning tree with spectral thinness $O(\max_{e\in E} \mathcal{R}\mathrm{eff}_{L_G}(e))$.*

See [HO14, Appendix E] for a detailed proof of the above theorem. It follows from the above theorem that every k-edge-connected edge-transitive graph has an $O(1/k)$-spectrally thin tree. This is because in any edge-transitive graph, by symmetry, the effective resistances of all edges are equal.

Let us summarize the relationship between spectrally thin trees and combinatorially thin trees that has been in the literature before our work. Goddyn conjectured that every k-edge-connected graph has an $O(1/k)$-thin tree. The result of [MSS13b] shows that a stronger assumption implies an stronger conclusion, i.e., if the maximum effective resistance of edges of G is at most $1/k$, then G has an $O(1/k)$-spectrally thin tree (see fig. 3.2).

We emphasize that $\max_{e\in E} \mathcal{R}\mathrm{eff}_{L_G}(e) \le 1/k$ is a stronger assumption than k-edge-connectivity. If $\mathcal{R}\mathrm{eff}_{L_G}(u, v) \le 1/k$, it means that when we send one unit of flow from u to v, the electric current divides and goes through at least k parallel paths connecting u to v, so, there are k edge-disjoint paths between u, v. But the converse of this does not necessarily hold. If there are k edge-disjoint paths from u to v, the electric current may just use one of these paths if the rest are very long, so the effective resistance can be very close to 1. Therefore, if $\max_{e\in E} \mathcal{R}\mathrm{eff}_{L_G}(e) \leq 1/k$, there are k edge-disjoint paths between each pair of vertices of G, and G is k-edge-connected, but the converse does not necessarily hold. For example in the graph in the top of fig. 3.3, even though there are k edge-disjoint paths from u_1 to v_1, a unit electrical flow from u_1 to v_1 almost entirely goes through the edge $\{u_1, v_1\}$, so $\mathcal{R}\mathrm{eff}(u_1, v_1) \approx 1$.

As a side remark, note that the sum of effective resistances of all edges of any connected graph G is $n-1$,

$$\sum_{e\in E} \chi_e^\intercal L_G^\dagger \chi_e = \sum_{e\in E} \mathrm{Tr}(L_G^{\dagger/2}\chi_e\chi_e^\intercal L_G^{\dagger/2}) = \mathrm{Tr}\left(\sum_{e\in E} L_G^{\dagger/2}\chi_e\chi_e^\intercal L_G^{\dagger/2}\right) = \mathrm{Tr}(L_G^{\dagger/2} L_G L_G^{\dagger/2}) = n-1.$$

In the last identity we use that $L_G^{\dagger/2} L_G L_G^{\dagger/2}$ is an identity matrix on the space of vectors that are orthogonal to the all-1s vector.

If G is k-edge-connected, by Markov's inequality, at most a quarter of the edges have effective resistance more than $8/k$. Therefore, by an application of [MSS13b], any k-edge-connected graph G has an $O(1/k)$-spectrally thin set of edges, $F \subset E$ where $|F| \geq \Omega(n)$ [HO14]. Unfortunately, the corresponding subgraph (V, F) may have $\Omega(n/k)$ connected components. So, this does not give any improved bounds on the approximability of ATSP.

3.3 Contribution to ATSP and Goddyn's Thin Tree Conjecture

In this dissertation we introduce a procedure to "transform" graphs that do not admit spectrally thin trees into those that *provably* have these trees. Then, we use the results of chapter 4 to find spectrally thin trees in the transformed "graph". Finally, we show that any spectrally thin tree of the transformed "graph" is a (combinatorially) thin tree in the original graph. From a high level perspective, our transformation massages the graph to equalize the effective resistance of the edges, while keeping the cut structure of the graph intact.

For two matrices $A, B \in \mathbb{R}^{n\times n}$, we write $A \preceq_\square B$, if for any set $\emptyset \subset S \subsetneq V$,

$$\mathbf{1}_S^\intercal A \mathbf{1}_S \leq \mathbf{1}_S^\intercal B \mathbf{1}_S.$$

Note that $A \preceq B$ implies $A \preceq_\square B$, but the converse is not necessarily true. We say a graph D is a *shortcut* graph with respect to G if $L_D \preceq_\square L_G$. We say a positive definite (PD) matrix D is a shortcut matrix with respect to G if $D \preceq_\square L_G$.

Our ideal plan is as follows: Show that there is a (weighted) shortcut graph D such that for any edge $e \in E$, $\mathcal{R}\mathrm{eff}_{L_D}(e) \leq \tilde{O}(1/k)$. Then, use a simple extension of theorem 4.1 such as [AW13] to show that there is a spanning tree $T \subseteq E$ such that

$$L_T \preceq_\square \alpha \cdot (L_G + L_D),$$

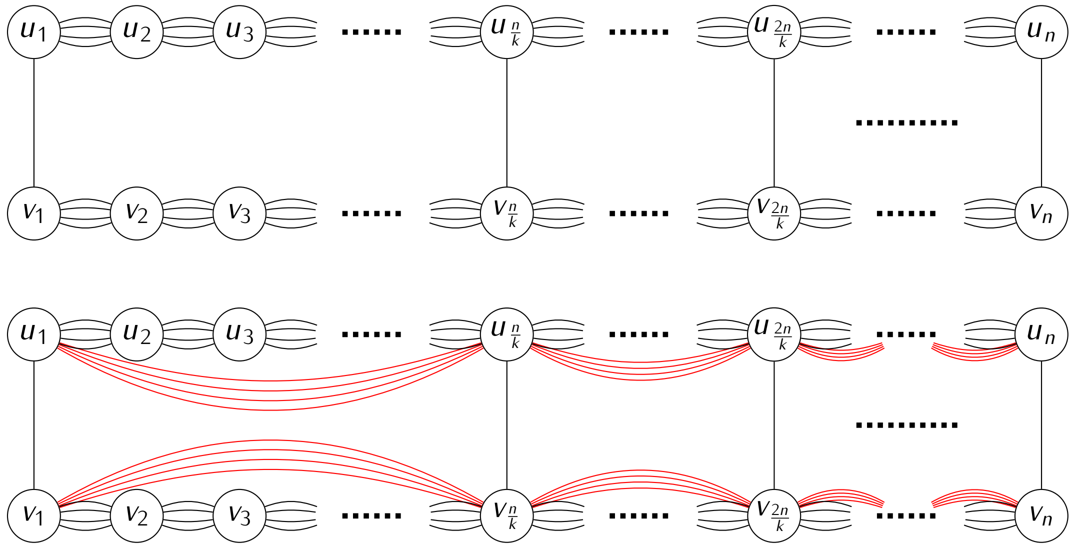

Figure 3.3: The top shows a k-edge-connected planar graph that has no spectrally thin tree. There are $k+1$ vertical edges, $(u_1, v_1), (u_{n/k}, v_{n/k}), \ldots, (u_n, v_n)$. For each $1 \le i \le n-1$ there are k parallel edges between u_i, u_{i+1} and v_i, v_{i+1}. The effective resistances of all vertical edges are $1 - O(k^2/n)$. The bottom shows a graph $G + D$ where the effective resistance of every black edge is $O(1/\sqrt{k})$. The red edges are edges in D and there are k parallel edges between the endpoints of consecutive vertical edges. Note that $L_D \preceq_\square L_G$ by construction.

for $\alpha = O(\max_{e \in E} \mathcal{R}\mathrm{eff}_{L_G + L_D}(e)) = \tilde{O}(1/k)$. But, since $L_D \preceq_\square L_G$, any α-spectrally thin tree of $D + G$ is a 2α-combinatorially thin tree of G. In summary, the graph D allows us to bypass the spectral thinness barrier that we described in lemma 3.4.

Let us give a clarifying example. Consider the k-edge-connected planar graph G illustrated at the top of fig. 3.3. In this graph, all edges in the cut $(\{v_1, \ldots, v_n\}, \{u_1, \ldots, u_n\})$ have effective resistance very close to 1. Now, let D consist of the red edges shown at the bottom. Observe that $L_D \preceq_\square L_G$. The effective resistance of every *black* edge in $G + D$ is $O(1/\sqrt{k})$. Roughly speaking, this is because the red edges *shortcut* the long paths between the endpoints of vertical edges. This reduces the energy of the corresponding electrical flows. So, $G + D$ has a spectrally thin tree $T \subseteq E$. Such a tree is combinatorially thin with respect to G.

It turns out that there are k-edge-connected graphs where it is impossible to reduce the effective resistance of all edges by a shortcut graph D (see section 7.1 for details). So, in our main theorem, we prove a weaker version of the above ideal plan. Firstly, instead of finding a shortcut graph D, we find a PD shortcut matrix D. The matrix D does not necessarily represent the Laplacian matrix of a graph as it may have positive off-diagonal entries. Secondly, the shortcut matrix reduces the effective resistance of only a set $F \subseteq E$ of edges, that we call *good* edges, where (V, F) is $\Omega(k)$-edge-connected.

Theorem 3.6 (Main). *For any k-edge-connected graph $G = (V, E)$ where $k \ge 7\log(n)$, there* *is a shortcut matrix $0 \prec D \preceq_{\square} L_G$ and a set of good edges $F \subseteq E$ such that the graph (V, F) is $\Omega(k)$-edge-connected and that for any edge $e \in F$,*

$$\mathcal{R}\text{eff}_D(e) \leq \tilde{O}(1/k),^{2}$$

where $\mathcal{R}\text{eff}_D(e) = \chi_e^{\intercal} D^{-1} \chi_e$.

Note that in the above we upper bound the effective resistance of good edges with respect to D as opposed to $D + L_G$; this is sufficient because $\mathcal{R}\text{eff}_{L_G+D}(e) \leq \mathcal{R}\text{eff}_D(e)$. We remark that the dependency on $\log(n)$ in the statement of the theorem is because of a limitation of our current proof techniques. We expect that a corresponding statement without any dependency on n holds for any k-edge-connected graph G. Such a statement would resolve Goddyn's thin tree conjecture 3.3 and may lead to improved bounds on the integrality gap of LP (3.1). Finally, the logarithmic dependency on k in the upper bound on the effective resistance of the edges of F is necessary.

Unfortunately, the good edges in the above theorem may be very sparse with respect to G, i.e., G may have cuts $(S, \overline{S})$ such that

$$|F(S, \overline{S})| \ll |E(S, \overline{S})|.$$

So, if we use theorem 4.1 or its simple extensions as in [AW13], we get a thin set of edges $T \subseteq E$ that may have $\Omega_k(n)$ many connected components. Instead, we use a theorem, that we proved in our recent extension of [MSS13b], that shows that as long as F is $\Omega(k)$-edge-connected, G has a spanning tree T that is $\tilde{O}(1/k)$-spectrally thin with respect to $D + L_G$.

Theorem 3.7 (see chapter 4). *Given a graph $G = (V, E)$, a PD matrix D and $F \subseteq E$ such that (V, F) is k-edge-connected, if for $\epsilon > 0$,*

$$\max_{e \in F} \mathcal{R}\text{eff}_D(e) \leq \epsilon,$$

then G has a spanning tree $T \subseteq F$ s.t.,

$$L_T \preceq O(\epsilon + 1/k)(D + L_G).$$

Putting theorem 3.6 and theorem 3.7 together implies that any k-edge-connected graph has a $\text{poly}\log\log(n)/k$-thin tree.

Corollary 3.8. *Any k-edge-connected graph $G = (V, E)$, has a* $\text{poly}\log\log(n)/k$*-thin tree.*

Proof. First, observe that by theorems 3.6, 3.7 any $7\log(n)$ connected graph contains a $\text{poly}\log\log(n)/\log(n)$-thin tree.

[2]For functions $f(.), g(.)$ we write $g = \tilde{O}(f)$ if $g(n) \leq \text{poly}\log(f(n)) \cdot f(n)$ for all sufficiently large n.

Now, if G is k-edge-connected and $k \gg \log(n)$, then we simply construct a $7\log(n)$ connected subgraph of G that is $7\log(n)/k$ thin by sampling each edge independently with probability $\Theta(\log n/k)$ (see the proof of theorem 3.2 for the details of the analysis). Then, we use the aforementioned statement to prove the existence of a thin tree in the sampled graph.

Otherwise, if $k \ll \log(n)$, then we add $7\log(n)/k$ copies of each edge of G and make a new graph H that is $7\log(n)$ connected, then we use the previous corollary to find a $\operatorname{poly}\log\log(n)/\log(n)$-thin tree of H. Such a tree is $\operatorname{poly}\log\log(n)/k$-thin with respect to G. $\square$

We remark that, the above theorems do not resolve Goddyn's thin tree conjecture because of the dependency on n.

At first inspection, it would seem that there are two nonalgorithmic ingredients in our proof. The first one is the exponential-sized convex program that we will use to find the shortcut matrix D; this is because verifying $D \preceq_{\square} L_G$ is equivalent to 2^n many linear constraints. Secondly, we need to have a constructive (in polynomial time) proof of theorem 3.7. The following theorem shows we can get around the first barrier.

Theorem 3.9. *Assume that there is an oracle that takes an input graph $G = (V, E)$, PD matrix D, and a k-edge-connected $F \subseteq E$, such that $\max_{e\in F} \mathcal{R}\text{eff}_D(e) \le \epsilon$, and returns the spanning tree T promised by theorem 3.7, i.e., $L_T \preceq O(\epsilon + 1/k)(D + L_G)$. For any $\ell \le \log\log n$, there is a $\operatorname{poly}\log\log(n)\cdot\log(n)^{1/\ell}$-approximation algorithm for ATSP that runs in time $n^{O(\ell)}$ (and makes at most $n^{O(\ell)}$ oracle calls).*

We will prove this theorem in section 8.1.

3.4 Main Components of the Proof

Our proof has three main components, namely the thin basis problem, the effective resistance reducing convex programs, and the locally connected hierarchies. In this section we summarize the high-level interaction of these three components.

The Thin Basis Problem. Let us start by an overview of the proof of theorem 3.7 which also appears in chapter 4.

The thin basis problem is defined as follows: Given a set of vectors $\{x_e\}_{e\in E} \in \mathbb{R}^d$, what is a sufficient condition for the existence of an α-thin basis, namely, d linearly independent set of vectors $T \subseteq E$ such that

$$\left\| \sum_{e\in T} x_e x_e^{\intercal} \right\| \le \alpha?$$

It follows from the work of Marcus, Spielman, and Srivastava [MSS13b] that a sufficient condition for the existence of an α-thin basis is that the vectors are in isotropic position,

$$\sum_{e\in E} x_e x_e^{\intercal} = I,$$

and for all $e \in E$, $\|x_e\|^2 \le c \cdot \alpha$ for some universal constant $c < 1$.

The thin basis problem is closely related to the existential problem of spectrally thin trees. Say we want to see if a given graph $G = (V, E)$ has a spectrally thin tree. We can define a vector $y_e = L_G^{\dagger/2}\chi_e$ for each edge $e \in E$. It turns out that these vectors are in isotropic position; in addition, if all edges of G have effective resistance at most ϵ, then $\|y_e\|^2 = \chi_e^\intercal L_G^\dagger \chi_e \le \epsilon$. So, these vectors contain an $O(\epsilon)$-thin basis. It is easy to see that such a basis corresponds to an $O(\epsilon)$-spectrally thin tree of G (see [AO14] for details).

As alluded to in the introduction, if G is a k-edge-connected graph, it may have many edges of large effective resistance, so $\|y_e\|^2$ in the above argument may be very close to 1. We use the shortcut matrix D that is promised in theorem 3.6 to reduce the squared norm of the vectors. We assign a vector $y_e = (L_G + D)^{-1/2}\chi_e$ to any good edge $e \in F$. It follows that

$$\|y_e\|^2 \le \chi_e^\intercal D^{-1}\chi_e \le \tilde{O}(1/k).$$

But, since the good edges are only a subset of the edges of G, the set of vectors $\{y_e\}_{e\in F}$ are not necessarily in an isotropic position; they are rather in a sub-isotropic position,

$$\sum_{e\in F} y_e y_e^\intercal \preceq I.$$

In chapter 4 we prove a weaker sufficient condition for the existence of a thin basis. If the vectors $\{x_e\}_{e\in E}$ are in a sub-isotropic position, each of them has a squared norm at most ϵ, and they contain k disjoint bases, then there exists an $O(\epsilon + 1/k)$-thin basis $T \subset E$

$$\left\|\sum_{e\in E} x_e x_e^\intercal\right\| \le O(\epsilon + 1/k).$$

Since, the set F of good edges promised in theorem 3.6 is $\Omega(k)$-edge-connected, it contains $\Omega(k)$ edge-disjoint spanning trees, so the set of vectors $\{y_e\}_{e\in F}$ defined above contains $\Omega(k)$ disjoint bases. So, $\{y_e\}_{e\in F}$ contains a $\tilde{O}(1/k)$-thin basis T; this corresponds to a $\tilde{O}(1/k)$-spectrally thin tree of $L_G + D$ and a $\tilde{O}(1/k)$-thin tree of G.

Effective Resistance Reducing Convex Programs. As illustrated in the previous section, at the heart of our proof we find a PD shortcut matrix D to reduce the effective resistance of a subset of edges of G.

It turns out that the problem of finding the best shortcut matrix D that reduces the maximum effective resistance of the edges of G is convex. This is because for any fixed vector x and $D \succ 0$, $x^\intercal D^{-1} x$ is a convex function of D. See lemma 2.4 for the proof. The problem of minimizing the sum of effective resistances of all pairs of vertices in a given graph was previously studied in [GBS08].

The following (exponentially sized) convex program finds the best shortcut matrix D that minimizes the maximum effective resistance of the edges of G while preserving the cut structure of G.

Max-CP:

$$\begin{aligned} \min \quad & \mathcal{E}, \\ \text{s.t.} \quad & \mathcal{R}\text{eff}_D(e) \le \mathcal{E} \quad \forall e \in E, \\ & D \preceq_{\square} L_G, \\ & D \succ 0. \end{aligned}$$

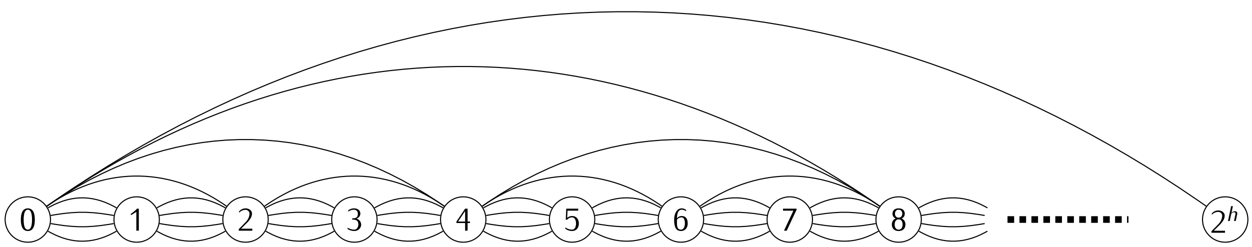

Figure 3.4: A tight example for theorem 7.3. The graph has $2^h + 1$ vertices labeled with $\{0, 1, \ldots, 2^h\}$. There are k parallel edges connecting each pair of consecutive vertices. In addition, for any $1 \le i \le h$ and any $0 \le j < 2^{h-i}$ there is an edge $\{j \cdot 2^i, (j+1) \cdot 2^i\}$.

Note that if we replace the constraint $D \preceq_{\square} L_G$ with $D \preceq L_G$, i.e., if we require D to be upper-bounded by L_G in the PSD sense, then the optimum D for any graph G is exactly L_G and the optimum value is the maximum effective resistance of the edges of G.

Unfortunately, the optimum of the above program can be very close to 1 even if the input graph G is $\log(n)$-edge-connected. A bad graph is shown in fig. 3.4. In theorem 7.3 we show that the optimum of the above convex program for the family of graphs in fig. 3.4 is close to 1 by constructing a feasible solution of the dual.

To prove our main theorem, we study a variant of the above convex program that reduces the effective resistance of only a subset of edges of G to $\tilde{O}(1/k)$. We will use combinatorial objects called locally connected hierarchies as discussed in the next paragraph to feed a carefully chosen set of edges into the convex program. To show that the optimum value of the program is $\tilde{O}(1/k)$, we analyze its dual. The dual problem corresponds to proving an upper bound on the ratio involving distances of pairs of vertices of G with respect to an L_1 embedding of the vertices in a high-dimensional space. We refrain from going into the details at this point. We will provide a more detailed overview in section 7.1.

Locally Connected Hierarchies. The main difficulty in proving theorem 3.6 is that the good edges, F, are unknown a priori. If we knew F then we could use Max-CP to minimize the maximum effective resistance of edges of F as opposed to E. In addition, the k-th smallest effective resistance of the edges of a cut of G is not a convex function of D. So, we cannot write a single program that gives us the best matrix D for which there are at least $\Omega(k)$ edges of small effective resistance in every cut of G.

So, we take a detour. We use combinatorial structures that we call locally connected hierarchies that allow us to find an $\Omega(k)$-edge-connected set of good edges that may be very sparse with respect to G in some of the cuts. Let us give an informal definition of locally connected hierarchies. Consider a *laminar* structure on the vertices of G, say $S_1, S_2, \dots \subseteq V$, where by a laminar structure we mean that there is no $i \neq j$ such that $S_i \cap S_j, S_i \setminus S_j, S_j \setminus S_i \neq \emptyset$. Modulo some technical conditions, if for all i, the induced subgraph on S_i, $G[S_i]$, is k-edge-connected, then we call $S_1, S_2, \dots$ a locally connected hierarchy.

Let S_{i^*} be the smallest set that is a superset of S_i in the family, and let $\mathcal{O}(S_i) = E(S_i, S_{i^*} \setminus S_i)$ be the set of edges leaving S_i in the induced graph $G[S_{i^*}]$. In our main technical theorem we show that for any locally connected hierarchy we can find a shortcut matrix D that reduces the maximum of the average effective resistance of all $\mathcal{O}(S_i)$'s. In other words, the shortcut matrix D reduces the effective resistance of at least half of the edges of each $\mathcal{O}(S_i)$. Unfortunately, these small effective resistance edges may have $\Omega(n)$ connected components.

To prove theorem 3.6 we choose $\text{poly}\log\log(n)$ many locally connected hierarchies adaptively, such that the following holds: Let the laminar family $S_1^j, S_2^j, \dots$ be the j-th locally connected hierarchy, and D_j be a shortcut matrix that reduces the maximum average effective resistance of $\mathcal{O}(S_i^j)$'s. We let F_j be the set of small effective resistance edges in $\cup_i \mathcal{O}(S_i^j)$. We choose our locally connected hierarchies such that $F = \cup_j F_j$ is $\Omega(k)$-edge-connected in G. To ensure this we use several tools in graph partitioning.

3.5 Overview of Approach

In this section we give a high-level overview of our approach. We will motivate and formally define locally connected hierarchies and we describe our main technical theorem. In this section we will not overview the proof of the main technical theorem 4.2, see section 7.1 for the explanation.

As alluded to in the introduction, in theorem 7.3 we will show that it is not possible to reduce the maximum effective resistance of the edges of every k-edge-connected graph using a shortcut matrix.

The first idea that comes to mind is to reduce the maximum average effective resistance amongst all cuts of G. We can use the following convex program to find the best such shortcut matrix.

Average-CP:

$$
\begin{aligned}
\min \quad & \mathcal{E} \\
\text{s.t.} \quad & \mathop{\mathbb{E}}_{e \sim E(S,\overline{S})} \mathcal{R}\text{eff}_D(e) \leq \mathcal{E} \qquad \forall \emptyset \subsetneq S \subsetneq V, \\
& D \preceq_{\square} L_G, \\
& D \succ 0.
\end{aligned}
$$

Note that if the optimum is small, it means that there are at least $k/2$ good edges in every cut of G, so the set F of good edges is $\Omega(k)$-edge-connected and we are done. Unfortunately, as we will show in theorem 7.3 the same example shows that the optimum of the above convex program is very close to 1 for an $\Omega(\log(n))$-edge-connected graph. In fact, in the proof of theorem 7.3, we lower-bound the optimum of Average-CP.

The above impossibility result shows that it is not possible to reduce the average effective resistance of all cuts of G. Our approach is to recognize families of subsets of edges for which it is possible to reduce the maximum average effective resistance.

In the first step, we observe that for any partitioning of the vertices of a k-edge-connected graph G into $S_1, S_2, \ldots$ we can use a variant of the above convex program to reduce the maximum average effective resistance of the sets

$$E(S_1, \overline{S_1}), E(S_2, \overline{S_2}), \text{ and so on}$$

to $\tilde{O}(1/k)$. Next, we illustrate why this is useful using an example. Later, we will see that our main technical theorem implies a stronger version of this statement.

Example 3.10. *Assume that G is defined as follows: Start with a k-regular ϵ-expander on $\sqrt{n}$ vertices and replace each vertex with a cycle of length $\sqrt{n}$ repeated k times where the endpoints of the expander edges incident to each cycle are equidistantly distributed. This graph is k-edge-connected by definition and all expander edges have effective resistance close to 1.*

If we use the $\sqrt{n}$ cycles as our partition, by the above observation, we can reduce the average effective resistance of edges coming out of each cycle to some $\alpha = \tilde{O}(1/k)$. Let F be the union of all of the cycle edges and the expander edges of effective resistance at most $2\alpha/\epsilon$. Now, we show that F is $\Omega(k)$-edge-connected. For any cut that cuts at least one of the cycles, obviously there are at least k cycle edges in F. For the rest of the cuts, at least ϵ-fraction of the expander edges incident to the cycles on the small side of the cut cross the cut; among these edges at least half of them are in F, so F has at least $\Omega(k)$ edges in the cut.

We can use the above observation in any k-edge-connected graph repeatedly to gradually make F $\Omega(k)$-edge-connected as follows: Start with partitioning into singletons; let D_1 be a shortcut matrix that reduces the average effective resistance of degree cuts to $\alpha = \tilde{O}(1/k)$, and let F_1 be the edges of effective resistance at most 2α. In the next step, let the partitioning $S_1, S_2, \ldots$ be a natural decomposition of (V, F_1) into $k/2$-edge-connected components. Similarly, define D_2 and let F_2 be the edges connecting $S_1, S_2, \ldots$ of effective resistance at most 2α. This procedure ends in $\ell = O(\log n)$ iterations. It follows that $\cup_{i=1}^{\ell} F_i$ is $\Omega(k)$-edge-connected and the average of shortcut matrices, $\mathbb{E}_i D_i$, is a shortcut matrix that reduces the effective resistance of all edges of F to $O(\ell \cdot \alpha)$. Therefore, if $\ell = \text{poly} \log \log(n)$ we are done.

Unfortunately there are k-edge-connected graphs where the above procedure ends in $\Theta(\log n)$ steps because each time the size of the partition may reduce only by a factor of 2. Note that this procedure defines a laminar family over the vertices. Let $S_1, S_2, \ldots$ be all of

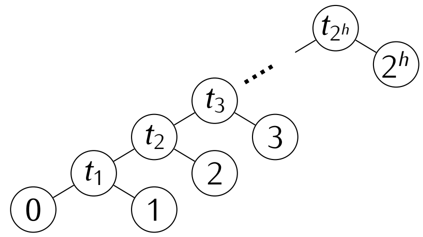

Figure 3.5: A $\mathcal{T}(k, 1/2, \{1, 2, \ldots, 2^h\})$-locally connected hierarchy of the graph of fig. 3.4.

the sets in all partitions; observe that they form a laminar family; let S_{i^*} be the smallest set that is a superset of S_i. Also, let $\mathcal{O}(S_i) = E(S_i, S_{i^*} \setminus S_i)$.

Suppose we write a convex program to *simultaneously* reduce the maximum average effective resistance of all $\mathcal{O}(S_i)$'s; then we may obtain a k-edge-connected set F of good edges in a single shot. As we will see next, modulo some technical conditions, this is what we prove in our main technical theorem. Such a statement is not enough to get a k-edge-connected set of good edges, but it is enough to get F in $\text{poly} \log \log(n)$ steps.

Locally Connected Hierarchies

For a graph $G = (V, E)$, a hierarchy, $\mathcal{T}$, is a tree where every non-leaf node has at least two children and each leaf corresponds to a unique vertex of G. We use the terminology *node* to refer to vertices of $\mathcal{T}$. For each node $t \in \mathcal{T}$ let $V(t) \subseteq V$ be the set of vertices of G that are mapped to the leaves of the subtree of t, $E(t)$ be the set of edges between the vertices of $V(t)$, and

$$G(t) = G[V(t), E(t)],$$

be the induced subgraph of G on $V(t)$. Let $\mathcal{P}(t) := E(V(t), \overline{V(t)})$ be the set of edges that leave $V(t)$ in G. Throughout the dissertation we use t^* to denote the parent of a node t. We define $\mathcal{O}(t) := E(V(t), V(t^*) \setminus V(t))$ as the set of edges that leave $V(t)$ in $G(t^*)$. We abuse notation and use $\mathcal{T}$ to also denote the set of nodes of $\mathcal{T}$.

Let us give a clarifying example. Say G is the "bad" graph of fig. 3.4. In fig. 3.5 we give a locally connected hierarchy of G. For each node t_i, $V(t_i) = \{0, 1, \ldots, i\}$. For each $1 \le i \le 2^h$, the set $\mathcal{O}(i)$ is the set of edges from vertex i to all vertices j with $j < i$. In addition, since t_i has exactly two children, $\mathcal{O}(i) = \mathcal{O}(t_{i-1})$. Finally, $\mathcal{P}(i)$ is all edges incident to vertex i and $\mathcal{P}(t_i)$ is the set of edges $E(\{0, 1, \ldots, i\}, \{i + 1, \ldots, 2^h\})$.

For an integer $k > 1$, $0 < \lambda < 1$, and $T \subseteq \mathcal{T}$, we say $\mathcal{T}$ is a (k, λ, T)-locally connected hierarchy of G, or (k, λ, T)-LCH if

1. For each node $t \in \mathcal{T}$, the induced graph $G(t)$ is k-edge-connected.

2. For any node $t \in \mathcal{T}$ that is not the root, $|\mathcal{O}(t)| \ge k$. This property follows from 1 because $\mathcal{O}(t) = E(V(t), V(t^*) \setminus V(t))$ is a cut of $G(t^*)$.

3. For any node $t \in T$, $|\mathcal{O}(t)| \geq \lambda \cdot |\mathcal{P}(t)|$. Note that unlike the other two properties, this one only holds for a subset T of the nodes of $\mathcal{T}$.

We say $\mathcal{T}$ is a (k, λ, T)-LCH if T is the set of all nodes of $\mathcal{T}$. For example, the hierarchy of fig. 3.5 is a $(k, 1/2, \{1, 2, \ldots, 2^h\})$-LCH of the graph illustrated in fig. 3.4. Condition 1 holds because there are k parallel edges between any pair of vertices $i-1, i$, so $G(V(t_i))$ is k-edge-connected. Condition 2 holds because,

$$|\mathcal{O}(i)| = |\mathcal{O}(t_{i-1})| = |E(\{0, \ldots, i-1\}, \{i\})| \geq k.$$

Lastly, it is easy to see that condition 3 holds for any leaf node $i \in T$, $|\mathcal{O}(i)| \geq d(i)/2 = |\mathcal{P}(i)|/2$.

We will use the following terminology mostly in section 7.3. For two nodes t, t' of an locally connected hierarchy, $\mathcal{T}$, we say t is an *ancestor* of t', if $t \neq t'$ and t' is a node of a subtree of t. We say t is a *weak ancestor* of t' if either $t = t'$ or t is an ancestor of t. We say t is a *descendant* of t' if t' is an ancestor of t. We say $t, t' \in \mathcal{T}$ are *ancestor-descendant* if either t is a weak ancestor of t' or t' is a weak ancestor of t.

Locally Connected Hierarchies and Good Edges. Let $\mathcal{T}$ be a hierarchy of G. Let $t \in \mathcal{T}$ have children $t_1, \ldots, t_j$. Define

$$G\{t\} := G(t)/V(t_1)/V(t_2)/\ldots/V(t_j)$$

to be the graph obtained from $G(t)$ by contracting each $V(t_i)$ into a single vertex. We may call $G\{t\}$ an internal subgraph of G. Let $V\{t\}$ be the vertex set of $G\{t\}$; we can also identify this set with the children of t in $\mathcal{T}$. Also, let $E\{t\}$ be the edge set of $V\{t\}$.

The following property of locally connected hierarchies is crucial in our proof. Roughly speaking, if a subset F of edges of G is k-edge-connected in each internal subgraph, then it is globally k-edge-connected.

Lemma 3.11. *Let $\mathcal{T}$ be a hierarchy of a graph $G = (V, E)$ and $F \subseteq E$. If for any internal node t, the subgraph $(V\{t\}, F \cap E\{t\})$ is k-edge-connected, then (V, F) is k-edge-connected.*

Proof. Consider any cut $(S, \overline{S})$ of G. Observe that there exists an internal node $t \in \mathcal{T}$ such that S crosses $V(t)$. Let t_0 be the deepest such node in $\mathcal{T}$ (root has depth 0). But then,

$$F(S, \overline{S}) \supseteq F(S \cap V(t_0), \overline{S} \cap V(t_0)),$$

and the size of the set on the RHS is at least k by the assumption of the lemma. □

To prove theorem 3.6 we will find a good set of edges which satisfy the assumption of the above lemma. Note that the assumption of the above lemma does not imply that F is dense in G. This is crucial because theorem 7.3 shows that there is no shortcut matrix D which has a dense set of good edges.

Construction of an LCH for Planar Graphs. In this section we give a universal construction of locally connected hierarchies for k-edge-connected planar graphs.

Lemma 3.12. *Any k-edge-connected planar graph $G = (V, E)$ has a $(k/5, 1/5, T)$-LCH $\mathcal{T}$ where $\mathcal{T}$ is a binary tree, and T contains at least one child of each nonleaf node of $\mathcal{T}$.*

We will use the following fact about planar graphs, whose proof easily follows from the fact that *simple* planar graphs have at least one vertex with degree at most 5.

Fact 3.13. *In any k-edge-connected planar graph $G = (V, E)$, there is a pair of vertices $u, v \in V$ with at least $k/5$ parallel edges between them.*

The details of the construction are given in algorithm 1. Observe that the algorithm

---

Algorithm 1 Construction of a locally connected hierarchy for planar graphs.

Input: A k-edge-connected planar graph G.
Output: A $(k/5, ., .)$-LCH of G.

For each vertex $v \in V$, add a unique leaf node to $\mathcal{T}$ and map v to it. Let W be the set of these leaf nodes. ▷ We keep the invariant that W is the nodes of $\mathcal{T}$ that do not have a parent yet, but their subtree is fixed, i.e., $V(t)$ is well-defined for any $t \in W$.
while $|W| > 1$ **do**
    Add a new node t^* to W.
    Let G_{t^*} be the graph where for each node $t \in W$, $V(t)$ is contracted to a single vertex; identify each $t \in W$ with the corresponding contracted vertex. ▷ Note that G_{t^*} is also a planar graph, because for any $t \in W$, the induced graph $G[V(t)]$ is connected.
    Let t_1 be a vertex with at most 5 neighbors in G_{t^*}. ▷ t_1 exists by Fact 3.13.
    Let t_2 be a neighbor of t_1 such that $\{t_1, t_2\}$ has the largest number of parallel edges among all neighbors of t_1. ▷ Note that t_1, t_2 are not necessarily vertices of G, so parallel edges between them do not correspond to parallel edges of G.
    Make t^* the parent of t_1, t_2; remove t_1, t_2 from W, and add t_1 to T. ▷ So, $V(t^*) = V(t_1) \cup V(t_2)$.
end while
return $\mathcal{T}$.

---

terminates after exactly $n - 1$ iterations of the loop, because any non-leaf node of $\mathcal{T}$ has exactly two children, so $|W|$ decreases by 1 in each iteration. We show that $\mathcal{T}$ is $\mathcal{T}(k/5, 1/5, T)$-LCH. First of all, for any non-leaf node t of $\mathcal{T}$, $G(t)$ is $k/5$-edge-connected. We prove this by induction. Say, t_1, t_2 are the two children of t^*, and by induction, $G(t_1)$ and $G(t_2)$ are $k/5$-edge-connected. By the selection of t_2, there are at least $k/5$ parallel edges between t_1, t_2, so $G(t^*)$ is $k/5$-edge-connected. Secondly, we need to show that $\mathcal{O}(t_1) \geq \mathcal{P}(t_1)/5$. This is because by the selection of t_2, 1/5 of the edges incident to t_1 in G_{t^*} are $\{t_1, t_2\}$. This completes the proof of lemma 3.12.

Main Technical Theorem

Given a (k, λ, T)-LCH $\mathcal{T}$ of G, in our main technical theorem we minimize the maximum average effective resistance of $\mathcal{O}(t)$'s among all nodes $t \in T$.

The following convex program finds a shortcut matrix $0 \prec D \preceq L_G$ that minimizes the maximum of the average effective resistance of edges in $\mathcal{O}(t)$ for all $t \in T$.

$$\begin{array}{ll} \multicolumn{2}{l}{\textbf{Tree-CP}(\mathcal{T} \in (k,\lambda,T)\text{-LCH}):} \\ \min & \mathcal{E} \\ \text{s.t.} & \mathbb{E}_{e \sim \mathcal{O}(t)} \mathcal{R}\text{eff}_D(e) \le \mathcal{E} \quad \forall t \in T, \\ & D \preceq_{\square} L_G, \\ & D \succ 0. \end{array}$$

Theorem 3.14 (Main Technical). *For any k-edge-connected graph G, and any $\mathcal{T}(k, \lambda, T)$-LCH, $\mathcal{T}$, of G, there is a PD shortcut matrix D such that for any $t \in T$,*

$$\mathbb{E}_{e \sim \mathcal{O}(t)} \mathcal{R}\text{eff}_D(e) \le \frac{f_1(k, \lambda)}{k},$$

where $f_1(k, \lambda)$ is a poly-logarithmic function of $k, 1/\lambda$.

Note that the statement of the above theorem does not have any dependency on the size of G.

If we apply the above theorem to the $(k/5, 1/5, T)$-LCH $\mathcal{T}$ of a k-edge-connected planar graph as constructed in algorithm 1, we obtain a shortcut matrix D for which the small effective resistance edges are $\Omega(k)$-edge-connected. Let us elaborate on this. Let $F = \{e : \mathcal{R}\text{eff}_D(e) \le \frac{2f_1(k/5,1/5)}{k/5}\}$. First, note that by lemma 3.12, $\mathcal{T}$ is a binary tree and at least one child of each internal node of $\mathcal{T}$ is in T. Say t is an internal node with children t_1, t_2 and $t_1 \in T$. Then, by Markov's inequality

$$|F \cap \mathcal{O}(t_1)| \ge |\mathcal{O}(t_1)|/2 \ge \frac{k/5}{2}.$$

Since t has only two children, this implies $G(V\{t\}, F \cap E\{t\})$ is $k/10$-edge-connected. Now, by lemma 3.11, (V, F) is $k/10$-edge-connected.

It is natural to expect that for every k-edge-connected graph G, one can find a locally connected hierarchy $\mathcal{T}$ such that one application of the above theorem produces a set F of good edges such that for any $t \in \mathcal{T}$, $G(V\{t\}, F \cap E\{t\})$ is $\Omega(k)$-edge-connected. By lemma 3.11 this would imply (V, F) is $\Omega(k)$-edge-connected. However, the following example shows that this may not be the case.

Example 3.15. *Let $G = (V, E)$ be the k-dimensional hypercube ($n = 2^k$). Note that G is k-edge-connected. Let $\mathcal{T}$ be a $(\Omega(k), ., .)$-LCH for G. Consider an internal node $t_0 \in \mathcal{T}$, all of whose children are leaves. By definition $G(t_0)$ is $\Omega(k)$-edge-connected. Consider a dimension* *cut of the hypercube that cuts $G(t_0)$ into $(S, V(t_0) \setminus S)$. Imagine a solution D of* Tree-CP($\mathcal{T}$) *which reduces the effective resistance of all edges except those in the cut $(S, V(t_0) \setminus S)$. In such a solution, $\mathbb{E}_{e \sim \mathcal{O}(t)} \mathcal{R}\text{eff}_D(e)$ is small for all t. This is because each vertex $v \in G(t)$ has at most one of its $\Omega(k)$ neighboring edges in the cut $(S, V(t_0) \setminus S)$. But note that the small effective resistance edges are disconnected in $G\{t_0\} = G(t_0)$.*

Consider a $(\Omega(k), ., .)$-LCH $\mathcal{T}$ of G and let t be an internal node. Theorem 4.2 promises that the average effective resistance of all degree cuts of the internal graph $G\{t\}$ are small. If $G\{t\}$ is an *expander* this implies that the good edges are $\Omega(k)$-edge-connected in $G\{t\}$. Therefore, if we can find a locally connected hierarchy whose internal subgraphs are expanding we can find an $\Omega(k)$-edge-connected set of good edges by a single application of theorem 4.2. This is exactly what we proved in the case of planar graphs. The above hypercube example shows that such a locally connected hierarchy does not necessarily exist in all k-edge-connected graphs.

Expanding Locally Connected Hierarchies

In this section we define expanding locally connected hierarchies and we describe our plan to prove theorem 3.6 using the main technical theorem.

Definition 3.16 (Expanding Locally Connected Hierarchies). *For a node t with children $t_1, \dots, t_j$ in a locally connected hierarchy $\mathcal{T}$ of a graph $G = (V, E)$, an internal node t (or the internal subgraph $G\{t\}$) is called (α, β)-expanding, if $G\{t\}$ is an α-expander and is β-edge-connected. A subset of the nodes T is called (α, β)-expanding iff each one of them is (α, β)-expanding and similarly the locally connected hierarchy, $\mathcal{T}$, is (α, β)-expanding iff all of its nodes are (α, β)-expanding.*

Recall that locally connected hierarchies already guarantee k-edge-connectivity of the internal subgraphs for some k. So, we always have $\beta \geq k$. If $\beta = k$, we omit it from the notation and write $(\alpha, .)$-expanding; otherwise, the (α, β)-expanding property guarantees slightly stronger connectivity for a *subset* of the internal subgraphs.

For example, observe that the locally connected hierarchies that we constructed in algorithm 1 for k-edge-connected planar graphs are $(1, k/5)$-expanding. In theorem 6.1 we construct an $(\Omega(1/k), \Omega(k))$-expanding $(\Omega(k), \Omega(1), \mathcal{T})$-LCH for any k-edge-connected graph where $k \geq 7 \log n$. But example 3.15 shows that this is essentially the best possible, as the k-dimensional hypercube does not have any $(\omega(1/k), \Omega(k))$-expanding locally connected hierarchy.

It follows that if G has an $(\alpha, \Omega(k))$-expanding locally connected hierarchy then there is a shortcut matrix D and an $\Omega(k)$-edge-connected set F of edges such that

$$\max_{e \in F} \mathcal{R}\text{eff}_D(e) \leq O(\text{Tree-CP}(\mathcal{T})/\alpha).$$

Recall the argument in example 3.10 for details. Since the best α we can hope for is $O(1/\log n)$ this argument by itself does not work.

Our approach is to apply theorem 4.2 to an adaptively chosen sequence of locally connected hierarchies. Each time we recognize the internal subgraphs of the locally connected hierarchy in which the set of good edges found so far are not $\Omega(k)$-edge-connected. Then, we apply theorem 4.2 to the nodes in these internal subgraphs. We "refine" these internal subgraphs by a natural decomposition of the newly found good edges to get the next locally connected hierarchy. At the heart of the argument we show that this refinement procedure improves the expansion of the aforementioned internal subgraphs by a constant factor. Therefore, this procedure stops after $O(\log(1/\alpha)) = \text{poly}\log\log(n)$ steps in the worst case.

We conclude this section by describing an instantiation of the above procedure in the special case of a k-dimensional hypercube for demonstration purposes. Let G be a k-dimensional hypercube. We let $\mathcal{T}_1$ be a star, i.e., it has only one internal node and the vertices of G are the leaves. This means that in Tree-CP($\mathcal{T}_1$) we minimize the maximum average effective resistance of degree cuts of G. Let F_1 be the edges of effective resistance at most twice the optimum of Tree-CP($\mathcal{T}_1$). It follows that half the edges incident to each vertex are in F_1. Now, we find a natural decomposition of the good edges F_1. In the "worst case", edges of F_1 form $k/2$ dimensional sub-hypercubes and all edges connecting these sub-hypercubes are not in F_1. Note that if we contract these sub-hypercubes, we get a $k/2$-dimensional hypercube which is a $2/k$-expander, twice more expanding than G. Of course, we cannot contract, because we need good edges having small effective resistance with respect to the original vertex set, but the expansion is our measure of progress.

In the next iteration, we construct a $(.,., T_2)$-LCH $\mathcal{T}_2$ where the vertices of each $k/2$-dimensional sub-hypercube are connected to a unique internal node, and the root is connecting all internal nodes, i.e., $\mathcal{T}_2$ has height 2. We let T_2 be the set of all internal nodes (except the root). Note that if we delete the leaves, then $\mathcal{T}_2$ would be the same as $\mathcal{T}_1$ for a $k/2$-dimensional sub-hypercube. Similarly, we solve Tree-CP($\mathcal{T}_2$), and in the worst case the new good edges form $k/4$ dimensional sub-hypercubes. Continuing this procedure after $\log(k) = \log\log n$ iterations the good edges span an $\Omega(k)$-edge-connected subset of G.

Missing Proofs from theorem 3.2

Theorem 3.2. *For any $\alpha > 0$ (which can be a function of n), and $k \geq \log n$, a polynomial-time construction of an α/k-thin tree in any k-edge-connected graph gives an $O(\alpha)$-approximation algorithm for ATSP. In addition, even an existential proof gives an $O(\alpha)$ upper bound on the integrality gap of the LP relaxation.*

Proof. For a feasible vector x of LP (3.1), let $c(x) = \sum_{u,v} c(u,v) \cdot x_{u,v}$. For two disjoint sets A, B and a set of arcs T let

$$\vec{T}(A,B) := \{(u,v) : u \in A, v \in B\},$$

be the set of arcs from A to B. We use the following theorem that is proved in [Asa+10].

Theorem 3.17. *For a feasible solution x of LP* (3.1) *and a spanning tree T such that for any $S \subseteq V$,*

$$|\vec{T}(S,\overline{S})| - |\vec{T}(\overline{S},S)| \leq \alpha \cdot \sum_{u \notin S, v \in S} x_{u,v} + x_{v,u} =: \alpha \cdot x(S,\overline{S}), \tag{3.3}$$

and $\sum_{(u,v)\in T} c(u,v) \leq \beta \cdot c(x)$, there is a polynomial time algorithm that finds a tour of length $O(\alpha+\beta)\cdot c(x)$.

Given a feasible solution x of LP (3.1), for a constant $C \geq 4$, we sample $Ck \cdot n$ arcs where the probability of choosing each arc (u,v) is proportional to $x_{u,v}$. We drop the direction of the arcs and we call the sampled graph $G=(V,E)$. Since $x(S,\overline{S}) \geq 2$ for all $S \subseteq V$, and $k \geq \log n$, it follows by the seminal work of Karger [Kar99] that for a sufficiently large C, with high probability, for any $S \subseteq V$, $|E(S,\overline{S})|$ is between 1/2 and 2 times $Ck \cdot x(S,\overline{S})$. Since this happens with high probability, by Markov's inequality we can also assume that

$$c(E) \leq 2C \cdot k \cdot c(x),$$

where for a set $F \subseteq E$,

$$c(F) := \sum_{\{u,v\}\in F} \min\{c(u,v), c(v,u)\}.$$

Since $x(S,\overline{S}) \geq 2$ and $C \geq 4$, G is $2k$-edge-connected. Let $\beta = \alpha/k$. By the assumption of the theorem, G has a β-thin tree, say T_1. Because of the thinness of T_1, $G(V, E \setminus T_1)$ is $2k(1-\beta) \geq k$-edge-connected. Therefore, it also has a β-thin tree. By repeating this argument, we can find $j = \frac{1}{2\beta}$ edge-disjoint β-thin spanning trees in G, $T_1, \ldots, T_j$.

Without loss of generality, assume that $c(T_1) = \min_{1\leq i\leq j} c(T_i)$. We show that T_1 satisfies the conditions of the above theorem. First, since $c(T_1) = \min_{1\leq i\leq j} c(T_i)$,

$$c(T_1) \leq \frac{c(E)}{j} \leq \frac{2C \cdot k \cdot c(x)}{j} = 4C \cdot \alpha \cdot c(x).$$

On the other hand, since T_1 is β-thin with respect to G, for any set $S \subseteq V$,

$$|T_1(S,\overline{S})| \leq \beta \cdot |E(S,\overline{S})| \leq 2C \cdot k \cdot \beta \cdot x(S,\overline{S}) = 2C \cdot \alpha \cdot x(S,\overline{S}).$$

Therefore, the theorem follows from an application of Theorem 3.17. □

Chapter 4

Kadison-Singer for Strongly Rayleigh Measures

Marcus, Spielman and Srivastava [MSS13b] in a breakthrough work proved the Kadison-Singer conjecture [KS59]; they proved Weaver's conjecture KS_2 which was known to imply the Kadison-Singer conjecture [Wea04].

The following is their main technical contribution.

Theorem 4.1 ([MSS13b]). *If $\epsilon > 0$ and $v_1, \ldots, v_m$ are independent random vectors in $\mathbb{R}^d$ with finite support where,*

$$\sum_{i=1}^{m} \mathbb{E} v_i v_i^{\intercal} = I,$$

such that for all i,

$$\mathbb{E} \left\| v_i \right\|^2 \leq \epsilon,$$

then

$$\mathbb{P}\left[\left\| \sum_{i=1}^{m} v_i v_i^{\intercal} \right\| \leq (1+\sqrt{\epsilon})^2 \right] > 0.$$

In this chapter, we prove an extension of the above theorem to families of vectors assigned to elements of a not necessarily independent distribution.

4.1 Strongly Rayleigh Measures

Let $\mu : 2^{[m]} \to \mathbb{R}_{>0}$ be a probability distribution on the subsets of the set $[m] = \{1, 2, \ldots, m\}$. In particular, we assume that $\mu(.)$ is nonnegative and,

$$\sum_{S \subseteq [m]} \mu(S) = 1.$$

We assign a multi-affine polynomial with variables $z_1, \ldots, z_m$ to μ,

$$g_\mu(z) = \sum_{S \subseteq [m]} \mu(S) \cdot z^S,$$

where for a set $S \subseteq [m]$, $z^S = \prod_{i \in S} z_i$. The polynomial g_μ is also known as the *generating polynomial* of μ. We say μ is a *homogeneous* probability distribution if g_μ is a homogeneous polynomial.

We say that μ is a *strongly Rayleigh* distribution if g_μ is a real stable polynomial. See section 2.5 for the definition of real stability. Strongly Rayleigh measures are introduced and deeply studied in the seminal work of Borcea, Brändén and Liggett [BBL09]. They are natural generalizations of product distributions and cover several interesting families of probability distributions including determinantal measures and random spanning tree distributions. We refer interested readers to [OSS11; PP14] for applications of these probability measures.

The main theorem of this chapter extends theorem 4.1 to families of vectors assigned to the elements of a strongly Rayleigh distribution. This can be seen as a generalization because independent distributions are special classes of strongly Rayleigh measures. To state the main theorem we need another definition. The *marginal* probability of an element i with respect to a probability distribution μ, is the probability that i is in a sample of μ,

$$\mathbb{P}_{S \sim \mu}[i \in S] = \partial_{z_i} g_\mu(z)\big|_{z_1 = \ldots = z_m = 1}. \tag{4.1}$$

4.2 The Extension of Kadison-Singer

We extend Kadison-Singer to work for strongly Rayleigh measures. More formally:

Theorem 4.2. *Let μ be a homogeneous strongly Rayleigh probability distribution on $[m]$ such that the marginal probability of each element is at most ϵ_1, and let $v_1, \ldots, v_m \in \mathbb{R}^d$ be vectors in isotropic position, i.e.,*

$$\sum_{i=1}^m v_i v_i^\intercal = I,$$

such that for all i, $\left\|v_i\right\|^2 \le \epsilon_2$. Then,

$$\mathbb{P}_{S \sim \mu}\left[\left\|\sum_{i \in S} v_i v_i^\intercal\right\| \le 4(\epsilon_1 + \epsilon_2) + 2(\epsilon_1 + \epsilon_2)^2\right] > 0.$$

The above theorem does not directly generalize theorem 4.1, but it can be seen as a variant of theorem 4.1 that works in the case where the vectors $v_1, \ldots, v_m$ are chosen according to a distribution with negative dependence. We expect to see several applications of our main theorem that are not realizable by the original proof of [MSS13b].

Let us conclude this part by proving a simple application of the above theorem to prove KS_r for $r \ge 5$.

Corollary 4.3. *Given a set of vectors $v_1, \dots, v_m \in \mathbb{R}^d$ in isotropic position,*

$$\sum_{i=1}^{m} v_i v_i^{\intercal} = I,$$

if for all i, $\|v_i\|^2 \le \epsilon$, then for any r, there is an r-partitioning of $[m]$ into $S_1, \dots, S_r$ such that for any $j \le r$,

$$\left\| \sum_{i \in S_j} v_i v_i^{\intercal} \right\| \le 4(1/r + \epsilon) + 2(1/r + \epsilon)^2.$$

Proof. The proof is inspired by the lifting idea in [MSS13b]. For $i \in [m]$ and $j \in [r]$ let $w_{i,j} \in \mathbb{R}^{d \cdot r}$ be the directed sum of r vectors all of which are 0^d except the j-th one which is v_i, i.e.,

$$w_{i,1} = \begin{pmatrix} v_i \\ 0^d \\ \vdots \\ 0^d \end{pmatrix}, w_{i,2} = \begin{pmatrix} 0^d \\ v_i \\ \vdots \\ 0^d \end{pmatrix}, \text{ and so on.}$$

Let $E = \{(i,j) : i \in [m], j \in [r]\}$ and let $\mu : 2^E \to \mathbb{R}_+$ be a product distribution defined in a way that selects exactly one pair $(i,j) \in E$ for any $i \in [m]$ uniformly at random. Observe that there are m^r sets in the support of μ each of size exactly m and each has probability $1/r^m$. Therefore, μ is a homogeneous probability distribution and the marginal probability of each element of E is exactly $1/r$. In addition, since product distributions are strongly Rayleigh, μ is strongly Rayleigh. Therefore, by theorem 4.2, there is a set S in the support of μ such that

$$\left\| \sum_{(i,j) \in S} w_{i,j} w_{i,j}^{\intercal} \right\| \le \alpha,$$

for $\alpha = 4(1/r + \epsilon) + 2(1/r + \epsilon)^2$. Now, let $S_j = \{i : (i,j) \in S\}$. It follows that for any $j \in [r]$,

$$\left\| \sum_{i \in S_j} v_i v_i^{\intercal} \right\| \le \alpha,$$

as desired. □

4.3 The Thin Basis Problem

In this section we use the main theorem to prove the existence of a thin basis among a given set of isotropic vectors. In the next section, we will use this theorem to prove the existence of thin trees in graphs, i.e., trees which are “sparse” in all cuts of a given graph.

Given a set of vectors $v_1, \dots, v_m \in \mathbb{R}^d$ in the isotropic position,

$$\sum_{i=1}^{m} v_i v_i^{\intercal} = I,$$

we want to find a sufficient condition for the existence of a *thin basis*. Recall that a set $T \subset [m]$ is a basis if $|T| = d$ and all vectors indexed by T are linearly independent. We say T is α-thin if

$$\left\| \sum_{i \in T} v_i v_i^{\intercal} \right\| \leq \alpha.$$

An obvious necessary condition for the existence of an α-thin basis is that the set

$$V(\alpha) := \{v_i : \left\|v_i\right\|^2 \leq \alpha\},$$

contains a basis. We show that there exist universal constants $C_1, C_2 > 0$ such that the existence of C_1/α disjoint bases in $V(C_2 \cdot \alpha)$ is a sufficient condition.

Theorem 4.4. *Given a set of vectors $v_1, \dots, v_m \in \mathbb{R}^d$ in the sub-isotropic position*

$$\sum_{i=1}^{m} v_i v_i^{\intercal} \preceq I,$$

if for all $1 \leq i \leq m$, $\left\|v_i\right\|^2 \leq \epsilon$, and the set $\{v_1, \dots, v_m\}$ contains k disjoint bases, then there exists an $O(\epsilon + 1/k)$-thin basis $T \subseteq [m]$.

We will use theorem 4.2 to prove the above theorem. To use theorem 4.2 we need to define a strongly Rayleigh distribution on $[m]$ with small marginal probabilities. This is proved in the following proposition.

Proposition 4.5. *Given a set of vectors $v_1, \dots, v_m \in \mathbb{R}^d$ that contains k disjoint bases, there is a strongly Rayleigh probability distribution $\mu : 2^{[m]} \to \mathbb{R}_+$ supported on the bases such that the marginal probability of each element is at most $O(1/k)$.*

Now, theorem 4.4 follows simply from the above proposition. Letting μ be defined as above, we get $\epsilon_1 = \epsilon$ and $\epsilon_2 = O(1/k)$ in theorem 4.2 which implies the existence of a basis $T \subseteq [m]$ such that

$$\left\| \sum_{i \in T} v_i v_i^{\intercal} \right\| \leq O(\epsilon + 1/k),$$

as desired.

In the rest of this section we prove the above proposition. In our proof μ will in fact be a homogeneous *determinantal* probability distribution. We say $\mu : 2^{[m]} \to \mathbb{R}_+$ is a determinantal probability distribution if there is a PSD matrix $M \in \mathbb{R}^{m \times m}$ such that for any set $T \subseteq [m]$,

$$\mathbb{P}_{S \sim \mu}[T \subseteq S] = \det(M_{T,T}),$$

where $M_{T,T}$ is the principal submatrix of M whose rows and columns are indexed by T. It is proved in [BBL09] that any determinantal probability distribution is a strongly Rayleigh measure, so this is sufficient for our purpose. In fact, we will find nonnegative weights $\lambda : [m] \to \mathbb{R}_+$ and for any basis T we will let

$$\mu_\lambda(T) \propto \det\left(\sum_{i \in T} \lambda_i v_i v_i^\intercal\right). \tag{4.2}$$

It follows by the Cauchy-Binet identity that for any λ, such a distribution is determinantal with respect to the gram matrix

$$M(i,j) = \sqrt{\lambda_i \lambda_j}\left\langle B^{-1/2} v_i, B^{-1/2} v_j\right\rangle$$

where $B = \sum_{i=1}^m \lambda_i v_i v_i^\intercal$. So, all we need to do is find $\{\lambda_i\}_{1 \le i \le m}$ such that the marginal probability of each element in μ_λ is $O(1/k)$.

For any basis $T \subset [m]$ let $\mathbf{1}_T \in \mathbb{R}^m$ be the indicator vector of the set T. Let P be the convex hull of bases' indicator vectors,

$$P := \operatorname{conv}\{\mathbf{1}_T : T \text{ is a basis}\}.$$

Recall that a point x is in the *relative interior* of P, $x \in \operatorname{relint} P$, if and only if x can be written as a convex combination of all of the vertices of P with strictly positive coefficients.

We find the weights in two steps. First, we show that for any point $x \in \operatorname{relint} P$, there exist weights $\lambda : [m] \to \mathbb{R}$ such that for any i,

$$\mathbb{P}_{S \sim \mu_\lambda}[i \in S] = x(i),$$

where $x(i)$ is the i-th coordinate of x and μ_λ is defined as in (4.2). Then, we show that there exists a point $x \in \operatorname{relint} P$ such that for all i, $x(i) \le O(1/k)$.

Lemma 4.6. *For any $x \in \operatorname{relint} P$ there exist $\lambda : [m] \to \mathbb{R}_+$ such that the marginal probability of each element i in μ_λ is $x(i)$.*

Proof. Let $\mu^* := \mu_1$ be the (determinantal) distribution where $\lambda_i = 1$ for all i. The idea is to find a distribution $p(.)$ maximizing the relative entropy with respect to μ^* and preserves x as the marginal probabilities. This is analogous to the recent applications of maximum entropy distributions in approximation algorithms [Asa+10; SV14].

Consider the following entropy maximization convex program.

$$\begin{aligned} \min \quad & \sum_T p(T) \cdot \log \frac{p(T)}{\mu^*(T)} \\ \text{s.t.} \quad & \sum_{T : i \in T} p(T) = x(i) \qquad \forall i, \\ & p(T) \ge 0. \end{aligned} \tag{4.3}$$

Note that any feasible solution satisfies $\sum_T p(T) = 1$ so we do not need to add this as a constraint. First of all, since $x \in \operatorname{relint} P$, there exists a distribution $p(.)$ such that for all bases T, $p(T) > 0$. So, the Slater condition holds and the duality gap of the above program is zero.

Secondly, we use the Lagrange duality to characterize the optimum solution of the above convex program. For any element i let γ_i be the Lagrange dual variable of the first constraint. The Lagrangian $L(p, \gamma)$ is defined as follows:

$$L(p, \gamma) = \inf_{p \geq 0} \sum_T p(T) \cdot \log \frac{p(T)}{\mu^*(T)} - \sum_i \gamma_i \sum_{T : e \in T} (p(T) - x(i))$$

Let p^* be the optimum p, letting the gradient of the RHS equal to zero we obtain, for any bases T,

$$\log \frac{p^*(T)}{\mu^*(T)} + 1 = \sum_{i \in T} \gamma_i.$$

For all i, let $\lambda_i = \exp(\gamma_i - 1/d)$, where d is the dimension of the v_i's. Then, we get

$$\begin{aligned} p^*(T) &= \prod_{i \in T} \lambda_i \cdot \mu^*(T) \\ &= \prod_{i \in T} \lambda_i \cdot \det\left(\sum_{i \in T} v_i v_i^{\intercal} \right) \\ &= \det\left(\sum_{i \in T} \lambda_i v_i v_i^{\intercal} \right). \end{aligned}$$

Therefore $p^* \equiv \mu_\lambda$. Since the duality gap is zero, the above p^* is indeed an optimal solution of the convex program (4.3). Therefore, the marginal probability of every element i with respect to p^* (μ_λ) is equal to $x(i)$. □

Note that this lemma does not use any property of the vectors and can be readily generalized to arbitrary strongly Rayleigh distributions. However the following lemma does not generalize as easily.

Lemma 4.7. *If $\{v_1, \dots, v_m\}$ contains k disjoint bases, then there exists a point $x \in \operatorname{relint} P$, such that $x(i) = O(1/k)$ for all i.*

Proof. Let $T_1, \dots, T_k$ be the promised disjoint bases. Let

$$x_0 = \frac{\mathbf{1}_{T_1} + \cdots + \mathbf{1}_{T_k}}{k}.$$

The above is a convex combination of the vertices of P; so $x_0 \in P$. We now perturb x_0 by a small amount to find a point in $\operatorname{relint} P$. Let x_1 be an arbitrary point in $\operatorname{relint} P$ (such as the average of all vertices). For any $0 < \epsilon < 1$, the point $x = (1 - \epsilon)x_0 + \epsilon x_1 \in \operatorname{relint} P$. If ϵ is small enough, we get $x(i) = O(1/k)$ which proves the claim. □

This completes the proof of proposition 4.5.

4.4 Implications for Spectrally Thin Trees

Remember that for a graph $G = (V, E)$, the Laplacian of G, L_G, is defined as follows: For a vertex $i \in V$ let $\mathbf{1}_i \in \mathbb{R}^V$ be the vector that is one at i and zero everywhere else. Fix an arbitrary orientation on the edges of E and let $b_e = \mathbf{1}_i - \mathbf{1}_j$ for an edge e oriented from i to j. Then,

$$L_G = \sum_{e \in E} b_e b_e^\intercal.$$

We use $L_G^\dagger$ to denote the pseudo-inverse of L_G. Also, for a set $T \subseteq E$, we write

$$L_T = \sum_{e \in T} b_e b_e^\intercal.$$

Remember that a spanning tree T is α-spectrally thin with respect to G if

$$L_T \preceq \alpha \cdot L_G.$$

Remember that spectral thinness is more strict than combinatorial thinness, i.e., if T is α-spectrally thin it is also α-thin.

It turns out that the existence of spectrally thin trees is significantly easier to prove than combinatorially thin trees thanks to theorem 4.1 of [MSS13b]. Given a graph $G = (V, E)$, Harvey and Olver [HO14] employ a recursive application of [MSS13b] and show that if for all edges $e \in E$, $b_e^\intercal L_G^\dagger b_e \leq \alpha$, then G has an $O(\alpha)$-spectrally thin tree. The quantity $b_e L_G^\dagger b_e$ is the effective resistance between the endpoints of e when we replace every edge of G with a resistor of resistance 1 [LP13, Ch. 2]. Unfortunately, k-edge-connectivity is a significantly weaker property than $\max_e b_e L_G^\dagger b_e \leq \alpha$. So, this does not resolve the thin tree problem.

The main idea of this dissertation is to slightly change the graph G in order to decrease the effective resistance of edges while maintaining the size of the cuts intact. More specifically, to add a "few" edges E' to G such that in the new graph $G' = (V, E \cup E')$, the effective resistance of every edge of E is small and the size of every cut of G' is at most twice of that cut in G. If we can prove that G' has a spectrally thin tree $T \subseteq E$ such a tree is combinatorially thin with respect to G because G, G' have the same cut structure. To show that G' has a spectrally thin tree we need to answer the following question.

Problem 4.8. *Given a graph $G = (V, E)$, suppose there is a set $F \subseteq E$ such that (V, F) is k-edge-connected, and that for all $e \in F$, $b_e^\intercal L_G^\dagger b_e \leq \alpha$. Can we say that G has a $C \cdot \max\{\alpha, 1/k\}$-spectrally thin tree for a universal constant C?*

We use theorem 4.4 to answer the above question affirmatively.

Note that the above question cannot be answered by theorem 4.1. One can use theorem 4.1 to show that the set F can be partitioned into two sets F_1, F_2 such that each F_i is $1/2 + O(\alpha)$-spectrally thin, but theorem 4.1 gives no guarantee on the connectivity of F_i's. On the other hand, once we apply our main theorem to a strongly Rayleigh distribution supported on connected subgraphs of G, e.g. the spanning trees of G, we get connectivity for free.

Corollary 4.9. *Given a graph $G = (V, E)$ and a set $F \subseteq E$ such that (V, F) is k-edge-connected, if for $\epsilon > 0$ and any edge $e \in F$, $b_e^\intercal L_G^\dagger b_e \leq \epsilon$, then G has an $O(1/k + \epsilon)$ spectrally thin tree.*

Proof. Let $L_G^{\dagger/2}$ be the square root of $L_G^\dagger$. Note that since $L_G^\dagger \succeq 0$, its square root is well defined. For all $e \in F$, let

$$v_e = L_G^{\dagger/2} b_e.$$

Then, by the corollary's assumption, for each $e \in F$,

$$\left\|v_e\right\|^2 = b_e L_G^\dagger b_e \leq \epsilon,$$

and the vectors $\{v_e\}_{e\in F}$ are in sub-isotropic position,

$$\begin{aligned} \sum_{e\in F} v_e v_e^\intercal &= L_G^{\dagger/2} \left(\sum_{e\in F} b_e b_e^\intercal \right) L_G^{\dagger/2} \\ &= L_G^{\dagger/2} L_F L_G^{\dagger/2} \preceq I. \end{aligned}$$

In addition, we can show that $\{v_e\}_{e\in F}$ contains $k/2$ disjoint bases. First of all, note that each basis of the vectors $\{v_e\}_{e\in F}$ corresponds to a spanning tree of the graph (V, F). Nash-Williams [Nas61] proved that any k-edge-connected graph has $k/2$ edge-disjoint spanning trees. Since (V, F) is k-edge-connected, it has $k/2$ edge-disjoint spanning trees, and equivalently, $\{v_e\}_{e\in F}$ contains $k/2$ disjoint bases.

Therefore, by theorem 4.4, there exists a basis (i.e., a spanning tree) $T \subseteq F$ such that

$$\left\| \sum_{e\in T} v_e v_e^\intercal \right\| \leq \alpha, \tag{4.4}$$

for $\alpha = O(\epsilon + 1/k)$. Fix an arbitrary vector $y \in \mathbb{R}^V$. We show that

$$y^\intercal L_T y \leq \alpha \cdot y^\intercal L_G y, \tag{4.5}$$

and this completes the proof. By (4.4) for any $x \in \mathbb{R}^V$,

$$x^\intercal \left(\sum_{e\in T} v_e v_e^\intercal \right) x \leq \alpha \cdot ||x||^2.$$

Let $x = L_G^{1/2} y$, we get

$$y^\intercal L_G^{1/2} \left(L_G^{\dagger/2} \sum_{e\in T} b_e b_e^\intercal L_G^{\dagger/2} \right) L_G^{1/2} y \leq \alpha \cdot y^\intercal L_G y.$$

The above is the same as (4.5) and we are done. $\square$

The above corollary completely answers Problem 4.8 but it is not enough for our purpose in this dissertation; we need a slightly stronger statement. For a matrix $D \in \mathbb{R}^{V\times V}$ we say $D \preceq_{\square} L_G$, if for any set $S \subset V$,

$$\mathbf{1}_S^\intercal D \mathbf{1}_S \leq \mathbf{1}_S^\intercal L_G \mathbf{1}_S,$$

where as usual $\mathbf{1}_S \in \mathbb{R}^V$ is the indicator vector of the set S. In the main technical theorem of this dissertation we show that for any k-edge-connected graph G with $k = \Omega(\log n)$, there is a positive definite (PD) matrix $D \preceq_{\square} L_G$ and a set $F \subseteq E$ such that (V, F) is $\Omega(k)$-edge-connected and

$$\max_{e\in F} b_e^\intercal D^{-1} b_e \leq \frac{\operatorname{poly}\log(k)}{k}.$$

To show that G has a combinatorially thin tree it is enough to show that there is a tree $T \subseteq E$ that is α-spectrally thin w.r.t. $L_G + D$ for $\alpha = \operatorname{poly}\log(k)/k$, i.e.,

$$L_T \preceq \frac{\operatorname{poly}\log(k)}{k}(L_G + D).$$

Such a tree is 2α-combinatorially thin w.r.t. G because $D \preceq_{\square} L_G$. Note that the above corollary does not imply $L_G + D$ has a spectrally thin tree because D is not necessarily a Laplacian matrix. Nonetheless, we can prove the existence of a spectrally thin tree with another application of theorem 4.4.

Corollary 4.10. *Given a graph $G = (V, E)$, a matrix $D \succ 0$, and $F \subseteq E$ such that (V, F) is k-edge-connected, if for any edge $e \in F$,*

$$b_e^\intercal D^{-1} b_e \leq \epsilon,$$

then G has a spanning tree $T \subseteq F$ such that

$$L_T \preceq O(\epsilon + 1/k) \cdot (L_G + D).$$

Proof. The proof is very similar to corollary 4.17. For any edge $e \in F$, let $v_e = (D + L_G)^{-1/2} b_e$. Note that since D is PD, $D + L_G$ is PD and $(D + L_G)^{-1/2}$ is well defined. By the assumption,

$$\left\|v_e\right\|^2 = b_e^\intercal (D + L_G)^{-1} b_e \leq b_e^\intercal D^{-1} b_e = \epsilon,$$

where the inequality uses lemma 2.2. In addition, the vectors are in sub-isotropic position,

$$\sum_{e\in F} v_e v_e^\intercal = (D + L_G)^{\dagger/2} L_F (D + L_G)^{\dagger/2} \preceq I.$$

The matrix PSD inequality uses that $L_F \preceq L_G \preceq D + L_G$. Furthermore, every basis of $\{v_e\}_{e\in E}$ is a spanning tree of G and by $\Omega(k)$-connectivity of F, there are $\Omega(k)$-edge disjoint bases. Therefore, by theorem 4.4, there is a tree $T \subseteq F$ such that

$$\left\|\sum_{e\in T} v_e v_e^\intercal\right\| \leq \alpha,$$

for $\alpha = O(\epsilon + 1/k)$. Similar to corollary 4.17 this tree satisfies

$$L_T \preceq \alpha \cdot (L_G + D),$$

and this completes the proof. □

4.5 Proof Overview

We build on the method of interlacing polynomials of [MSS13a; MSS13b]. Recall that an interlacing family of polynomials has the property that there is always a polynomial whose largest root is at most the largest root of the sum of the polynomials in the family. First, we show that for any set of vectors assigned to the elements of a homogeneous strongly Rayleigh measure, the characteristic polynomials of natural quadratic forms associated with the samples of the distribution form an interlacing family. This implies that there is a sample of the distribution such that the largest root of its characteristic polynomial is at most the largest root of the average of the characteristic polynomials of all samples of μ. Then, we use the multivariate barrier argument of [MSS13b] to upper-bound the largest root of our expected characteristic polynomial.

Our proof has two main ingredients. The first one is the construction of a new class of expected characteristic polynomials which are the weighted average of the characteristic polynomials of the natural quadratic forms associated to the samples of the strongly Rayleigh distribution, where the weight of each polynomial is proportional to the probability of the corresponding sample set in the distribution. To show that the expected characteristic polynomial is real rooted we appeal to the theory of real stability. We show that our expected characteristic polynomial can be realized by applying $\prod_{i=1}^{m}(1 - \partial/\partial_{z_i}^2)$ operator to the real stable polynomial $g_\mu(z) \cdot \det(\sum_{i=1}^{m} z_i v_i v_i^\intercal)$, and then projecting all variables onto x.

Our second ingredient is the extension of the multivariate barrier argument. Unlike [MSS13b], here we need to prove an upper bound on the largest root of the mixed characteristic polynomial which is very close to zero. It turns out that the original idea of [BSS14] that studies the behavior of the roots of a (univariate) polynomial $p(x)$ under the operator $1 - \partial/\partial_x$ cannot establish upper bounds that are less than one. Fortunately, here we need to study the behavior of the roots of a (multivariate) polynomial $p(z)$ under the operators $1 - \partial/\partial_{z_i}^2$. The $1 - \partial/\partial_{z_i}^2$ operators allow us to impose very small shifts on the multivariate upper barrier assuming the barrier functions are sufficiently small. The intuition is that, since

$$1 - \frac{\partial}{\partial_{z_i}^2} = \left(1 - \frac{\partial}{\partial_{z_i}}\right) \cdot \left(1 + \frac{\partial}{\partial_{z_i}}\right),$$

we expect $(1 - \partial/\partial_{z_i})$ to shift the upper barrier by $1 + \Theta(\delta)$ (for some δ depending on the value of the i-th barrier function) as proved in [MSS13b] and $(1 + \partial/\partial_{z_i})$ to shift the upper barrier by $1 - \Theta(\delta)$. Therefore, applying both operators the upper barrier must be moved by no more than $\Theta(\delta)$.

4.6 Interlacing Families

We recall the definition of interlacing families of polynomials from [MSS13a], and its main consequence.

Definition 4.11. *We say that a real rooted polynomial $g(x) = \alpha_0 \prod_{i=1}^{m-1}(x - \alpha_i)$ interlaces a real rooted polynomial $f(x) = \beta_0 \prod_{i=1}^{m}(x - \beta_i)$ if*

$$\beta_1 \leq \alpha_1 \leq \beta_2 \leq \alpha_2 \leq \ldots \leq \alpha_{m-1} \leq \beta_m.$$

We say that polynomials $f_1, \ldots, f_k$ have a common interlacing if there is a polynomial g such that g interlaces all f_i. The following lemma is proved in [MSS13a].

Lemma 4.12. *Let $f_1, \ldots, f_k$ be polynomials of the same degree that are real rooted and have positive leading coefficients. Define*

$$f_\emptyset = \sum_{i=1}^{k} f_i.$$

If $f_1, \ldots, f_k$ have a common interlacing, then there is an i such that the largest root of f_i is at most the largest root of $f_\emptyset$.

Definition 4.13. *Let $\mathcal{F} \subseteq 2^{[m]}$ be nonempty. For any $S \in \mathcal{F}$, let $f_S(x)$ be a real rooted polynomial of degree d with a positive leading coefficient. For $s_1, \ldots, s_k \in \{0, 1\}$ with $k < m$, let*

$$\mathcal{F}_{s_1,\ldots,s_k} := \{S \in \mathcal{F} : i \in S \Leftrightarrow s_i = 1\}.$$

Note that $\mathcal{F} = \mathcal{F}_\emptyset$. Define

$$f_{s_1,\ldots,s_k} = \sum_{S \in \mathcal{F}_{s_1,\ldots,s_k}} f_S,$$

and

$$f_\emptyset = \sum_{S \in \mathcal{F}} f_S.$$

We say polynomials $\{f_S\}_{S \in \mathcal{F}}$ form an interlacing family *if for all $0 \leq k < m$ and all $s_1, \ldots, s_k \in \{0, 1\}$ the following holds: If both of $\mathcal{F}_{s_1,\ldots,s_k,0}$ and $\mathcal{F}_{s_1,\ldots,s_k,1}$ are nonempty, $f_{s_1,\ldots,s_k,0}$ and $f_{s_1,\ldots,s_k,1}$ have a common interlacing.*

The following is analogous to [MSS13b, Thm 3.4].

Theorem 4.14. *Let $\mathcal{F} \subseteq 2^{[m]}$ and let $\{f_S\}_{S \in \mathcal{F}}$ be an interlacing family of polynomials. Then, there exists $S \in \mathcal{F}$ such that the largest root of $f(S)$ is at most the largest root of $f_\emptyset$.*

Proof. We prove by induction. Assume that for some choice of $s_1, \ldots, s_k \in \{0, 1\}$ (possibly with $k = 0$), $\mathcal{F}_{s_1,\ldots,s_k}$ is nonempty and the largest root of $f_{s_1,\ldots,s_k}$ is at most the largest root of $f_\emptyset$. If $\mathcal{F}_{s_1,\ldots,s_k,0} = \emptyset$, then $f_{s_1,\ldots,s_k} = f_{s_1,\ldots,s_k,1}$, so we let $s_{k+1} = 1$ and we are done. Similarly, if

$\mathcal{F}_{s_1,\ldots,s_k,1} = \emptyset$, then we let $s_{k+1} = 0$ and we are done with the induction. If both of these sets are nonempty, then $f_{s_1,\ldots,s_k,0}$ and $f_{s_1,\ldots,s_k,1}$ have a common interlacing. So, by lemma 4.12, for some choice of $s_{k+1} \in \{0, 1\}$, the largest root of $f_{s_1,\ldots,s_{k+1}}$ is at most the largest root of $f_\emptyset$. □

We use the following lemma which appeared as Theorem 2.1 of [Ded92] to prove that a certain family of polynomials that we construct in section 4.7 form an interlacing family.

Lemma 4.15. *Let $f_1, \ldots, f_k$ be univariate polynomials of the same degree with positive leading coefficients. Then, $f_1, \ldots, f_k$ have a common interlacing if and only if $\sum_{i=1}^k \lambda_i f_i$ is real rooted for all convex combinations $\lambda_i \geq 0$, $\sum_{i=1}^k \lambda_i = 1$.*

4.7 The Mixed Characteristic Polynomial

For a probability distribution μ, let d_μ be the degree of the polynomial g_μ.

Theorem 4.16. *For $v_1, \ldots, v_m \in \mathbb{R}^d$ and a homogeneous probability distribution $\mu : [m] \to \mathbb{R}_+$,*

$$x^{d_\mu - d} \mathop{\mathbb{E}}_{S\sim\mu} \chi\left[\sum_{i\in S} 2 v_i v_i^\intercal\right](x^2) = \prod_{i=1}^m \left(1 - \partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1} + z) \cdot \det\left(xI + \sum_{i=1}^m z_i v_i v_i^\intercal\right)\right)\Bigg|_{z_1=\ldots=z_m=0}. \tag{4.6}$$

We call the polynomial $\mathbb{E}_{S\sim\mu}\chi[\sum_{i\in S} 2 v_i v_i^\intercal](x^2)$ the *mixed characteristic polynomial* and we denote it by $\mu[v_1, \ldots, v_m](x)$.

Proof. For $S \subseteq [m]$, let $z^{2S} = \prod_{i\in S} z_i^2$. By lemma 2.8, the coefficient of z^{2S} in

$$g_\mu(x\mathbf{1} + z) \cdot \det(xI + \sum_{i=1}^m z_i v_i v_i^\intercal)$$

is equal to

$$\left(\prod_{i\in S} \partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1} + z) \cdot \det\left(xI + \sum_{i=1}^m z_i v_i v_i^\intercal\right)\right)\Bigg|_{z_1=\ldots=z_m=0}.$$

Each of the two polynomials $g_\mu(x\mathbf{1} + z)$ and $\det(xI + \sum_{i=1}^m z_i v_i v_i^\intercal)$ is multi-linear in $z_1, \ldots, z_m$. Therefore, for $k = |S|$, the above is equal to

$$2^k \cdot \left(\prod_{i\in S} \partial_{z_i}\right) g_\mu(x\mathbf{1} + z)\Bigg|_{z_1=\ldots=z_m=0} \cdot \left(\prod_{i\in S} \partial_{z_i}\right) \det\left(xI + \sum_{i=1}^m z_i v_i v_i^\intercal\right)\Bigg|_{z_1=\ldots=z_m=0}. \tag{4.7}$$

Since g_μ is a homogeneous polynomial of degree d_μ, the first term in the above is equal to

$$x^{d_\mu - k}\mathbb{P}_{T\sim\mu}[S \subseteq T].$$

And, by lemma 2.8, the second term of (4.7) is equal to

$$x^{d-k}\sigma_k\left(\sum_{i\in S} v_i v_i^{\intercal}\right).$$

Applying the above identities for all $S \subseteq [m]$,

$$\begin{aligned}
&\prod_{i=1}^{m}\left(1-\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI+\sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)\right)\Bigg|_{z_1=\ldots=z_m=0}\\
&\qquad=\sum_{k=0}^{m}(-1)^k\sum_{S\subseteq\binom{[m]}{k}}\left(\prod_{i\in S}\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI+\sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)\right)\Bigg|_{z_1=\ldots=z_m=0}\\
&\qquad=\sum_{k=0}^{d}(-1)^k 2^k x^{d_\mu+d-2k}\sum_{S\in\binom{[m]}{k}}\mathbb{P}_{T\sim\mu}[S\subseteq T]\cdot\sigma_k\left(\sum_{i\in S} v_i v_i^{\intercal}\right)\\
&\qquad=x^{d_\mu-d}\mathop{\mathbb{E}}_{S\sim\mu}\chi\left[\sum_{i\in S}2v_i v_i^{\intercal}\right](x^2).
\end{aligned}$$

The last identity uses lemma 2.8. ∎

Corollary 4.17. *If μ is a strongly Rayleigh probability distribution, then the mixed characteristic polynomial is real-rooted.*

Proof. First, by theorem 2.15,

$$\det\left(xI+\sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)$$

is real stable. Since μ is strongly Rayleigh, $g_\mu(z)$ is real stable. So, by lemma 2.19, $g_\mu(x\mathbf{1}+z)$ is real stable. The product of two real stable polynomials is also real stable, so

$$g_\mu(x\mathbf{1}+z)\cdot\det\left(xI+\sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)$$

is real stable. Corollary 2.17 implies that

$$\prod_{i=1}^{m}\left(1-\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI+\sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)\right)$$

is real stable as well. Wagner [Wag11, Lemma 2.4(d)] tells us that real stability is preserved under setting variables to real numbers, so

$$\prod_{i=1}^{m}\left(1-\partial_{z_i}^2\right)\left(g_\mu(x\mathbf{1}+z)\cdot\det\left(xI+\sum_{i=1}^{m} z_i v_i v_i^{\intercal}\right)\right)\Bigg|_{z_1=\ldots=z_m=0}$$

is a univariate real-rooted polynomial. The mixed characteristic polynomial is equal to the above polynomial up to a term $x^{d_\mu - d}$. So, the mixed characteristic polynomial is also real rooted. □

Now, we use the real-rootedness of the mixed characteristic polynomial to show that the characteristic polynomials of the set of vectors assigned to any set S with nonzero probability in μ form an interlacing family. For a homogeneous strongly Rayleigh measure μ, let

$$\mathcal{F} = \{S : \mu(S) > 0\},$$

and for $s_1, \dots, s_k \in \{0, 1\}$ let $\mathcal{F}_{s_1,\dots,s_k}$ be as defined in definition 4.13. For any $S \in \mathcal{F}$, let

$$q_S(x) = \mu(S) \cdot \chi\left[\sum_{i \in S} 2 v_i v_i^{\intercal}\right](x^2).$$

Theorem 4.18. *The polynomials $\{q_S\}_{S \in \mathcal{F}}$ form an interlacing family.*

Proof. For $1 \le k \le m$ and $s_1, \dots, s_k \in \{0, 1\}$, let $\mu_{s_1,\dots,s_k}$ be μ conditioned on the sets $S \in \mathcal{F}_{s_1,\dots,s_k}$, i.e., μ conditioned on $i \in S$ for all $i \le k$ where $s_i = 1$ and $i \notin S$ for all $i \le k$ where $s_i = 0$. We inductively write the generating polynomial of $\mu_{s_1,\dots,s_k}$ in terms of g_μ. Say we have written $g_{\mu_{s_1,\dots,s_k}}$ in terms of g_μ. Then, we can write,

$$g_{\mu_{s_1,\dots,s_k,1}}(z) = \frac{z_{k+1} \cdot \partial_{z_{k+1}} g_{\mu_{s_1,\dots,s_k}}(z)}{\partial_{z_{k+1}} g_{\mu_{s_1,\dots,s_k}}(z)\big|_{z_i=1}}, \tag{4.8}$$

$$g_{\mu_{s_1,\dots,s_k,0}}(z) = \frac{g_{\mu_{s_1,\dots,s_k}}(z)\big|_{z_{k+1}=0}}{g_{\mu_{s_1,\dots,s_k}}(z)\big|_{z_{k+1}=0, z_i=1 \text{ for } i \neq k+1}}. \tag{4.9}$$

Note that the denominators of both equations are just normalizing constants. The above polynomials are well defined if the normalizing constants are nonzero, i.e., if the set $\mathcal{F}_{s_1,\dots,s_k,s_{k+1}}$ is nonempty. Since the real stable polynomials are closed under differentiation and substitution, for any $1 \le k \le m$, and $s_1, \dots, s_k \in \{0, 1\}$, if $g_{\mu_{s_1,\dots,s_k}}$ is well defined, it is real stable, so $\mu_{s_1,\dots,s_k}$ is a strongly Rayleigh distribution.

Now, for $s_1, \dots, s_k \in \{0, 1\}$, let

$$q_{s_1,\dots,s_k}(x) = \sum_{S \in \mathcal{F}_{s_1,\dots,s_k}} q_S(x).$$

Since $\mu_{s_1,\dots,s_k}$ is strongly Rayleigh, by corollary 4.17, $q_{s_1,\dots,s_k}(x)$ is real rooted.

By lemma 4.15, to prove the theorem it is enough to show that if $\mathcal{F}_{s_1,\dots,s_k,0}$ and $\mathcal{F}_{s_1,\dots,s_k,1}$ are nonempty, then for any $0 < \lambda < 1$,

$$\lambda \cdot q_{s_1,\dots,s_k,1}(x) + (1 - \lambda) \cdot q_{s_1,\dots,s_k,0}(x)$$

is real rooted. Equivalently, by corollary 4.17, it is enough to show that for any $0 < \lambda < 1$,

$$\lambda \cdot g_{\mu_{s_1,\ldots,s_k,1}}(z) + (1-\lambda) \cdot g_{\mu_{s_1,\ldots,s_k,0}}(z) \tag{4.10}$$

is real stable. Let us write,

$$\begin{aligned} g_{\mu_{s_1,\ldots,s_k}}(z) &= z_{k+1} \cdot \partial_{z_{k+1}} g_{\mu_{s_1,\ldots,s_k}}(z) + g_{\mu_{s_1,\ldots,s_k}}(z)\big|_{z_{k+1}=0} \\ &= \alpha \cdot g_{\mu_{s_1,\ldots,s_k,1}}(z) + \beta \cdot g_{\mu_{s_1,\ldots,s_k,0}}(z), \end{aligned}$$

for some $\alpha, \beta > 0$. The second identity follows by (4.8) and (4.9). Let $\lambda_{k+1} > 0$ such that

$$\frac{\lambda_{k+1} \cdot \alpha}{\lambda} = \frac{\beta}{1-\lambda}. \tag{4.11}$$

Since $g_{\mu_{s_1,\ldots,s_k}}$ is real stable, by lemma 2.18

$$g_{\mu_{s_1,\ldots,s_k}}(z_1, \ldots, z_k, \lambda_{k+1} \cdot z_{k+1}, z_{k+2}, \ldots, z_m)$$

is real stable. But, by (4.11) the above polynomial is just a multiple of (4.10). So, (4.10) is real stable. □

4.8 The Multivariate Barrier Argument

In this section we upper-bound the roots of the mixed characteristic polynomial in terms of the marginal probabilities of elements of $[m]$ in μ and the maximum of the squared norm of vectors $v_1, \ldots, v_m$.

Theorem 4.19. *Given vectors $v_1, \ldots, v_m \in \mathbb{R}^d$, and a homogeneous strongly Rayleigh probability distribution $\mu : [m] \to \mathbb{R}_+$, such that the marginal probability of each element $i \in [m]$ is at most ϵ_1, $\sum_{i=1}^m v_i v_i^\intercal = I$ and $\left\|v_i\right\|^2 \leq \epsilon_2$, the largest root of $\mu[v_1, \ldots, v_m](x)$ is at most $4(2\epsilon + \epsilon^2)$, where $\epsilon = \epsilon_1 + \epsilon_2$,*

First, similar to [MSS13b] we derive a slightly different expression.

Lemma 4.20. *For any probability distribution μ and vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ such that $\sum_{i=1}^m v_i v_i^\intercal = I$,*

$$x^{d_\mu - d} \mu[v_1, \ldots, v_m](x) = \prod_{i=1}^m \left(1 - \partial_{y_i}^2\right) \left(g_\mu(y) \cdot \det\left(\sum_{i=1}^m y_i v_i v_i^\intercal \right) \right) \Bigg|_{y_1 = \ldots = y_m = x}.$$

Proof. This is because for any differentiable function f, $\partial_{y_i} f(y_i)|_{y_i = z_i + x} = \partial_{z_i} f(z_i + x)$. □

Let

$$Q(y_1, \dots, y_m) = \prod_{i=1}^{m} \left(1 - \partial_{y_i}^2\right) \left(g_\mu(y) \cdot \det \left(\sum_{i=1}^{m} y_i v_i v_i^\intercal \right) \right).$$

Then, by the above lemma, the maximum root of $Q(x, \dots, x)$ is the same as the maximum root of $\mu[v_1, \dots, v_m](x)$. In the rest of this section we upper-bound the maximum root of $Q(x, \dots, x)$.

It directly follows from the proof of Theorem 5.1 in [MSS13b] that the maximum root of $Q(x, \dots, x)$ is at most $(1 + \sqrt{\epsilon})^2$. But, in our setting, any upper-bound that is more than 1 obviously holds, as for any $S \subseteq [m]$,

$$\left\| \sum_{i=1}^{m} v_i v_i^\intercal \right\| \leq 1.$$

The main difficulty that we are facing is to prove an upper-bound of $O(\epsilon)$ on the maximum root of $Q(x, \dots, x)$.

We use an extension of the multivariate barrier argument of [MSS13b] to upper-bound the maximum root of Q. We manage to prove a significantly smaller upper-bound because we apply $1 - \partial_{y_i}^2$ operators as opposed to the $1 - \partial_{y_i}$ operators used in [MSS13b]. This allows us to impose significantly smaller shifts on the barrier upper-bound in our inductive argument.

Definition 4.21. *For a multivariate polynomial $p(z_1, \dots, z_m)$, we say $z \in \mathbb{R}^m$ is above all roots of p if for all $t \in \mathbb{R}_+^m$,*

$$p(z + t) > 0.$$

We use Ab_p to denote the set of points which are above all roots of p.

We use the same barrier function defined in [MSS13b].

Definition 4.22. *For a real stable polynomial p, and $z \in \mathrm{Ab}_p$, the barrier function of p in direction i at z is*

$$\Phi_p^i(z) := \frac{\partial_{z_i} p(z)}{p(z)} = \partial_{z_i} \log p(z).$$

To analyze the rate of change of the barrier function with respect to the $1 - \partial_{z_i}^2$ operator, we need to work with the second derivative of p as well. We define,

$$\Psi_p^i(z) := \frac{\partial_{z_i}^2 p(z)}{p(z)}.$$

Equivalently, for a univariate restriction $q_{z,i}(t) = p(z_1, \dots, z_{i-1}, t, z_{i+1}, \dots, z_m)$, with real roots $\lambda_1, \dots, \lambda_r$ we can write,

$$\Phi_p^i(z) = \frac{q'_{z,i}(z_i)}{q_{z,i}(z_i)} = \sum_{j=1}^{r} \frac{1}{z_i - \lambda_j},$$

$$\Psi_p^i(z) = \frac{q''_{z,i}(z_i)}{q_{z,i}(z_i)} = \sum_{1 \leq j < k \leq r} \frac{2}{(z_i - \lambda_j)(z_i - \lambda_k)}.$$

The following lemma is immediate from the above definition.

Lemma 4.23. *If p is real stable and $z \in \mathrm{Ab}_p$, then for all $i \leq m$,*

$$\Psi_p^i(z) \leq \Phi_p^i(z)^2.$$

Proof. Since $z \in \mathrm{Ab}_p$, $z_i > \lambda_j$ for all $1 \leq j \leq r$, so,

$$\Phi_p^i(z)^2 - \Psi_p^i(z) = \left(\sum_{j=1}^r \frac{1}{z_i - \lambda_j}\right)^2 - \sum_{1 \leq j < k \leq r} \frac{2}{(z_i - \lambda_j)(z_i - \lambda_k)} = \sum_{j=1}^r \frac{1}{(z_i - \lambda_j)^2} > 0.$$

□

The following monotonicity and convexity properties of the barrier functions are proved in [MSS13b].

Lemma 4.24. *Suppose $p(.)$ is a real stable polynomial and $z \in \mathrm{Ab}_p$. Then, for all $i, j \leq m$ and $\delta \geq 0$,*

$$\Phi_p^i(z + \delta \mathbf{1}_j) \leq \Phi_p^i(z) \text{ and,} \qquad \text{(monotonicity)} \tag{4.12}$$

$$\Phi_p^i(z + \delta \mathbf{1}_j) \leq \Phi_p^i(z) + \delta \cdot \partial_{z_j} \Phi_p^i(z + \delta \mathbf{1}_j) \qquad \text{(convexity).} \tag{4.13}$$

Recall that the purpose of the barrier functions Φ_p^i is to allow us to reason about the relationship between Ab_p and $\mathrm{Ab}_{p - \partial_{z_i}^2 p}$; the monotonicity property and lemma 4.23 imply the following lemma.

Lemma 4.25. *If p is real stable and $z \in \mathrm{Ab}_p$ is such that $\Phi_p^i(z) < 1$, then $z \in \mathrm{Ab}_{p - \partial_{z_i}^2 p}$.*

Proof. Fix a nonnegative vector t. Since Φ is nonincreasing in each coordinate,

$$\Phi_p^i(z + t) \leq \Phi_p^i(z) < 1.$$

Since $z + t \in \mathrm{Ab}_p$, by lemma 4.23,

$$\Psi_p^i(z + t) \leq \Phi_p^i(z + t)^2 < 1.$$

Therefore,

$$\partial_{z_i}^2 p(z + t) < p(z + t) \Rightarrow (1 - \partial_{z_i}^2) p(z + t) > 0,$$

as desired. □

We use an inductive argument similar to [MSS13b]. We argue that when we apply each operator $(1 - \partial_{z_j}^2)$, the barrier functions, $\Phi_p^i(z)$, do not increase by shifting the upper bound along the direction $\mathbf{1}_j$. As we would like to prove a significantly smaller upper bound on the maximum root of the mixed characteristic polynomial, we may only shift along direction $\mathbf{1}_j$ by a small amount. In the following lemma we show that when we apply the $(1 - \partial_{z_j}^2)$ operator we only need to shift the upper bound proportional to $\Phi_p^j(z)$ along the direction $\mathbf{1}_j$.

Lemma 4.26. *Suppose that $p(z_1, \ldots, z_m)$ is real stable and $z \in \mathrm{Ab}_p$. If for $\delta > 0$,*

$$\frac{2}{\delta}\Phi_p^j(z) + \Phi_p^j(z)^2 \leq 1,$$

then, for all i,

$$\Phi^i_{p-\partial^2_{z_j}p}(z + \delta \cdot \mathbf{1}_j) \leq \Phi^i_p(z).$$

To prove the above lemma we first need to prove a technical lemma to upper-bound $\frac{\partial_{z_i}\Psi_p^j(z)}{\partial_{z_i}\Phi_p^j(z)}$. We use the following characterization of the bivariate real stable polynomials proved by Lewis, Parrilo, and Ramana [LPR05]. The following form is stated in [BB10, Cor 6.7].

Lemma 4.27. *If $p(z_1, z_2)$ is a bivariate real stable polynomial of degree d, then there exist $d \times d$ positive semidefinite matrices A, B and a Hermitian matrix C such that*

$$p(z_1, z_2) = \pm \det(z_1 A + z_2 B + C).$$

Lemma 4.28. *Suppose that p is real stable and $z \in \mathrm{Ab}_p$, then for all $i, j \leq m$,*

$$\frac{\partial_{z_i}\Psi_p^j(z)}{\partial_{z_i}\Phi_p^j(z)} \leq 2\Phi_p^j(z).$$

Proof. If $i = j$, then we consider the univariate restriction $q_{z,i}(z_i) = \prod_{k=1}^r (z_i - \lambda_k)$. Then,

$$\frac{\partial_{z_i}\sum_{1\leq k<\ell\leq r}\frac{2}{(z_i-\lambda_k)(z_i-\lambda_\ell)}}{\partial_{z_i}\sum_{k=1}^r\frac{1}{(z_i-\lambda_k)}} = \frac{\sum_{k\neq \ell}\frac{-2}{(z_i-\lambda_k)^2(z_i-\lambda_\ell)}}{\sum_{k=1}^r \frac{-1}{(z_i-\lambda_k)^2}} \leq \sum_{\ell=1}^r \frac{2}{(z_i-\lambda_\ell)} = 2\Phi_p^j(z).$$

The inequality uses the assumption that $z \in \mathrm{Ab}_p$.

If $i \neq j$, we fix all variables other than z_i, z_j and we consider the bivariate restriction

$$q_{z,ij}(z_i, z_j) = p(z_1, \ldots, z_m).$$

By lemma 4.27, there are Hermitian positive semidefinite matrices B_i, B_j, and a Hermitian matrix C such that

$$q_{z,ij}(z_i, z_j) = \pm \det(z_i B_i + z_j B_j + C).$$

Let $M = z_i B_i + z_j B_j + C$. Marcus, Spielman, and Srivastava [MSS13b, Lem 5.7] observed that the sign is always positive, that $B_i + B_j$ is positive definite. In addition, M is positive definite since $B_i + B_j$ is positive definite and $z \in \mathrm{Ab}_p$.

By theorem 2.9, the barrier function in direction j can be expressed as

$$\Phi_p^j(z) = \frac{\partial_{z_j}\det(M)}{\det(M)} = \frac{\det(M)\,\mathrm{Tr}(M^{-1}B_j)}{\det(M)} = \mathrm{Tr}(M^{-1}B_j). \tag{4.14}$$

By another application of theorem 2.9,

$$\begin{aligned}
\Psi_p^j(z) = \frac{\partial_{z_j}^2 \det(M)}{\det(M)} &= \frac{\partial_{z_j}(\det(M)\operatorname{Tr}(M^{-1}B_j))}{\det(M)} \\
&= \frac{\det(M)\operatorname{Tr}(M^{-1}B_j)^2}{\det(M)} + \frac{\det(M)\operatorname{Tr}((\partial_{z_j}M^{-1})B_j)}{\det(M)} \\
&= \operatorname{Tr}(M^{-1}B_j)^2 + \operatorname{Tr}(-M^{-1}B_jM^{-1}B_j) \\
&= \operatorname{Tr}(M^{-1}B_j)^2 - \operatorname{Tr}((M^{-1}B_j)^2).
\end{aligned}$$

The second to last identity uses lemma 2.10. Next, we calculate $\partial_{z_i}\Phi_p^j$ and $\partial_{z_i}\Psi_p^j$. First, by another application of lemma 2.10,

$$\partial_{z_i}M^{-1}B_j = -M^{-1}B_iM^{-1}B_j =: L.$$

Therefore,

$$\partial_{z_i}\Phi_p^j(z) = \partial_{z_i}\operatorname{Tr}(M^{-1}B_j) = \operatorname{Tr}(L),$$

and

$$\begin{aligned}
\partial_{z_i}\Psi_p^j(z) &= \partial_{z_i}\operatorname{Tr}(M^{-1}B_j)^2 - \partial_{z_i}\operatorname{Tr}((M^{-1}B_j)^2) \\
&= 2\operatorname{Tr}(M^{-1}B_j)\operatorname{Tr}(L) - \operatorname{Tr}\left(L(M^{-1}B_j) + (M^{-1}B_j)L\right) \\
&= 2\operatorname{Tr}(M^{-1}B_j)\operatorname{Tr}(L) - 2\operatorname{Tr}(LM^{-1}B_j).
\end{aligned}$$

Putting above equations together we get

$$\begin{aligned}
\frac{\partial_{z_i}\Psi_p^j(z)}{\partial_{z_i}\Phi_p^j(z)} &= 2\frac{\operatorname{Tr}(M^{-1}B_j)\operatorname{Tr}(L) - \operatorname{Tr}(LM^{-1}B_j)}{\operatorname{Tr}(L)} \\
&= 2\operatorname{Tr}(M^{-1}B_j) - 2\frac{\operatorname{Tr}(LM^{-1}B_j)}{\operatorname{Tr}(L)} \\
&= 2\Phi_p^j(z) - 2\frac{\operatorname{Tr}(LM^{-1}B_j)}{\operatorname{Tr}(L)}
\end{aligned}$$

where we used (4.14).

To prove the lemma it is enough to show that $\frac{\operatorname{Tr}(LM^{-1}B_j)}{\operatorname{Tr}(L)} \geq 0$. We show that both the numerator and the denominator are nonpositive. First,

$$\operatorname{Tr}(L) = -\operatorname{Tr}(M^{-1}B_iM^{-1}B_j) \leq 0$$

where we used that $M^{-1}B_iM^{-1}$ and B_j are positive semidefinite and the fact that the trace of the product of positive semidefinite matrices is nonnegative. Secondly,

$$\operatorname{Tr}(LM^{-1}B_j) = \operatorname{Tr}(-M^{-1}B_iM^{-1}B_jM^{-1}B_j) = -\operatorname{Tr}(B_iM^{-1}B_jM^{-1}B_jM^{-1}) \leq 0,$$

where we again used that $M^{-1}B_jM^{-1}B_jM^{-1}$ and B_i are positive semidefinite and the trace of the product of two positive semidefinite matrices is nonnegative. $\square$

Proof of lemma 4.26. We write ∂_i instead of ∂_{z_i} for the ease of notation. First, we write $\Phi^i_{p-\partial_j^2 p}$ in terms of Φ^i_p and Ψ^j_p and $\partial_i \Psi^j_p$.

$$\begin{aligned}\Phi^i_{p-\partial_j^2 p} &= \frac{\partial_i(p - \partial_j^2 p)}{p - \partial_j^2 p} \\ &= \frac{\partial_i((1-\Psi^j_p)p)}{(1-\Psi^j_p)p} \\ &= \frac{(1-\Psi^j_p)(\partial_i p)}{(1-\Psi^j_p)p} + \frac{(\partial_i(1-\Psi^j_p))p}{(1-\Psi^j_p)p} \\ &= \Phi^i_p - \frac{\partial_i \Psi^j_p}{1-\Psi^j_p}.\end{aligned}$$

We would like to show that $\Phi^i_{p-\partial_j^2 p}(z+\delta\mathbf{1}_j) \le \Phi^i_p(z)$. Equivalently, it is enough to show that

$$-\frac{\partial_i \Psi^j_p(z+\delta\mathbf{1}_j)}{1-\Psi^j_p(z+\delta\mathbf{1}_j)} \le \Phi^i_p(z) - \Phi^i_p(z+\delta\mathbf{1}_j).$$

By (4.13) of lemma 4.24, it is enough to show that

$$-\frac{\partial_i \Psi^j_p(z+\delta\mathbf{1}_j)}{1-\Psi^j_p(z+\delta\mathbf{1}_j)} \le \delta \cdot (-\partial_j \Phi^i_p(z+\delta\mathbf{1}_j)).$$

By (4.12) of lemma 4.24, $\delta \cdot (-\partial_j \Phi^i_p(z+\delta\mathbf{1}_j)) > 0$ so we may divide both sides of the above inequality by this term and obtain

$$\frac{-\partial_i \Psi^j_p(z+\delta\mathbf{1}_j)}{-\delta \cdot \partial_i \Phi^j_p(z+\delta\mathbf{1}_j)} \cdot \frac{1}{1-\Psi^j_p(z+\delta\mathbf{1}_j)} \le 1,$$

where we also used $\partial_j \Phi^i_p = \partial_i \Phi^j_p$. By lemma 4.28, $\frac{\partial_i \Psi^j_p}{\partial_i \Phi^j_p} \le 2\Phi^j_p$. So, we can write,

$$\frac{2}{\delta}\Phi^j_p(z+\delta\mathbf{1}_j) \cdot \frac{1}{1-\Psi^j_p(z+\delta\mathbf{1}_j)} \le 1.$$

By lemma 4.23 and (4.12) of lemma 4.24,

$$\begin{aligned}&\Phi^j_p(z+\delta\mathbf{1}_j) \le \Phi^j_p(z), \\ &\Psi^j_p(z+\delta\mathbf{1}_j) \le \Phi^j_p(z+\delta\mathbf{1}_j)^2 \le \Phi^j_p(z)^2.\end{aligned}$$

So, it is enough to show that

$$\frac{2}{\delta}\Phi^j_p(z) \cdot \frac{1}{1-\Phi^j_p(z)^2} \le 1$$

Using $\Phi_p^j(z) < 1$ we may multiply both sides with $1 - \Phi_p^j(z)$ and we obtain,

$$\frac{2}{\delta}\Phi_p^j(z) + \Phi_p^j(z)^2 \leq 1,$$

as desired. $\square$

Now, we are read to prove theorem 4.19.

Proof of theorem 4.19. Let

$$p(y_1, \ldots, y_m) = g_\mu(y) \cdot \det\left(\sum_{i=1}^m y_i v_i v_i^\intercal\right).$$

Set $\epsilon = \epsilon_1 + \epsilon_2$ and

$$\delta = t = \sqrt{2\epsilon + \epsilon^2}.$$

For any $z \in \mathbb{R}^m$ with positive coordinates, $g_\mu(z) > 0$, and additionally

$$\det\left(\sum_{i=1}^m z_i v_i v_i^\intercal\right) > 0.$$

Therefore, for every $t > 0$, $t\mathbf{1} \in \mathrm{Ab}_p$.

Now, by theorem 2.9,

$$\begin{aligned}\Phi_p^i(y) &= \frac{(\partial_i g_\mu(y)) \cdot \det(\sum_{i=1}^m y_i v_i v_i^\intercal)}{g_\mu(y) \cdot \det(\sum_{i=1}^m y_i v_i v_i^\intercal)} + \frac{g_\mu(y) \cdot (\partial_i \det(\sum_{i=1}^m y_i v_i v_i^\intercal))}{g_\mu(y) \cdot \det(\sum_{i=1}^m y_i v_i v_i^\intercal)} \\ &= \frac{\partial_i g_\mu(y)}{g_\mu(y)} + \mathrm{Tr}\left(\left(\sum_{i=1}^m y_i v_i v_i^\intercal\right)^{-1} v_i v_i^\intercal\right)\end{aligned}$$

Therefore, since g_μ is homogeneous,

$$\begin{aligned}\Phi_p^i(t\mathbf{1}) &= \frac{1}{t} \cdot \frac{\partial_i g_\mu(\mathbf{1})}{g_\mu(\mathbf{1})} + \frac{\left\|v_i\right\|^2}{t} \\ &= \frac{\mathbb{P}_{S\sim\mu}[i \in S]}{t} + \frac{\left\|v_i\right\|^2}{t} \leq \frac{\epsilon_1}{t} + \frac{\epsilon_2}{t} = \frac{\epsilon}{t}.\end{aligned}$$

The second identity uses (4.1). Let $\phi = \epsilon/t$. Using $t = \delta$, it follows that

$$\frac{2}{\delta}\phi + \phi^2 = \frac{2\epsilon}{t^2} + \frac{\epsilon^2}{t^2} = 1.$$

For $k \in [m]$ define

$$p_k(y_1, \ldots, y_m) = \prod_{i=1}^k \left(1 - \partial_{y_i}^2\right)\left(g_\mu(y) \cdot \det\left(\sum_{i=1}^m y_i v_i v_i^\intercal\right)\right),$$

and note that $p_m = Q$. Let x^0 be the all-t vector and x^k be the vector that is $t+\delta$ in the first k coordinates and t in the rest. By inductively applying lemma 4.25 and lemma 4.26 for any $k \in [m]$, x^k is above all roots of p_k and for all i,

$$\Phi^i_{p_k}(x_k) \le \phi \Rightarrow \frac{2}{\delta}\Phi^i_{p_k}(x_i) + \Phi^i_{p_k}(x_i)^2 \le 1.$$

Therefore, the largest root of $\mu[v_1, \dots, v_m](x)$ is at most

$$t + \delta = 2\sqrt{2\epsilon + \epsilon^2}.$$

$\square$

Proof of theorem 4.2. Let $\epsilon = \epsilon_1 + \epsilon_2$ as always. Theorem 4.19 implies that the largest root of the mixed characteristic polynomial, $\mu[v_1, \dots, v_m](x)$, is at most $2\sqrt{2\epsilon + \epsilon^2}$. Theorem 4.18 tells us that the polynomials $\{q_S\}_{S:\mu(S)>0}$ form an interlacing family. So, by theorem 4.14 there is a set $S \subseteq [m]$ with $\mu(S) > 0$ such that the largest root of

$$\det\left(x^2 I - \sum_{i \in S} 2 v_i v_i^\intercal\right)$$

is at most $2\sqrt{2\epsilon + \epsilon^2}$. This implies that the largest root of

$$\det\left(x I - \sum_{i \in S} 2 v_i v_i^\intercal\right)$$

is at most $(2\sqrt{2\epsilon + \epsilon^2})^2$. Therefore,

$$\left\|\sum_{i \in S} v_i v_i^\intercal\right\| = \frac{1}{2}\left\|\sum_{i \in S} 2 v_i v_i^\intercal\right\| \le \frac{1}{2}(2\sqrt{2\epsilon + \epsilon^2})^2 = 4\epsilon + 2\epsilon^2.$$

$\square$

4.9 Remarks about the Extension

Similar to [MSS13b] our main theorem is not algorithmic, i.e., we are not aware of any polynomial time algorithm that for a given homogeneous strongly Rayleigh distribution with small marginal probabilities and for a set of vectors assigned to the underlying elements with small norm finds a sample of the distribution with spectral norm bounded away from 1. We will however touch on possible directions for creating an algorithm in section 8.2.

Although our theorem can be seen as a generalization of [MSS13b], the bound that we prove on the maximum root of the mixed characteristic polynomial is incomparable to the bound of theorem 4.1. In corollary 4.3 we used our theorem to prove Weaver's KS_r conjecture [Wea04] for $r > 4$. It is an interesting question to see if the dependency on ϵ in our multivariate barrier can be improved, and if one can reprove KS_2 using our machinery.

Chapter 5

Preconditioning for Kadison-Singer

In this chapter we will try to design a framework based on our idea of reducing L_1 problems to L_2 problems. For concreteness we will stick with the Kadison-Singer problem, but this technique has already been used for problems such as the Sparsest Cut problem [ARV09].

Remember that our formulation of the Kadison-Singer problem is roughly the following: The goal is to find a subset S from the given rank 1 matrices $A_1, \ldots, A_n \succeq 0$ such that

$$\sum_{i \in S} A_i \preceq \sum_{i=1}^{n} A_i. \tag{5.1}$$

This subset may have to be chosen from a strongly Rayleigh distribution, but in this chapter our focus is on (5.1).

5.1 Kadison-Singer for Given Test Vectors

Assume that we do not necessarily want (5.1) to be satisfied, but rather our goal is to have the following

$$\forall x \in W \quad x^\intercal \left(\sum_{i \in S} A_i \right) x \leq x^\intercal \left(\sum_{i=1}^{n} A_i \right) x, \tag{5.2}$$

where W is a given collection of test vectors. For example if we want thin trees, W would be the set $\{\mathbf{1}_S : S \subseteq V\}$. If $W = \mathbb{R}^n$, then (5.2) reduces to (5.1).

We can generalize our ideas from thin trees to such general test vectors.

Definition 5.1. *Given a set of test vectors W and rank 1 matrices $A_1, \ldots, A_n \succeq 0$, a matrix $D \succeq 0$ is an ϵ-preconditioner for Kadison-Singer if and only if*

$$\forall x \in W \quad x^\intercal D x \leq x^\intercal \left(\sum_{i=1}^{n} A_i \right) x,$$

and

$$\forall i \quad A_i \preceq \epsilon \cdot D.$$

If we are able to find a preconditioner D, we can add it to the set of A_i's without changing the law of the random subset S. This makes sure that the random subset S does not contain any part of the preconditioner. If Kadison-Singer's conditions are met, we are left with a subset that satisfies (5.2).

We will demonstrate this technique using an example problem in the next section.

5.2 Bounded Degree Spanning Trees

In this section we provide a simple application of corollary 4.10. We show that k-regular k-edge-connected graphs have constant degree spanning trees.

Lemma 5.2. *Any k-edge-connected graph $G = (V, E)$ has a spanning tree $T \subseteq E$ such that for any $i \in V$,*

$$d_T(i) \leq O(1/k) d_G(i),$$

where $d_T(i), d_G(i)$ are the degree of v in T and G respectively.

In two beautiful works Goemans [Goe06] and Lau, Singh [SL07] used combinatorial and polyhedral techniques to prove a generalization of the above lemma. Next, we use corollary 4.10 to give a *spectral* proof of the above lemma.

Our proof of the above lemma follows the high level plan of our approach to the ATSP that we discussed in the previous chapters. The main difference is that because here we are only interested in finding a spanning three which in "only thin with respect to degree cuts", we can find the best possible matrix D independently from the cut structure of graph G. In particular, we simply let $D = kI$. Then, for any edge $e \in F$, we have

$$b_e^\intercal D^{-1} b_e = b_e^\intercal \frac{I}{k} b_e = \frac{\left\|b_e\right\|^2}{k} = \frac{2}{k},$$

where we used that for any edge e oriented from i to j, $\left\|\mathbf{1}_e\right\|^2 = \left\|\mathbf{1}_i\right\|^2 + \left\|\mathbf{1}_j\right\|^2 = 2$. Since E is k-edge-connected, by corollary 4.10, there is a spanning tree $T \subseteq E$ such that

$$L_T \preceq O(2/k + 1/k) \cdot (L_G + kI).$$

Fix a vertex $i \in V$. We show that $d_T(i) \leq O(1/k) d_G(i)$. Multiplying both sides of the above equation with $\mathbf{1}_i$ we get

$$d_T(i) = \mathbf{1}_i^\intercal L_T \mathbf{1}_i \leq O(1/k) \mathbf{1}_i^\intercal (L_G + kI) \mathbf{1}_i \tag{5.3}$$

We can upper bound the RHS as follows,

$$\mathbf{1}_i^\intercal (L_G + kI) \mathbf{1}_i = \mathbf{1}_i^\intercal L_G \mathbf{1}_i + k \mathbf{1}_i^\intercal I \mathbf{1}_i = d_G(i) + k \left\|\mathbf{1}_i\right\|^2 = d_G(i) + k \leq 2 d_G(i),$$

where the inequality uses that G is k-edge-connected, so $d_G(i) \geq k$. Putting the above equation together with (5.3) proves the lemma.

Chapter 6

Hierarchical Decompositions

6.1 Proof of the Main Theorem via Hierarchical Decompositions

In this section we prove our main theorem 3.6 assuming the main technical theorem 4.2. Lastly, we will prove the algorithmic theorem 3.9. First, in section 6.2 we show that for $k \geq 7 \log n$, any k-edge-connected graph has a $(1/k, .)$-expanding $(k/20, 1/4, \mathcal{T})$-LCH. Then, in section 6.3, we show that if a given graph $G = (V, E)$ has an $(\alpha, .)$-expanding $(k, ., .)$-LCH, then there exists a PD shortcut matrix D, and an $\Omega(k)$-edge-connected subset F of good edges, such that for any $e \in F$,

$$\mathcal{R}\text{eff}_D(e) \leq \frac{\text{poly}\log(k, 1/\alpha)}{k}.$$

6.2 Construction of Locally Connected Hierarchies

In this section, we prove the following theorem. We remark that this is the only place in the entire dissertation where we depend on k being $\Omega(\log(n))$.

Theorem 6.1. *Given a k-edge-connected graph G, with $k \geq 7\log(n)$, one can construct a $(\frac{1}{k}, .)$-expanding $(\frac{k}{20}, \frac{1}{4}, \mathcal{T})$-LCH $\mathcal{T}$.*

The proof of the theorem will be an adaptation of the proof for the special case of k-edge-connected planar graphs that we saw in lemma 3.12. Given a graph G, we iteratively find $\Omega(k)$-edge-connected $\Omega(1/k)$ induced expanders, i.e., a set $S \subseteq V$ where $G[S]$ is $\Omega(k)$-edge-connected and $\phi(G[S]) \geq \Omega(1/k)$. We also need to make sure that $G[S]$ satisfies the following definition to ensure that we get a $(., \lambda, .)$-LCH.

Definition 6.2. *An induced subgraph H of an unweighted graph $G = (V, E)$ is λ-dense if for any $v \in V(H)$,*

$$d_H(v) \geq \lambda \cdot d_G(v),$$

where we use $V(H)$ to denote the vertex set of H.

The following proposition is the main technical statement that we need for the proof.

Proposition 6.3. *Any $k \geq 7 \log n$-edge-connected graph $G = (V, E)$ (with n vertices) has an induced $k/20$-edge-connected, $1/4$-dense subgraph $G[S]$ that is an $1/k$-expander.*

Note that for every edge $\{u, v\} \in E$, the induced graph $G[\{u, v\}]$ is a 1-expander. But, if there is only one edge between u, v in G, then this induced graph is only 1-edge-connected and $O(1/k)$-dense. It is instructive to compare the statement of the above proposition to the planar case. Recall that Fact 3.13 asserts that in any k-edge-connected planar graph there is a pair of vertices with $k/5$ parallel edges. Such an induced graph is a $k/5$-edge-connected 1-expander. Of course, this fact does not necessarily hold for a general k-edge-connected graph as G may not have any parallel edges at all.

Note that, in the above proposition, the condition $k \geq 7 \log n$ is necessary up to a constant; a tight example is the $\log n$-dimensional hypercube, which is a k-edge-connected for any $k \leq \log n$, but every $\Omega(1)$-dense induced subgraph is no better than $O(1/\log n)$-expanding.

We use proof by contradiction. Suppose G does not have any induced subgraph satisfying the statement of the proposition. Then, invoking the following lemma with $H = G$ and $\phi^* = 1/k$, we obtain that G must have more than $2^{3k/20}$ vertices. But this contradicts the fact that $k \geq 7 \log n$.

Lemma 6.4. *Given a k-edge-connected graph G, if every $k/20$-edge-connected $1/4$-dense subgraph $G[S]$ of G satisfies $\phi(G[S]) < \phi^*$, then for any induced subgraph H of G,*

$$\log_2(|V(H)|) \geq \frac{3/10 - \phi_G(V(H))}{2\phi^*}.$$

Proof. We prove the lemma by induction on the number of vertices of H. Fix an induced subgraph $H = G[U]$. Without loss of generality, assume that $\phi_G(U) < 3/10$. We consider two cases, and in the end we show that one of them always happens.

Case 1: There is a vertex $v \in U$ such that $d_H(v) \leq 7d_G(v)/20$. We show that $\phi_G(U)$ decreases when we remove v from U.

$$\phi_G(U) = \frac{\partial_G(U \setminus \{v\}) + d_G(v) - 2d_H(v)}{d_G(U \setminus \{v\}) + d_G(v)} \geq \frac{\partial_G(U \setminus \{v\}) + 6d_G(v)/20}{d_G(U \setminus \{v\}) + d_G(v)} \geq \phi_G(U \setminus \{v\})$$

The last inequality uses that $\phi_G(U) < 3/10$. By induction,

$$\log_2(|U|) \geq \log_2(|U - \{v\}|) \geq \frac{3/10 - \phi_G(U - \{v\})}{2\phi^*} \geq \frac{3/10 - \phi_G(U)}{2\phi^*},$$

and we are done. Note that if this case does not happen, then H is $\frac{7}{20}$-dense in G.

Case 2: For some $S \subset U$, $\max\{\phi_H(S), \phi_H(U \setminus S)\} < \phi^*$. Let $T := U \setminus S$. Observe that if $\phi_G(S) \leq \phi_G(U)$ or $\phi_G(T) \leq \phi_G(U)$, then we are done by induction. So assume that none of the two conditions hold. We show that $\phi_G(S), \phi_G(T) \leq \phi_G(U) + 2\phi^*$.

First, it follows from
$$\phi_G(U) = \frac{\partial_G(S) + \partial_G(T) - 2\partial_H(T)}{d_G(S) + d_G(T)}$$
and $\frac{\partial_G(S)}{d_G(S)} = \phi_G(S) > \phi_G(U)$ that
$$\phi_G(U) > \frac{\partial_G(T) - 2\partial_H(T)}{d_G(T)} = \phi_G(T) - 2\frac{\partial_H(T)}{d_G(T)} \geq \phi_G(T) - 2\phi_H(T). \tag{6.1}$$
Therefore, $\phi_G(T) \leq \phi_G(U) + 2\phi^*$. Similarly, we can show $\phi_G(S) \leq \phi_G(U) + 2\phi^*$. So, by induction,
$$\log_2(|U|) = \log_2(|S|+|T|) \geq 1+\log_2(\min\{|S|, |T|\}) \geq 1+\frac{3/10 - \phi_G(U) - 2\phi^*}{2\phi^*} = \frac{3/10 - \phi_G(U)}{2\phi^*}.$$

We now show that one of the above cases (Case 1 and Case 2) need to happen. Suppose towards contradiction that none of the above cases happens. Then H is 7/20-dense and for all $S \subset U : \max\{\phi_H(S), \phi_H(U \setminus S)\} \geq \phi^*$. In other words, $\phi(H) \geq \phi^*$. Therefore, by the assumption of the lemma, there must be a set $S \subset U$ such that $\partial_H(S) < k/20$ (we can also assume that $\phi_H(S) \geq \phi_H(U \setminus S)$, otherwise just take the other side). We now show that this cannot happen.

Note that H is 7/20-dense in G, so for each $v \in U$,
$$d_H(v) \geq 7d_G(v)/20 \geq 7k/20, \tag{6.2}$$
where we used the k-edge-connectivity of G.

We start with a natural decomposition of the induced graph $G[S]$ into $k/20$-edge-connected subgraphs, $S_1, \ldots, S_\ell$, as defined in definition 2.13. We show that for each i, $\partial_H(S_i) \geq k/10$. This already gives a contradiction, because by lemma 2.14
$$\begin{aligned}\frac{k}{20} + 2(\ell-1)\frac{k}{20} &> \partial_H(S) + \sum_{i=1}^{\ell} \partial_{G[S]}(S_i) \\ &= \sum_{i=1}^{\ell} \partial_H(S_i) \geq \ell \cdot \frac{k}{10}.\end{aligned} \tag{6.3}$$

It remains to show that $\partial_H(S_i) \geq k/10$. For the sake of the contradiction, suppose that $\partial_H(S_i) < k/10$ for some i. First, observe that S_i cannot be a singleton, because the induced degree of each vertex of H is at least $7k/20 > k/10$. We reach a contradiction by showing that $G[S_i]$ is a 1/4-dense, $k/20$-edge-connected induced subgraph of G with expansion $\phi(G[S_i]) \geq \phi^*$. By definition, $G[S_i]$ is $k/20$-edge-connected. Next, we show $G[S_i]$ is dense. For every vertex $v \in S_i$,

$$d_{G[S_i]}(v) \geq d_H(v) - \partial_H(S_i) \geq d_H(v) - k/10 \geq \frac{7d_G(v)}{20} - \frac{d_G(v)}{10} \geq d_G(v)/4,$$

where the third inequality uses (6.2). Therefore $G[S_i]$ is 1/4-dense.

Finally, we show that $G[S_i]$ is a ϕ^*-expander. This is because for any set $T \subseteq S_i$,

$$\phi_{G[S_i]}(T) \geq \frac{\partial_{G[S_i]}(T)}{d_H(T)} \geq \frac{k/20}{d_H(T)} \geq \frac{\partial_H(S)}{d_H(S)} = \phi_H(S) \geq \phi^*.$$

Therefore, $G[S_i]$ is a $k/20$-edge-connected, 1/4-dense and ϕ^*-expander, which is a contradiction. So, $\partial_H(S_i) \geq k/10$, which gives a contradiction by (6.3). □

This completes the proof of proposition 6.3. We are now ready to prove theorem 6.1. The details of our construction are given in algorithm 2.

Algorithm 2 Construction of an locally connected hierarchy for a $7\log(n)$-edge-connected graph.

Input: A k-edge-connected graph $G = (V, E)$ where $k \geq 7\log(n)$.
Output: A $(1/k, .)$-expanding $(k/20, 1/4, \mathcal{T})$-LCH $\mathcal{T}$ of G.

1: For each vertex $v \in V$, add a unique (leaf) node to $\mathcal{T}$ and map v to it. Let W be the set of these leaf nodes. ▷ Throughout the algorithm, we keep the invariant that W consists of the nodes of $\mathcal{T}$ that do not have a parent yet, but their corresponding subtree is fixed, i.e., $V(t)$ is well-defined for any $t \in W$.
2: **while** $|W| > 1$ **do**
3: Add a new node t^* to W.
4: Let G_{t^*} be the graph where for each node $t \in W$, $V(t)$ is contracted to a single vertex, and identify t with the corresponding contracted vertex. ▷ G_{t^*} is k-edge-connected by Fact 2.12.
5: Let $H_{t^*} = G_{t^*}[U_{t^*}]$ be the $k/20$-edge-connected, 1/4-dense $1/k$-expanding induced subgraph of G_{t^*} promised by proposition 6.3.
6: Let $W = W \setminus U_{t^*}$, and make t^* the parent of all nodes in U_{t^*}. ▷ So, $V(t^*) = \cup_{t \in U_{t^*}} V(t)$ and $G\{t^*\} = H_{t^*}$.
7: **end while**
return $\mathcal{T}$.

First of all, observe that the algorithm always terminates in at most $n-1$ iterations of the loop, because in each iteration $|W|$ decreases by at least 1. The properties of H_{t^*} in step 5 translate to the properties of $\mathcal{T}$ as follows:

- $1/k$-expansion of H_{t^*} guarantees that $\mathcal{T}$ is $(1/k, .)$-expanding.
- The $k/20$-edge-connectivity of H_{t^*} implies that $\mathcal{T}$ is $(k/20, ., .)$-LCH.

- Finally, the fact that H_{t^*} is 1/4-dense with respect to G_{t^*} implies that $\mathcal{T}$ is $(., 1/4, \mathcal{T})$-LCH.

6.3 Extraction of an $\Omega(k)$-Edge-Connected Set of Low-Effective-Resistance Edges

In this part we prove the following theorem.

Theorem 6.5. *If $G = (V, E)$ has an $(\alpha, .)$-expanding $(k, \lambda, \mathcal{T})$-LCH, then there exists a PD shortcut matrix D, and a $k/4$-edge-connected set F of good edges such that*

$$\max_{e \in F} \mathcal{R}\text{eff}_D(e) \leq \frac{f_2(k, \lambda, \alpha)}{k},$$

where $f_2(k, \lambda, \alpha) = f_1(k, \lambda\alpha) \cdot O(\log(1/\alpha))$.

The main theorem of the dissertation , theorem 3.6, follows from the above theorem together with theorem 6.1.

Let $\mathcal{T}$ be the $(\alpha, .)$-expanding $(k, \lambda, \mathcal{T})$-LCH given to us. First, observe that it is very easy to prove a weaker version of the above theorem where

$$\mathcal{R}\text{eff}_D(e) \leq \frac{2f_1(k, \lambda)}{k \cdot \alpha}$$

for edges of F by a single application of theorem 4.2. Let D be the optimum of Tree-CP($\mathcal{T}$); we let $F \subseteq E$ be the edges where $\mathcal{R}\text{eff}_D(e) \leq \frac{2f_1(k,\lambda)}{k \cdot \alpha}$. Let $G' = (V, F)$. It follows that for any node t of $\mathcal{T}$, $G'\{t\}$ is $k/2$-edge-connected, so by lemma 3.11 G' is $k/2$-edge-connected and we are done.

The main difficulty in proving the above theorem is to reduce the inverse polynomial dependency on α in the above argument to a polylogarithmic function of α. To achieve that, we apply theorem 4.2 to $\log(1/\alpha)$ locally connected hierarchies, $\mathcal{T}_0, \ldots, \mathcal{T}_{\log(1/\alpha)}$, of our graph. For each $\mathcal{T}_i$, W_i is the set of bad internal nodes of W_{i-1}, i.e., those where their internal subgraph is not yet $\Omega(k)$-edge-connected with respect to the good edges found so far. Originally, W_0 contains all internal nodes of $\mathcal{T}_0$ and it is a $(1/k, .)$-expanding set. For each i, we will make sure that $\mathcal{T}_i$ is $(k/4, \lambda\alpha^i, T_i)$-LCH and W_i is $(2^i\alpha, k)$-expanding. In other words, each $\mathcal{T}_i$ is a "refinement" of $\mathcal{T}_{i-1}$ whose W_i nodes are twice more expanding.

Throughout the algorithm we also make sure that all (except possibly one) children of each node in W_i are in T_i. Let us elaborate on this statement. Let $t \in W_i$ and let $t_0, t_1, \ldots$ be the children of t. Since $W_i \subseteq W_0$, $t \in W_0$. Consider the graph $G_{\mathcal{T}_0}\{t\}$; by the theorem's assumptions $G_{\mathcal{T}_0}\{t\}$ is a k-edge-connected α-expander. It follows that $G_{\mathcal{T}_i}\{t\}$ can be obtained from $G_{\mathcal{T}_0}\{t\}$ by contracting a set $U_{t_j} \subset V_{\mathcal{T}_0}\{t\}$ corresponding to each children t_j of t. We use the notation

$$d_0(t_j) = \sum_{t' \in U_{t_j}} d_{G_{\mathcal{T}_0}\{t\}}(t')$$

Algorithm 3 Extracting Small Effective Resistance Edges

Input: A graph $G = (V, E)$ and a $(\alpha, .)$-expanding $(k, \lambda, \mathcal{T})$-LCH $\mathcal{T}$.

Output: A PD shortcut matrix D and a $k/4$-edge-connected set F of good edges.

Let W_0 be all internal nodes of $\mathcal{T}$, $W_i = \emptyset$ for $i > 0$, and $\mathcal{T}_0 = \mathcal{T}$, and $G' = (V, \emptyset)$.

for $i = 0 \to \log(1/\alpha)$ **do**

  Let D_i be the optimum of Tree-CP($\mathcal{T}_i$).

  Say $\mathcal{T}_i$ is a (k', λ', T_i)-LCH of G; let

$$F_i := \left\{ e \in E : \mathcal{R}\mathrm{eff}_{D_i}(e) \leq \frac{16 f_1(k', \lambda')}{k'} \right\}, \tag{6.4}$$

  add all edges of F_i to G'.

  For any node $t \in W_i$, let $S_{t,1}, \ldots, S_{t,\ell(t)}$ be a natural decomposition of $G'_{\mathcal{T}_i}\{t\}$ into $k/4$-edge-connected components as defined in definition 2.13. If $\ell(t) > 1$, then we add t to W_{i+1}. ▷ Note that if $\ell(t) = 1$ it means that $G'\{t\}$ is $k/4$-edge-connected.

  We construct a $(., ., T_{i+1})$-LCH of G, called $\mathcal{T}_{i+1}$, by modifying $\mathcal{T}_i$. For any node $t \in W_{i+1}$ we add $\ell(t)$ new nodes $s_{t,1}, \ldots, s_{t,\ell(t)}$ to $\mathcal{T}_{i+1}$ and we make all nodes of $S_{t,j}$ children of $s_{t,j}$ and we make t the parent of $s_{t,j}$. Therefore, t has exactly $\ell(t)$ children in $\mathcal{T}_{i+1}$. See fig. 6.1 for an example. The set T_{i+1} is the union of all nondominating nodes children of all nodes of W_i.

end for

return the PD shortcut matrix $\mathbb{E}_i D_i$ and the good edges $\cup_i F_i$.

to denote the sum of the degrees of nodes in S_{t_j} in the noncontracted graph $G_{\mathcal{T}_0}\{t\}$. We say a child t_ℓ of t is *dominating* if

$$d_0(t_\ell) > \frac{1}{2} \sum_j d_0(t_j).$$

It follows that each node $t \in W_i$ can have at most one dominating child. In addition, if t_ℓ is a dominating child, it may not satisfy $\mathcal{O}(t_\ell) \gtrsim \mathcal{P}(t_\ell)$, so we may not add t_ℓ to T_i. Because of this we need to treat the dominating children (of nodes of W_i) differently throughout the algorithm and the proof. In our construction T_i consists of all nondominating children of all nodes of W_i. It is easy to see that for any nondominating child t_ℓ of $t \in W_i$,

$$\mathcal{O}_{\mathcal{T}_i}(t_\ell) = \partial_{G_{\mathcal{T}_0}\{t\}}(U_{t_\ell}) \geq \alpha \cdot d_{G_{\mathcal{T}_0}\{t\}}(U_{t_\ell}) = \alpha \cdot \sum_{t' \in U_{t_\ell}} \mathcal{O}_{\mathcal{T}_0}(t') \geq \alpha \cdot \lambda \cdot \sum_{t' \in U_{t_\ell}} \mathcal{P}_{\mathcal{T}_0}(t') \geq \alpha \lambda \mathcal{P}_{\mathcal{T}_i}(t_\ell),$$

where the first inequality uses the fact that $G_{\mathcal{T}_0}\{t\}$ is an α-expander and the second inequality uses the fact that $\mathcal{T}_0$ is $(., \lambda, \mathcal{T}_0)$-LCH. The following claim is immediate

Claim 6.6. *If $\mathcal{T}_0$ is an $(\alpha, .)$-expanding $(., \lambda, \mathcal{T}_0)$-LCH, then for any $i \geq 1$, $\mathcal{T}_i$ is a $(., \lambda\alpha, T_i)$-LCH, where T_i consists of all nondominating children of the nodes of W_i, .*

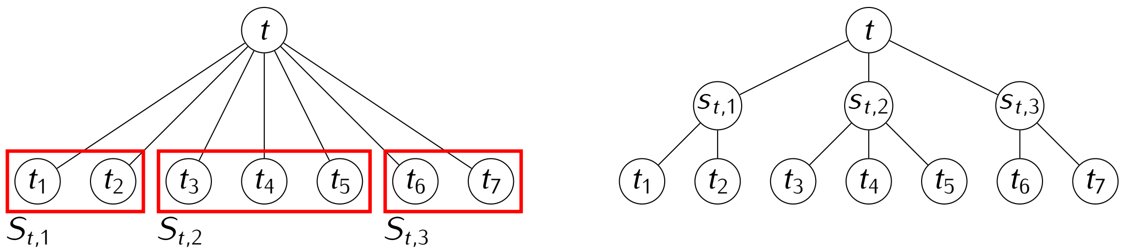

Figure 6.1: A node t and its children, $t_1, t_2, \ldots$, in $\mathcal{T}_{i-1}$ are illustrated in left. The right diagram shows the tree $\mathcal{T}_i$ when the new nodes $s_{t,1}, s_{t,2}, s_{t,3}$ corresponding to the sets $S_{t,1}, S_{t,2}, S_{t,3}$ are added.

At the end of the algorithm, we obtain PD shortcut matrices $D_0, \ldots, D_{\log(1/\alpha)}$ and sets $F_0, \ldots, F_{\log(1/\alpha)}$ such that the edges of each F_i have small effective resistance with respect to D_i, and $\cup_{i=0}^{\log(1/\alpha)} F_i$ is $\Omega(k)$-edge-connected. Then, we let D be the average of $D_0, \ldots, D_{\log(1/\alpha)}$ and F be the union of $F_0, \ldots, F_{\log(1/\alpha)}$. The details of the construction of these matrices and sets are given in algorithm 3.

We prove the claim by induction on i. In the first step we show $\mathcal{T}_{i+1}$ is a $(k/4, ., .)$-LCH. Then, we show that W_{i+1} is $(2^{i+1}\alpha, k)$-expanding. Then, we show that $W_{\log(1/\alpha)}$ is empty and we conclude by showing that $G' = (V, \cup_i F_i)$ is $\Omega(k)$-edge-connected.

Claim 6.7. *If $\mathcal{T}_i$ is a $(k/4, ., .)$-LCH of G, then $\mathcal{T}_{i+1}$ is a $(k/4, ., .)$-LCH of G. In addition, if W_i is $(., k)$-expanding, then W_{i+1} is $(., k)$-expanding.*

Proof. First, for any node $t \in \mathcal{T}_{i+1}$ that is also in $\mathcal{T}_i$, $G_{\mathcal{T}_{i+1}}(t) = G_{\mathcal{T}_i}(t)$; so, $G_{\mathcal{T}_{i+1}}(t)$ is $k/4$-edge-connected by induction. So, $G_{\mathcal{T}_{i+1}}\{t\}$ is also $k/4$-edge-connected. For any new node $s_{t,j} \in \mathcal{T}_{i+1}$, since $S_{t,j}$ is a $k/4$-edge-connected subgraph of $G'_{\mathcal{T}_i}\{t\}$, $G_{\mathcal{T}_{i+1}}(s_{t,j})$ is $k/4$-edge-connected. Therefore, $\mathcal{T}_{i+1}$ is a $(k/4, ., .)$-LCH of G.

Similarly, observe that W_{i+1} is $(., k)$-expanding, because $W_{i+1} \subseteq W_i$ and for any node $t \in \mathcal{T}_i$, $G_{\mathcal{T}_{i+1}}(t) = G_{\mathcal{T}_i}(t)$. □

We slightly strengthen our induction; instead of showing that $G_{\mathcal{T}_i}\{t\}$ is $(2^i\alpha, .)$-expanding for all $t \in W_i$, we show that for any $t \in W_i$ and any $S \subseteq V_{\mathcal{T}_i}\{t\}$ where $d_0(S) \leq \frac{1}{2} d_0(V_{\mathcal{T}_i}\{t\})$,

$$\phi_{G_{\mathcal{T}_i}\{t\}}(S) \geq 2^i \alpha.$$

For a set of indices $I \subseteq [\ell]$ we use $S_I = \cup_{i \in I} S_i$. The following is the key lemma of the proof of this section.

Claim 6.8. *For any $i \geq 0$, $t \in W_i$, and any $S \subseteq V_{\mathcal{T}_i}\{t\}$ where $d_0(S) \leq \frac{1}{2} d_0(V_{\mathcal{T}_i}\{t\})$,*

$$\phi_{G_{\mathcal{T}_i}\{t\}}(S) \geq \min\{2^i \alpha, 1/8\}.$$

Therefore, for any $i \geq 1$, W_i is $(2^i\alpha, .)$-expanding.

Proof. We prove this by induction. Note that the statement obviously holds for $i = 0$ because $G_{\mathcal{T}_0}\{t\}$ is an α-expander for all $t \in \mathcal{T}_0$. Suppose the statement holds for i. Fix a node $t \in W_{i+1}$ and let $S_{t,1}, \ldots, S_{t,\ell(t)}$ be the natural decomposition of $G_{\mathcal{T}_i}\{t\}$ into $k/4$-edge-connected components. We abuse notation and drop the subscript t and name these sets $S_1, \ldots, S_{\ell(t)}$. Choose $I \subset [\ell(t)]$ such that $d_0(S_I) \leq \frac{1}{2} d_0(V_{\mathcal{T}_i}\{t\})$. If $\phi_{G_{\mathcal{T}_i}\{t\}}(S_I) \geq 1/8$ there is nothing to prove. Otherwise, we invoke lemma 6.9 for the k-edge-connected graph $G = G_{\mathcal{T}_i}\{t\}$, $F = \cup_{j=1}^{i} F_j$ and the natural decomposition $S_1, \ldots, S_{\ell(t)}$ of $(V_{\mathcal{T}_i}\{t\}, F)$ into $k/4$-edge-connected components. The lemma shows that $\phi_{G_{\mathcal{T}_{i+1}}\{t\}}(S_I) \geq 2^{i+1}\alpha$.

We just need to verify the assumptions of the lemma. By the induction hypothesis $\phi_{G_{\mathcal{T}_i}\{t\}}(S_I) \geq 2^i\alpha$. In addition, S_I only contains nondominating nodes of t, i.e., $S_I \subset T_i$. Therefore, by the main technical theorem 4.2, equation (6.4), and the Markov inequality, at least 15/16 fraction of the edges incident to each $t' \in T_i$ are in F_i. So, $\partial_{F_i}(S_I) \geq \frac{15}{16} d(S_I) \geq \frac{7}{8} d(S_I)$. □

Lemma 6.9 (Expansion Boosting Lemma). *Given a k-edge-connected graph $G = (V, E)$, a set $F \subseteq E$ and a natural-decomposition of (V, F) into $k/4$-edge-connected components $S_1, \ldots, S_\ell$. For any $I \subseteq [\ell]$ if $d_F(S_I) \geq 7d(S_I)/8$, and $\phi(S_I) < 1/8$, then*

$$\frac{\partial(S_I)}{\sum_{i \in I} \partial(S_i)} \geq 2\phi(S_I).$$

Proof. Think of the edges in F as good edges and the edges not in F, $E \setminus F$ as the bad edges. We can write the denominator of the above as follows:

$$\sum_{i \in I} \partial(S_i) = \partial_F(S_I) + 2 \sum_{i,j \in I, i<j} |F(S_i, S_j)| + \sum_{i \in I} \partial_{E \setminus F}(S_i) \tag{6.5}$$

where we used $\partial_F(S)$ to denote the edges of F leaving a set S.

First, we observe that by the natural decomposition lemma 2.14, the middle term on the RHS, i.e., the number of good edges between $\{S_i\}_{i \in I}$ is small,

$$\sum_{i,j \in I, i<j} |F(S_i, S_j)| \leq (|I| - 1)(k/4) \leq \frac{1}{4} \sum_{i \in I} \partial(S_i),$$

where the second inequality follows by k-edge-connectivity of G. Subtracting twice the above inequality from (6.5) we get

$$\partial_F(S_I) + \sum_{i \in I} \partial_{E \setminus F}(S_i) \geq \frac{1}{2} \sum_{i \in I} \partial(S_i). \tag{6.6}$$

Secondly, by the lemma's assumption,

$$\sum_{i \in I} \partial_{E \setminus F}(S_i) \leq \sum_{i \in I} d_{E \setminus F}(S_i) = d_{E \setminus F}(S_I) = d(S_I) - d_F(S_I) \leq \frac{1}{8} d(S_I). \tag{6.7}$$

Putting the above two inequalities together we get,

$$\frac{1}{2}\sum_{i\in I}\partial(S_i) \leq \partial_F(S_I) + \frac{1}{8}d(S_I)$$

Dividing both sides of the above inequality by $\partial(S_I)$ we get

$$\begin{aligned}\frac{1}{2\partial(S_I)}\sum_{i\in I}\partial(S_i) &\leq \frac{\partial_F(S_I)}{\partial(S_I)} + \frac{d(S_I)}{8\partial(S_I)}\\ &\leq 1 + \frac{1}{8\phi(S_I)} \leq \frac{1}{4\phi(S_I)},\end{aligned}$$

where the last inequality uses that $\alpha \leq 1/8$. □

Claim 6.10. *$W_{\log(1/\alpha)}$ is empty.*

Proof. Let i be the smallest integer such that $2^i\alpha \geq 1/8$. Note that $i < \log(1/\alpha)$. By Claim 6.8, for any $t \in W_i$,

$$\phi(G_{\mathcal{T}_i}\{t\}) \geq 1/8. \tag{6.8}$$

We show that W_{i+1} is empty. Fix a node $t \in W_i$. Similar to the previous claim, at least 15/16 fraction of the edges adjacent to any nondominating child of t are in F_i. For a set $I \subset [\ell(t)]$ such that $d_0(S_I) \leq \frac{1}{2}d_0(V_{\mathcal{T}_0}\{t\})$, we have $\phi_{G_{\mathcal{T}_i}\{t\}}(S_I) \geq 1/8$; therefore at least half of the edges in the cut $(S_I, V_{\mathcal{T}_i}\{t\} \setminus S_I)$ are in F_i. By k-edge-connectivity of $G_{\mathcal{T}_i}\{t\}$, F_i has at least $k/2$ edges in this cut. So, $(V_{\mathcal{T}_i}\{t\}, F_i)$ is $k/2$-edge-connected. □

Claim 6.11. *At the end of the algorithm G' is $k/4$-edge-connected.*

Proof. We show that for any i and any node $t \notin W_i$, $G'_{\mathcal{T}_i}\{t\}$ is $k/4$-edge-connected. Then, the claim follows by Claim 6.10.

At any iteration i, for any new node $s_{t,j}$, $G'_{\mathcal{T}_i}\{s_{t,j}\}$ is $k/4$-edge-connected because $S_{t,j}$ is a $k/4$-edge-connected component of $G_{\mathcal{T}_i}\{t\}$; this subgraph remains $k/4$-edge-connected in the rest of the algorithm because we never delete edges from G'. On the other hand, when we remove a node t from W_i, we are guaranteed that $G_{\mathcal{T}_i}\{t\}$ is $k/4$-edge-connected. □

Now, theorem 6.5 follows from the above claim and that for any $e \in \cup_i F_i$,

$$\mathcal{R}\mathrm{eff}_{\mathbb{E}_i D_i}(e) \leq \log(1/\alpha)\cdot\min_i \mathcal{R}\mathrm{eff}_{D_i}(e) \leq \frac{16 f_1(k/4, \lambda\cdot\alpha)\log(1/\alpha)}{k/4}.$$

Chapter 7

Effective Resistance Reduction via Spectral Flows

In this chapter we analyze the effective resistance reducing convex programs by analyzing their duals.

7.1 The Dual of Tree-CP

In this section we write down the dual of Tree-CP. Before explicitly writing down the dual, let us give a few lines of intuition. We do this by writing down the dual of a few convex programs computing the maximum or average effective resistance of a number of pairs of vertices.

For a pair of vertices, $a, b \in V$, the optimum value of the following expression,

$$\max_{x:V\to\mathbb{R}} \frac{(x(a)-x(b))^2}{\sum_{u\sim v}(x(u)-x(v))^2}. \tag{7.1}$$

is exactly equal to $\mathcal{R}\mathrm{eff}_G(a,b)$; in particular, if we fix $x(b)=0, x(a)=\mathcal{R}\mathrm{eff}(a,b)$, then the optimum x is the *potential* vector of the electrical flow that sends one unit of flow from a to b. It is an easy exercise to cast the above as a convex program.

Now, suppose we want to write a program which computes the maximum effective resistance of pairs of vertices $(a_1,b_1),\dots,(a_h,b_h)$. In this case we need to choose a separate potential vector for each pair, We use a matrix X where the i-th row of X is the potential vector associated to the i-th pair. The following program gives the maximum effective resistance of all pairs.

$$\max_{X\in\mathbb{R}^{h\times V}} \frac{\sum_{i=1}^h (X_{i,a_i}-X_{i,b_i})^2}{\sum_{i=1}^h\sum_{u\sim v}(X_{i,u}-X_{i,v})^2} = \max_{X\in\mathbb{R}^{h\times V}} \frac{\sum_{i=1}^h (X_{i,a_i}-X_{i,b_i})^2}{\sum_{u\sim v}(X_u-X_v)^2}$$

It follows by (2.1) that the optimum of the above is the maximum effective resistance of all pairs $(a_1,b_1),\dots,(a_h,b_h)$. Recall that X_u is the u-th column of X.

Note that the denominator of the RHS is coordinate independent, i.e., it is rotationally invariant. We can rewrite the numerator in the following way and make it rotationally invariant.

Instead of mapping the i-th pair to the i-th coordinate, we map the i-th pair to z_i where $\{z_1, \ldots, z_h\}$ are h-orthonormal vectors. In other words, to calculate the numerator we need to find a coordinate system of the space such that the sum of the square of the projection of the edges on the corresponding coordinates is as large as possible

$$\max_{\substack{X \in \mathbb{R}^{h \times V}, \\ \{z_1,\ldots,z_h\} \text{ are orthonormal}}} \frac{\sum_{i=1}^h \langle z_i, X_{a_i} - X_{b_i} \rangle^2}{\sum_{u \sim v} (X_u - X_v)^2}.$$

Instead of choosing $z_1, \ldots, z_h$ we can simply maximize over an orthogonal matrix $U \in \mathbb{R}^{h \times h}$ and let $z_1, \ldots, z_h$ be the first h rows of U,

$$\max_{X \in \mathbb{R}^{h \times V}, \text{Orthogonal } U} \frac{\sum_{i=1}^h \langle U^i, X_{a_i} - X_{b_i} \rangle^2}{\sum_{u \sim v} (X_u - X_v)^2}, \tag{7.2}$$

where U^i is the i-th row of the matrix U. The above program is equivalent to the dual of the following convex program

$$\begin{aligned} \min \quad & \mathcal{E}, \\ \text{s.t.} \quad & \mathcal{R}\text{eff}_D(a_i, b_i) \le \mathcal{E} \quad \forall 1 \le i \le h, \\ & D \preceq L_G. \end{aligned}$$

We will give a formal argument later. When we replace the constraint $D \preceq L_G$ with $D \preceq_\square L_G$, we get the additional assumption that X is a cut metric. This can significantly reduce the value of (7.2).

Next, we write a program which computes the expected effective resistance of pairs of vertices $(a_1, b_1), \ldots, (a_h, b_h)$ with respect to a distribution $\lambda_1, \ldots, \lambda_h$,

$$\sum_{i=1}^h \lambda_i \cdot \mathcal{R}\text{eff}(a_i, b_i) = \max_{X \in \mathbb{R}^{h \times V}} \sum_{i=1}^h \lambda_i \cdot \frac{(X_{i,a_i} - X_{i,b_i})^2}{\sum_{u \sim v} (X_{i,u} - X_{i,v})^2}. \tag{7.3}$$

where we simply used (7.1). Equivalently, we can write the above ratio as follows:

$$\max_{X \in \mathbb{R}^{h \times V}} \frac{\left(\sum_{i=1}^h \sqrt{\lambda_i} \cdot (X_{i,a_i} - X_{i,b_i}) \right)^2}{\sum_{u \sim v} (X_u - X_v)^2}, \tag{7.4}$$

To see that the above two are the same, first, assume X is normalized such that $\sum_{u \sim v} (X_{i,a_i} - X_{i,b_i})^2 = 1$ for all i. This simplifies (7.3) to $\sum_i \lambda_i (X_{i,a_i} - (X_i, b_i))^2$. Then let

$$Y^i = X^i \sqrt{\lambda_i} \cdot (X_{i,a_i} - X_{i,b_i}),$$

where as usual Y^i is the i-th row of Y. Plugging in Y in (7.4) gives the same value $\sum_i \lambda_i (X_{i,a_i} - X_{i,b_i})^2$.

Lastly, we can write a rotationally invariant formulation of (7.4) using an orthogonal matrix U.

$$\max_{\substack{X\in\mathbb{R}^{h\times V},\\ \text{Orthogonal } U}} \frac{\left(\sum_{i=1}^{h}\sqrt{\lambda_i}\cdot\langle U^i, X_{a_i}-X_{b_i}\rangle\right)^2}{\sum_{u\sim v}(X_u-X_v)^2}$$

Let $\chi_h \in \mathbb{R}^{n\times h}$ be the matrix where the i-th column is χ_{a_i,b_i}. It follows by lemma 2.6 that

$$\max_{\text{Orthogonal } U}\sum_{i=1}^{h}\langle U^i, X_{a_i}-X_{b_i}\rangle = \max_{\text{Orthogonal } U}\mathrm{Tr}(UX\chi_h) = \left\|X\chi_h\right\|_* .$$

This is is a key observation in the proof of the technical theorem.

In the rest of this section we will prove that a similar expression is equivalent to the dual of Tree-CP. Then, in section 7.1 we write the dual of Max-CP, Average-CP and we will prove theorem 7.3. The following lemma is the main statement that we prove in this section. Recall that for a mapping X of vertices of G, $\mathsf{X} = X\chi$ is the matrix where for every edge $e=\{u,v\}$, $\mathsf{X}_e = X_u - X_v$.

Lemma 7.1. *For any graph $G=(V,E)$ and any $(.,.,T)$-LCH of G, the optimum of* Tree-CP *(up to a multiplicative factor of 2) is equal to*

$$\sup_{U,X}\frac{\sum_{t\in T}\frac{1}{|\mathcal{O}(t)|}\left(\sum_{e\in\mathcal{O}(t)}\langle U^e, \mathsf{X}_e\rangle\right)^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2} \tag{7.5}$$

where the supremum is over all semiorthogonal matrices $U\in\mathbb{R}^{E\times h}$, and all cut metrics $X\in\{0,1\}^{h\times V}$, for arbitrary $h>0$.

Note that the dimension h in the above can be arbitrarily large because X is a cut metric. However, only the first $|E|$ rows of U matter. In addition, since X is a cut metric, for any edge $e=\{u,v\}\in E$, $\left\|\mathsf{X}_e\right\|^2 = \left\|\mathsf{X}_e\right\|_1$; so, throughout the dissertation, we may use either of the two norms.

Proof. First, we show Tree-CP satisfies Slater's condition, i.e., that Tree-CP has a nonempty interior. It is easy to see that $D=\frac{1}{2}L_G+\frac{1}{3n^2}J$ is a PD matrix that satisfies all constraints strictly. In particular, since G is connected, for any set S, $\mathbf{1}_S^\intercal L_G\mathbf{1}_S\geq 1$, so

$$\frac{1}{3n^2}\mathbf{1}_S J\mathbf{1}_S \leq \frac{1}{3} < \frac{1}{2}\mathbf{1}_S^\intercal L_G\mathbf{1}_S.$$

Therefore, $\mathbf{1}_S^\intercal D\mathbf{1}_S < \mathbf{1}_S^\intercal L_G\mathbf{1}_S$ for all S. Hence, Slater's condition is satisfied, and the strong duality is satisfied and the primal optimum is equal to the Lagrangian dual's optimum (see [BV06, Section 5.2.3] for more information).

For every $t\in T$ we associate a Lagrange multiplier λ_t corresponding to the first set of constraints, and for every set S we associate a nonnegative Lagrange multiplier y_S

corresponding to the second set of constraints of the Tree-CP. The Lagrange function is defined as follows:

$$g(\lambda, y) = \inf_{D \succ 0} \mathcal{E} + \sum_{t \in T} \lambda_t \Big(\frac{1}{|\mathcal{O}(t)|} \sum_{e \in \mathcal{O}(t)} \chi_e^\intercal D^{-1} \chi_e - \mathcal{E} \Big) + \sum_{S \subset V} y_S (\mathbf{1}_S^\intercal D \mathbf{1}_S - \mathbf{1}_S^\intercal L_G \mathbf{1}_S)$$

First, we differentiate the RHS with respect to $\mathcal{E}, D$ to eliminate the inf. This gives us the Lagrangian dual. Then, we homogenize the dual expression by normalizing the entries of y; finally we eliminate the dependency on λ by an application of the Cauchy-Schwarz inequality.

First of all, differentiating $g(\lambda, y)$ w.r.t. $\mathcal{E}$ we obtain that

$$\sum_{t \in T} \lambda_t = 1. \tag{7.6}$$

Let

$$A := \sum_{t \in T} \frac{\lambda_t}{|\mathcal{O}(t)|} \Big(\sum_{e \in \mathcal{O}(t)} \chi_e \chi_e^\intercal \Big) \text{ and } Z := \sum_{\emptyset \subset S \subset V} y_S \mathbf{1}_S \mathbf{1}_S^\intercal.$$

Note that by definition A and Z are symmetric PSD matrices. The Lagrange dual function simplifies to

$$g(A, Z) = \inf_{D \succ 0} A \bullet D^{-1} + Z \bullet D - Z \bullet L_G,$$

subject to $\sum_t \lambda_t = 1$. Now, we find the optimum D for fixed A, Z. First, we assume that A and Z are nonsingular. This is without loss of generality by the continuity of $g(.)$ and because the assumption $\sum_t \lambda_t = 1$ can be satisfied by adding arbitrarily small perturbations. Differentiating with respect to D we obtain

$$D^{-1} A D^{-1} = Z.$$

Since, A, D are nonsingular there is a unique solution to the above equation,

$$D = Z^{-1/2} (Z^{1/2} A Z^{1/2})^{1/2} Z^{-1/2}$$

We refer interested readers to [SLB74] to solve the above matrix equation. Using

$$D^{-1} = Z^{1/2} (Z^{1/2} A Z^{1/2})^{-1/2} Z^{1/2},$$

we have

$$\begin{aligned} A \bullet D^{-1} + Z \bullet D &= \mathrm{Tr}(A Z^{1/2} (Z^{1/2} A Z^{1/2})^{-1/2} Z^{1/2}) + \mathrm{Tr}(Z^{1/2} (Z^{1/2} A Z^{1/2})^{1/2} Z^{-1/2}) \\ &= 2 \,\mathrm{Tr}((Z^{1/2} A Z^{1/2})^{1/2}). \end{aligned}$$

Therefore,

$$g(A, Z) = 2 \,\mathrm{Tr}((Z^{1/2} A Z^{1/2})^{1/2}) - Z \bullet L_G$$

Let $\mathcal{E}^*$ be the optimum value of Tree-CP. By the strong duality,

$$\mathcal{E}^* = \sup_{\lambda,y\geq 0} g(A,Z) = \sup_{\lambda,y\geq 0} 2\operatorname{Tr}((Z^{1/2}AZ^{1/2})^{1/2}) - Z \bullet L_G.$$

It remains to characterize values of λ, y that maximize the above function. Let $W \in \mathbb{R}^{E\times E}$ be a diagonal matrix where for each edge $e \in E$,

$$W_{e,e} = \sqrt{\sum_{t\in T: e\in\mathcal{O}(t)} \frac{\lambda_t}{|\mathcal{O}(t)|}}. \tag{7.7}$$

Note that the above sum is over zero, one, or two terms because each edge is in at most two sets $\mathcal{O}(t)$. Observe that

$$A = \chi W^2 \chi^\intercal.$$

Furthermore the nonzero eigenvalues of $Z^{1/2}AZ^{1/2} = Z^{1/2}\chi W^2\chi^\intercal Z^{1/2}$ are the same as the nonzero eigenvalues of $W\chi^\intercal Z\chi W$. Therefore,

$$\mathcal{E}^* = \sup_{\lambda,y\geq 0} 2\operatorname{Tr}((W\chi^\intercal Z\chi W)^{1/2}) - Z \bullet L_G \tag{7.8}$$

Observe that the above quantity is not homogeneous in y as $Z \bullet L_G$ scales linearly with y and $\operatorname{Tr}((W\chi^\intercal Z\chi W)^{1/2})$ scales with $\sqrt{y}$. It is an easy exercise to see that by choosing the right scaling for y we can rewrite the above as follows:

$$\mathcal{E}^* = \sup_{\lambda,y\geq 0} \frac{\operatorname{Tr}((W\chi^\intercal Z\chi W)^{1/2})^2}{Z \bullet L_G}.$$

Note that although (7.8) is convex, the above quantity is not necessarily convex but we prefer to work with the above quantity because it is homogeneous.

Write $Z = X^\intercal X$ where $X \in \mathbb{R}^{2^n\times V}$ and each row of X corresponds to a vector $y_S\mathbf{1}_S$ for a set $S \subseteq V$. Observe that X defines a weighted cut metric on the vertices of G which can be embedded into an unweighted cut metric (see section 2.2 for properties of weighted/unweighted cut metrics). So, we assume $X \in \{0,1\}^{h\times V}$ for an h possibly larger than 2^n. If $h < |E|$ then we extend X by adding all zeros rows to make $h \geq |E|$. Let X_v be the mapping of v in that metric, i.e., X_v is the column v of X. By the definition of the nuclear norm,

$$\operatorname{Tr}((W\chi^\intercal Z\chi W)^{1/2})^2 = \left\|X\chi W\right\|_*^2 = \left\|\mathsf{X}W\right\|_*^2.$$

Therefore,

$$\mathcal{E}^* = \sup_{X,\lambda} \frac{\left\|\mathsf{X}W\right\|_*^2}{\sum_{\{u,v\}\in E}\left\|\mathsf{X}_e\right\|_2^2}$$

In the denominator we used the fact that $Z \bullet L_G = \sum_{\{u,v\}} \left\|X_u - X_v\right\|_2^2 = \sum_e \left\|\mathsf{X}_e\right\|^2$.

Note that $\mathsf{X} \in \mathbb{R}^{h\times E}$. Since the number of rows of X is at least the number of its columns, by lemma 2.6, we can rewrite the nuclear norm as $\sup_U \mathrm{Tr}(U\mathsf{X}W)$ over all semiorthogonal matrices $U \in \mathbb{R}^{E\times h}$, so

$$\begin{aligned}
\mathcal{E}^* &= \sup_{\substack{X\in\{0,1\}^h,\lambda\geq 0,\\ \text{Semiorthgonal } U}} \frac{\left(\sum_{t\in T}\sum_{e\in\mathcal{O}(t)} W_{e,e}\cdot\langle U^e, \mathsf{X}_e\rangle\right)^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2} \\
&\asymp \sup_{\substack{X\in\{0,1\}^h,\lambda\geq 0,\\ \text{Semiorthgonal } U}} \frac{\left(\sum_{t\in T}\sum_{e\in\mathcal{O}(t)} \sqrt{\lambda_t/|\mathcal{O}(t)|}\cdot\langle U^e, \mathsf{X}_e\rangle\right)^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2}
\end{aligned} \tag{7.9}$$

Note that the second equation is an equality up to a factor of 2 because each edge is contained in at most two sets $\mathcal{O}(t)$. In particular, by (7.7), for any edge e,

$$\frac{1}{\sqrt{2}}\sum_{t\in T:e\in\mathcal{O}(t)} \sqrt{\lambda_t/|\mathcal{O}(t)|} \leq W_{e,e} \leq \sum_{t\in T:e\in\mathcal{O}(t)} \sqrt{\lambda_t/|\mathcal{O}(t)|}.$$

Finally, using the Cauchy-Schwarz inequality we can write

$$\mathcal{E}^* \lesssim \sup_{X,U} \frac{\left(\sum_{t\in T}\lambda_t\right)\cdot\left(\sum_{t\in T}\frac{1}{|\mathcal{O}(t)|}\left(\sum_{e\in\mathcal{O}(t)}\langle U^e, \mathsf{X}_e\rangle\right)^2\right)}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2}$$

The above inequality is tight because in the worst case we can let

$$\lambda_t \propto \frac{1}{|\mathcal{O}(t)|}\left(\sum_{e\in\mathcal{O}(t)}\langle U^e, \mathsf{X}_e\rangle\right)^2,$$

such that $\sum_t \lambda_t = 1$. □

The Dual for Variants of the Problem

In the rest of this section we prove simple positive and negative results on the value of the dual. We will not use these results in the proof of the technical theorem; we present them to provide some intuition on how one can approach the dual.

First of all, using similar ideas as the proof of the above lemma, we can also write the dual of Max-CP and Average-CP. We write these quantities, without proof, as we do not need them in the proof of our main theorem. First, we write the dual of Max-CP.

$$\left\{\begin{array}{ll} \min & \max_e \mathcal{R}\mathrm{eff}_D(e), \\ \text{s.t.} & D \preceq_\square L_G \\ D \succ 0 & \end{array}\right\} = \sup_{\substack{X\in\{0,1\}^{h\times V}\\ \text{Semiorthogonal } U}} \frac{\sum_{e\in E}\langle U^e, \mathsf{X}_e\rangle^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2}. \tag{7.10}$$

Now, we write the dual of Average-CP.

$$\left\{\begin{array}{ll}\min & \max_{S\subset V} \mathbb{E}_{e\sim E(S,\overline{S})} \mathcal{R}\mathrm{eff}_D(e), \\ \text{s.t.} & D \preceq_\square L_G, \\ & D \succ 0\end{array}\right\} = \sup_{\substack{X\in\{0,1\}^{h\times V},\lambda \\ \text{Semiorthogonal } U}} \frac{\left(\sum_{e\in E} \sqrt{\gamma_e}\cdot\langle U^e, \mathsf{X}_e\rangle\right)^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2}, \tag{7.11}$$

where for any edge e, $\gamma_e = \sum_{S:e\in E(S,\overline{S})} \frac{\lambda_{(S,\overline{S})}}{|E(S,\overline{S})|}$ and $\lambda_{(S,\overline{S})}$ is a probability distribution on all cuts of G.

In the following lemma, we show that for any pair of vertices of a k-edge-connected graph there is a shortcut matrix that reduces the effective resistance of that pair to $1/k$.

Lemma 7.2. *For any k-edge-connected graph G and any pair of vertices a, b, there is a shortcut matrix D such that $\mathcal{R}\mathrm{eff}_D(a, b) \le 1/k$.*

Proof. The statement can be proven relatively easy in the primal. Since G is k-edge-connected we can simply shortcut the k edge-disjoint paths connecting a, b and $D = k\cdot L_{a,b}$. Then it is easy to see that $\mathcal{R}\mathrm{eff}_D(a, b) = 1/k$ and $D \preceq_\square L_G$ as desired.

By (7.10) it is enough to show that

$$\sup_{\substack{X\in\{0,1\}^{h\times V}, \\ \text{Semiorthogonal } U\in\mathbb{R}^{1\times h}}} \frac{\langle U^{\{a,b\}}, X_a - X_b\rangle^2}{\sum_{u\sim v}\left\|X_u - X_v\right\|^2} \le O(1/k),$$

First note that in the worst case the vector U^e is parallel to $X_a - X_b$. Therefore, the numerator is exactly $\left\|X_a - X_b\right\|^2$. The proof simply follows from the triangle inequality of the cut metrics.

Since G is k-edge-connected there are k edge-disjoint paths from a to b. For any such path P we have

$$\sum_{e\in P}\left\|\mathsf{X}_e\right\|_1 \ge \left\|X_a - X_b\right\|_1.$$

□

In the following theorem we show that there is no PD shortcut matrix D that reduces the average effective resistance of all cuts of the graph of fig. 3.4 to $o(1)$.

Theorem 7.3. *For any $h > k > 2$, the optimum of* Average-CP *for the graph of fig. 3.4, is at most*

$$\frac{h^2}{8(h+k)^2}.$$

Proof. Fix k, h and let G be the graph of fig. 3.4. By (7.11) it is enough to construct a cut metric X, a semiorthogonal matrix U, and a distribution λ on the cuts of G such that

$$\frac{\left(\sum_{e\in E} \sqrt{\gamma_e}\cdot\langle U^e, \mathsf{X}_e\rangle\right)^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2} \le \frac{h^2}{8(h+k)^2}. \tag{7.12}$$

First, we construct X and we calculate the denominator, then we define U and γ, λ and we upper bound the numerator. Let $n = 2^h$ (so G has $n+1$ vertices). Let $X \in \{0,1\}^{n\times(n+1)}$ where for any vertex $0 \le i \le 2^h$, $X_i := \mathbf{1}_{[i]}$, i.e., X_i is 1 in the first i coordinates and 0 otherwise. So, $X_0 = \mathbf{0}$. It follows that

$$\sum_{\{i,j\}\in E} \left\|X_i - X_j\right\|_1 = n \cdot k + n \cdot h.$$

So, it remains to upper bound the numerator. Next, we define the semiorthogonal matrix U.

We define a semiorhogonal matrix U by describing the vectors that we assign to a carefully chosen set E' of "long" edges of G. For each $1 \le i \le h$, we assign a vector to each of the edges $\{0, 2^i\}, \{2\cdot 2^i, 3\cdot 2^i\}, \{4\cdot 2^i, 5\cdot 2^i\}, \dots$; we assign the following vector to the edges $\{2j\cdot 2^i, (2j+1)\cdot 2^i\}$:

$$U^{\{2j\cdot 2^i,(2j+1)\cdot 2^i\}} = \begin{array}{c} \begin{array}{ccccccccccc} 1 & & & 2j\cdot 2^i & & & (2j+1)\cdot 2^i & & & (2j+2)\cdot 2^i & \end{array} \\ \left[\begin{array}{ccc|ccc|ccc|ccc} 0 & \dots & 0 & \frac{+1}{\sqrt{2^i}} & \dots & \frac{+1}{\sqrt{2^i}} & \frac{-1}{\sqrt{2^i}} & \dots & \frac{-1}{\sqrt{2^i}} & 0 & \dots & 0 \end{array}\right]. \end{array}$$

Note that the above vector is only nonzero in the coordinates $2i\cdot 2^j$ to $(2i+2)\cdot 2^j - 1$; it is equal to $1/\sqrt{2^j}$ in the first half of these coordinates and $-1/\sqrt{2^j}$ in the second half. For example the rows of U corresponding to the 3 top layers of long edges look as follows:

$$\begin{array}{c|c} & \begin{array}{ccccccc} 1 & & n/4 & & n/2 & & 3n/4 \end{array} \\ \begin{array}{c} U^{\{0,n/2\}} \\ U^{\{0,n/4\}} \\ U^{\{n/2,3n/4\}} \end{array} & \left[\begin{array}{ccccc|cccc} \frac{+1}{\sqrt{n}} & & \dots & & \frac{+1}{\sqrt{n}} & \frac{-1}{\sqrt{n}} & & \dots & \frac{-1}{\sqrt{n}} \\ \frac{+\sqrt{2}}{\sqrt{n}} & \dots & \Big| & \frac{-\sqrt{2}}{\sqrt{n}} & \dots & 0 & & \dots & 0 \\ 0 & & \dots & & 0 & \frac{+\sqrt{2}}{\sqrt{n}} & \dots \Big| & \frac{-\sqrt{2}}{\sqrt{n}} & \dots \end{array}\right] \end{array}$$

Note that X can be extended to a matrix in $\{0,1\}^{E\times(n+1)}$ by adding zero rows, and U can be extended to an orthogonal matrix in $\mathbb{R}^{E\times E}$.

By the above construction for each edge $e = \{2j\cdot 2^i, (2j+1)\cdot 2^i\} \in E'$,

$$\langle U^e, \mathsf{X}_e\rangle = \frac{2^i}{2^{-(i+1)/2}} = 2^{(i-1)/2}. \tag{7.13}$$

Therefore, we can write the LHS of (7.12) as follows:

$$\frac{\left(\sum_{e\in E}\sqrt{\gamma_e}\cdot\langle U^e, \mathsf{X}_e\rangle\right)^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2} \ge \frac{\left(\sum_{e=\{j\cdot 2^i,(j+1)\cdot 2^i\}\in E'}\sqrt{\gamma_e}\cdot 2^{(i-1)/2}\right)^2}{n\cdot k + n\cdot h}$$

Note that we have an inequality because edges not in E' may have nonzero projection on the corresponding rows of U.

Now, let us define the distribution λ. Let $\lambda_{(S,\overline{S})} = 1/n$ for every cut $(\{0,1,\dots,\ell\}, \{\ell+1,\dots,n+1\})$ for all $0 \le \ell \le n-1$. Then, for any edge $\{2j\cdot 2^i, (2j+1)\cdot 2^i\}$,

$$\gamma_{\{2j\cdot 2^i,(2j+1)\cdot 2^i\}} = \sum_{2j\cdot 2^i \le \ell < (2j+1)\cdot 2^i} \frac{1}{n\cdot |E(\{0,\dots,\ell\},\{\ell+1,\dots,n+1\})|} \ge \frac{2^i}{n\cdot(h+k)}$$

In the above inequality we use the fact that the sum is over 2^i many cuts, and each "threshold cut" $(\{0,1,\ldots,\ell\},\{\ell+1,\ldots,n\})$ cuts at most $k+h$ edges of G.

Therefore, the optimum of Average-CP is at least,

$$\frac{\left(\sum_{i=0}^{h-1}\sum_{0\le 2j<2^{h-i}}\sqrt{\frac{2^i}{n\cdot(h+k)}}\cdot 2^{(i-1)/2}\right)^2}{n\cdot(h+k)} \ge \frac{\left(\sum_{i=0}^{h-1} n\cdot 2^{-i-1}\cdot\sqrt{\frac{2^i}{n\cdot(h+k)}}\cdot 2^{(i-1)/2}\right)^2}{n\cdot(h+k)}$$
$$= \frac{\left(2^{-3/2}\cdot h\cdot\sqrt{\frac{n}{h+k}}\right)^2}{n\cdot(h+k)} = \frac{h^2}{8(h+k)^2}$$

□

Let us conclude this section by demonstrating that Tree-CP performs better than Average-CP for the graph of fig. 3.4 with respect to the locally connected hierarchy of fig. 3.5. Let $\mathcal{T}$ be the tree shown in fig. 3.5 and $T=\{1,2,\ldots,2^h\}$. Let X and U be the cut metric and the orthogonal matrix constructed in the proof of theorem 7.3, respectively. Let us estimate $\sum_{e\in\mathcal{O}(t)}\langle U^e,\mathsf{X}_e\rangle$ for nodes $2^i\in T$; the rest of the terms can be estimated similarly. For node 2^i, $\mathcal{O}(2^i)$ has k copies of the edge $\{2^i-1,2^i\}$ and for each $1\le j\le i$, it has an edge $\{2^i-2^j,2^i\}$. By (7.13), for each edge $e=\{2^i-2^j,2^i\}$, $\langle U^e,\mathsf{X}_e\rangle=2^{(j-1)/2}$. Therefore, for any node 2^i, $\left(\sum_{e\in\mathcal{O}(2^i)}\langle U^e,\mathsf{X}_e\rangle\right)^2$ is a geometric sum and we can approximate it with the largest term, i.e., $\max_{e\in\mathcal{O}(2^i)}\langle U^e,\mathsf{X}_e\rangle^2$. Therefore,

$$\sum_{t\in T}\frac{1}{|\mathcal{O}(t)|}\left(\sum_{e\in\mathcal{O}(t)}\langle U^e,\mathsf{X}_e\rangle\right)^2 \lesssim \sum_{t\in T}\frac{1}{|\mathcal{O}(t)|}\max_{e\in\mathcal{O}(t)}\langle U^e,\mathsf{X}_e\rangle^2$$
$$\lesssim \sum_{t\in T}\frac{1}{k}\max_{e\in\mathcal{O}(t)}\langle U^e,\mathsf{X}_e\rangle^2 \le \frac{1}{k}\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2.$$

In the second inequality we use the crucial fact that each edge e is contained in $\mathcal{O}(t)$ for at most two nodes of $\mathcal{T}$ and that $|\mathcal{O}(t)|\ge k$ for all t. So, Tree-CP$(\mathcal{T})\le O(1/k)$.

7.2 Upper-bounding the Numerator of the Dual

In the rest of the dissertation we prove the following theorem.

Theorem 7.4. *For any k-edge-connected graph $G=(V,E)$ and any (k,λ,T)-LCH, of G, and for $h>0$, any cut metric $X\in\{0,1\}^{h\times V}$, and any semiorthogonal matrix $U\in\mathbb{R}^{E\times h}$,*

$$\frac{\sum_{t\in T}\frac{1}{|\mathcal{O}(t)|}\left(\sum_{e\in\mathcal{O}(t)}\langle U^e,\mathsf{X}_e\rangle\right)^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2} \le \frac{f_1(k,\lambda)}{k}. \tag{7.14}$$

Recall that $f_1(k,\lambda)$ is a polylogarithmic function of $k, 1/\lambda$. Observe that the above theorem together with lemma 7.1 implies theorem 4.2.

In the rest of the paper we fix U, X and we upper-bound the above ratio by $\operatorname{poly}\log(k, 1/\lambda)/k$. We also identify every vertex v with its map X_v.

Before getting into the details of the proof let us describe how k-edge-connectivity blends into our proof. In the following simple fact we show that to lower bound the denominator it is enough to find many disjoint L_1 balls centered at the vertices of G with large radii.

Fact 7.5. *For any $X : V \to \{0,1\}^h$ and any set of $\ell \geq 2$ disjoint L_1 balls $B_1, \ldots, B_\ell$ centered at vertices of G with radii $r_1, \ldots, r_\ell$ we have*

$$\sum_{i=1}^{\ell} r_i \cdot k \leq \sum_{e \in E} \left\| \mathbf{X}_e \right\|^2 .$$

Since there are k edge-disjoint paths connecting the center of each ball to the outside (see fig. 7.1), by the triangle inequality, the sum of the L_1 length of the edges of the graph is at least k times the sum of the radii of the balls. Note that if $\ell = 1$, i.e., if we have only one ball, the conclusion does not necessarily hold. This is because B_1 may contain all vertices of G.

Now, let us give a high-level overview of the proof of theorem 7.4. The main proof consists of two steps; in the rest of this section we upper-bound the numerator of the ratio in (7.14) by a quantity defined on a geometric object which we call a sequence of bags of balls. Then, in the next section we lower-bound the denominator by $\Omega(k)$ times the same quantity. The main result of this section is proposition 7.18, in which we construct a geometric sequence of bags of L_1 balls, $\mathcal{B}_1, \mathcal{B}_2, \ldots$, centered at the vertices of G such that balls in each $\mathcal{B}_i$ are disjoint and their radii are exactly equal to δ_i, where δ_1, δ_2, $\ldots$ form a $\operatorname{poly}(k, 1/\lambda)$-decreasing geometric sequence. We guarantee that the numerator is within a $\operatorname{poly}\log(k, 1/\lambda)$ factor of the sum of the radii of balls in the geometric sequence.

In section 7.3 we lower-bound the denominator, i.e., the sum of the L_1 lengths of the edges by $\Omega(k)$ times the sum of radii of the balls in our geometric sequence. At the heart of our dual proof in section 7.3, we use an inductive argument with no loss in n. We prove that under some technical conditions on $\mathcal{B}_1$,$\mathcal{B}_2$,$\ldots$, we can construct a set of label-disjoint (*hollowed*) balls such that the sum of the radii of these (hollowed) balls is a constant factor of the sum of the radii of balls in the given geometric sequence; by label-disjoint balls we mean that we can assign a set of nodes $\mathcal{C}(B) \subset \mathcal{T}$ to each (hollowed) ball B, called the conflict set of B, such that for any two intersecting (hollowed) balls B and B', $\mathcal{C}(B) \cap \mathcal{C}(B') = \emptyset$. Furthermore, we use properties of the locally connected hierarchy to ensure that for each (hollowed) ball B, there are $\Omega(k)$ edge-disjoint paths, supported on the vertices of G in $\mathcal{C}(B)$, crossing B.

In the rest of this section we construct a geometric sequence of bags of balls such that the sum of the radii of balls in the sequence is at least the numerator of (7.14) up to $\operatorname{poly}\log(k, 1/\lambda)$ factors (see proposition 7.18 for the final result of this section). First, in section 7.2 we prove a technical lemma; we show that if the average projection of a set F of edges on U is

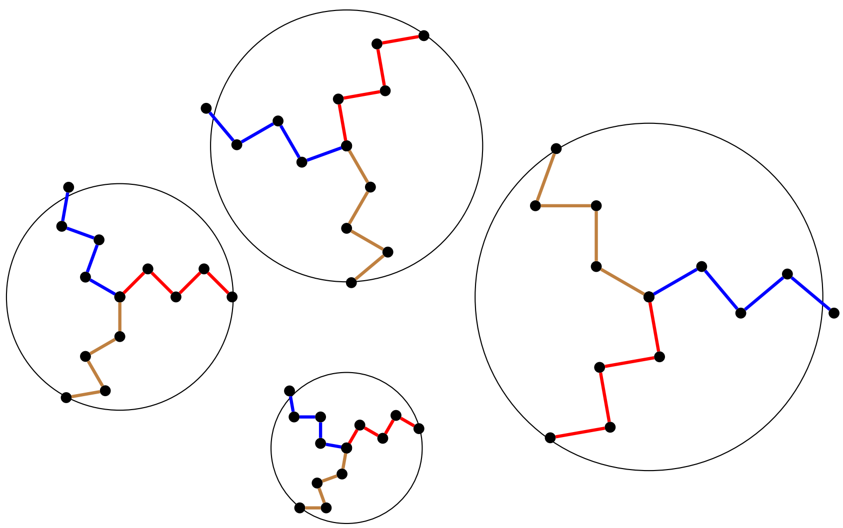

Figure 7.1: Sets of k edge-disjoint paths in disjoint L_1 balls.

"comparable" to the average squared norm of these edges, then we can construct a large number of disjoint balls centered at the endpoints of edges of F. We use this technical lemma to show that we can reduce the *average* effective resistance of any given set F of edges of any k-edge-connected graph to $\tilde{O}(1/k)$. Then, in section 7.2 we group these balls into several bags of balls. Finally, in section 7.2 we partition the edges of G into parts that have similar projections onto U and for each part we use the result of section 7.2 to find a family of bags of balls. Putting these families together we obtain a geometric sequence of families of bags of balls.

Construction of Disjoint L_1 Balls

In this section we prove the following proposition; although we do not directly use this proposition in the proof of our main technical theorem, we do use the main tool of the proof, lemma 7.7, as one of the key components of the proof for the main technical theorem.

Proposition 7.6. *For any k-edge-connected graph $G = (V, E)$ and any set $F \subseteq E$, there is a PD shortcut matrix D that reduces the average effective resistance of the edges of F to $\tilde{O}(1/k)$.*

By lemma 7.1 it is enough to show that for any $X \in \{0,1\}^{h \times V}$ and any semiorthogonal matrix $U \in \mathbb{R}^{E \times h}$,

$$\frac{\frac{1}{|F|}\left(\sum_{e\in F}\langle U^e, \mathsf{X}_e\rangle\right)^2}{\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2} = \frac{(\mathbb{E}_{e\sim F}\langle U^e, \mathsf{X}_e\rangle)^2}{\frac{1}{|F|}\sum_{e\in E}\left\|\mathsf{X}_e\right\|^2} \le \tilde{O}(1/k).$$

Let $Y = UX$ and $\mathbf{Y} = Y\chi = UX\chi$. Note that since U is semiorthogonal, $\left\|\mathbf{Y}_e\right\|^2 \leq \left\|\mathbf{X}_e\right\|^2$ for all e. Without loss of generality assume that

$$\frac{\left(\mathbb{E}_{e\sim F}\mathbf{Y}_{e,e}\right)^2}{\mathbb{E}_{e\sim F}\left\|\mathbf{Y}_e\right\|^2} \geq \alpha,$$

for $\alpha = \mathrm{poly}\log(k)/k$; otherwise we are done. In the following lemma we show that assuming the above inequality we can construct b disjoint L_2^2 balls of radius r centered at the vertices of the endpoints of edges of F such that

$$r \cdot b \geq \frac{\alpha^\epsilon}{\mathrm{poly}(\epsilon)} \cdot \left(\mathbb{E}_{e\sim F}\mathbf{Y}_{e,e}\right)^2 |F|.$$

On the other hand, since these balls are disjoint, by Fact 7.5,

$$r \cdot b \leq \frac{1}{k}\sum_{e\in E}\left\|\mathbf{X}_e\right\|^2.$$

Note that we really need to apply Fact 7.5 to balls in the space of X_v's, since Y_v's do not necessarily satisfy the triangle inequality. However, given disjoint balls centered around Y_v's, one can take the same balls around the corresponding X_v's and they will remain disjoint, since U, the mapping from X_v to Y_v, is a contraction.

Now, the above proposition simply follows by the above two inequalities for $\epsilon = \log k/\log\log k$.

Lemma 7.7. *Given $F \subseteq E$ and a mapping $Y \in \mathbb{R}^{E\times V}$ such that*

$$\Upsilon := \left(\mathop{\mathbb{E}}_{e\sim F}\mathbf{Y}_{e,e}\right)^2 \geq \alpha \cdot \mathop{\mathbb{E}}_{e\sim F}\left\|\mathbf{Y}_e\right\|_2^2, \tag{7.15}$$

for some $\alpha > 0$, for any $0 < \epsilon \leq 1$, there are b disjoint L_2^2 balls $B_1, \ldots, B_b$ with radius r such that the center of each ball is an endpoint of an edge in F, $b \geq \alpha|F|/C_1(\epsilon)$, and

$$r \cdot b \geq \frac{\alpha^\epsilon \cdot \Upsilon \cdot |F|}{C_1(\epsilon)},$$

where $C_1(\epsilon)$ is a polynomial function of $1/\epsilon$.

Before getting to the proof of the lemma, let us give an intuitive description of the statement of the lemma. The extreme case is for $\alpha \approx 1$. Observe that the inequality (7.15) enforces a very strong assumption on the mapping $\mathbf{Y}$. Since for any edge e, $\mathbf{Y}_{e,e} \leq \left\|\mathbf{Y}_e\right\|$, and $\alpha \approx 1$, the following two conditions must hold for $\mathbf{Y}$:

i) For most edges $e \in F$, $\mathbf{Y}_{e,e} \approx \left\|\mathbf{Y}_e\right\|$,

ii) For most pairs of edges $e, f \in F$, $\left\|\mathbf{Y}_e\right\| \approx \left\|\mathbf{Y}_f\right\|$.

The above two conditions essentially imply that the vectors $\{\mathbf{Y}_e\}_{e\in F}$ form an orthonormal basis up to normalizing the size of the vectors. It is an exercise to see that in this case one can select $\Omega(|F|)$ many L_2^2 balls of radius $\Omega(\Upsilon)$ around the endpoints of the edges in F; one can show that greedily picking balls that do not intersect each other works. Our proof can be interpreted as a robust version of this argument.

Proof of lemma 7.7. For a radius $r > 0$, run the following greedy algorithm. Scan the endpoints of the edges in an arbitrary order; for each point Y_u, if the L_2^2 ball $B(Y_u, r)$ doesn't touch the balls that we have already selected, select $B(Y_u, r)$. Suppose we manage to select b balls. We say the algorithm succeeds if both of the lemma's conclusions are satisfied. In the rest of the proof we show that this algorithm always succeeds for some value of r.

Without loss of generality, in the rest of the proof we drop the columns of $\mathbf{Y}$ corresponding to edges $e \notin F$ and their corresponding rows and we assume $Y \in \mathbb{R}^{F\times F}$. Note that by removing the rows, we are decreasing $\left\|\mathbf{Y}\right\|_F^2$, but this only weakens the assumption of the lemma. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{|F|}$ be the singular values of $\mathbf{Y}$. We can rewrite the assumption of the lemma as follows:

$$\left(\frac{1}{|F|}\sum_i \sigma_i\right)^2 \geq \left(\frac{\mathrm{Tr}(\mathbf{Y})}{|F|}\right)^2 = \left(\mathbb{E}_{e\sim F}\mathbf{Y}_{e,e}\right)^2 \geq \alpha \cdot \mathbb{E}_{e\sim F}\left\|\mathbf{Y}_e\right\|^2 = \frac{\alpha}{|F|}\left\|\mathbf{Y}\right\|_F^2 = \frac{\alpha}{|F|}\sum_{i=1}^{|F|}\sigma_i^2. \tag{7.16}$$

The first inequality follows by lemma 2.6. Note that, for $\alpha = 1$, the LHS is always less than or equal to the RHS by the Cauchy-Schwarz inequality with equality happening only when $\sigma_1 = \cdots = \sigma_{|F|}$. So, for large α the above inequality can be seen as a reverse Cauchy-Schwarz inequality.

In the next claim, we show that if the above algorithm finds a "small number" b of balls for a choice of r, this means that $\sigma_b, \ldots, \sigma_{|F|}$ are significantly smaller than $\sigma_1, \ldots, \sigma_{b-1}$. In the succeeding claim we use the above reverse Cauchy-Schwarz inequality to show that this is impossible.

Claim 7.8. *Given $r > 0$, suppose that the greedy algorithm finds b disjoint balls of radius r. Then*

$$r \geq \frac{1}{16|F|}\sum_{i=b}^{|F|}\sigma_i^2.$$

Proof. We construct a low-rank matrix $C \in \mathbb{R}^{F\times F}$. Then, we use theorem 2.7 to prove the claim. Let $Y_{w_1}, \ldots, Y_{w_b}$ be the centers of the chosen balls. Then, for any endpoint v of an edge in F, let $c(v)$ be the closest center to Y_v, i.e.,

$$c(v) := \mathrm{argmin}_{w_i}\left\|Y_{w_i} - Y_v\right\|_2^2$$

We construct a matrix $C \in \mathbb{R}^{F\times F}$ such that the e-th column of C is defined as follows: say the $\{u, v\}$-th column of $\mathbf{Y}$ is $Y_u - Y_v$ for $\{u, v\} \in F$; we let the $\{u, v\}$-th column of C be

$Y_{c(u)} - Y_{c(v)}$. By definition, $\operatorname{rank}(C) \leq b-1$, since C's columns are a subset of the differences between b points.

First, notice that

$$\begin{aligned}\left\|\Upsilon - C\right\|_F^2 &= \sum_{\{u,v\}\in F} \left\|(Y_u - Y_v) - (Y_{c(u)} - Y_{c(v)})\right\|_2^2 \\ &\leq \sum_{\{u,v\}\in F} (\left\|Y_u - Y_{c(u)}\right\|_2 + \left\|Y_v - Y_{c(v)}\right\|_2)^2 \\ &\leq \sum_{\{u,v\}\in F} 2\left\|Y_u - Y_{c(u)}\right\|_2^2 + 2\left\|Y_v - Y_{c(v)}\right\|_2^2 \leq 16r\cdot|F|,\end{aligned}$$

where the first inequality follows by the triangle inequality and the last inequality follows by the definition of greedy algorithm; in particular, for any point v, in the worst case there is a point p in the L_2^2 ball about $c(v)$ such that $\left\|p - Y_v\right\|^2 < r$, so

$$(\left\|Y_v - Y_{c(v)}\right\|_2)^2 \leq (\left\|Y_v - p\right\| + \left\|Y_{c(v)} - p\right\|)^2 \leq (\sqrt{r} + \sqrt{r})^2 \leq 4r.$$

Now by theorem 2.7,

$$16r\cdot|F| \geq \left\|\Upsilon - C\right\|_F^2 \geq \sum_{i=b}^{|F|} \sigma_i^2.$$

where the second inequality uses the fact that $\operatorname{rank}(C) \leq b-1$. □

All we need to show is that there is a value of $b \geq \alpha|F|/C_1(\epsilon)$ such that $\frac{b}{16|F|}\sum_{i=b}^{|F|}\sigma_i^2 \geq \frac{\alpha^\epsilon\Upsilon\cdot|F|}{C_1(\epsilon)}$.

Claim 7.9. *There is a universal function $C_1(\epsilon)$ that is polynomial in $1/\epsilon$ such that for any $0 < \epsilon \leq 1$ there is an integer $b \geq \alpha|F|/C_1(\epsilon)$ such that*

$$\frac{b}{16|F|}\sum_{i=b}^{|F|}\sigma_i^2 \geq \frac{\alpha^\epsilon\cdot\Upsilon\cdot|F|}{C_1(\epsilon)}.$$

Proof. Let us first prove the claim for $\epsilon = 1$; this special case reveals the meat of the argument. We show the claim holds for $b = \alpha|F|/4$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\left(\frac{1}{|F|}\sum_{i=1}^{|F|}\sigma_i\right)^2 &\leq 2\left(\frac{1}{|F|}\sum_{i=1}^{b-1}\sigma_i\right)^2 + 2\left(\frac{1}{|F|}\sum_{i=b}^{|F|}\sigma_i\right)^2 \\ &\leq \frac{2b}{|F|^2}\sum_{i=1}^{b-1}\sigma_i^2 + \frac{2}{|F|}\sum_{i=b}^{|F|}\sigma_i^2 \\ &= \frac{\alpha}{2|F|}\sum_{i=1}^{b-1}\sigma_i^2 + \frac{8b}{\alpha|F|^2}\sum_{i=b}^{|F|}\sigma_i^2 \leq \frac{1}{2}\left(\frac{1}{|F|}\sum_{i=1}^{|F|}\sigma_i\right)^2 + \frac{8b}{\alpha|F|^2}\sum_{i=b}^{|F|}\sigma_i^2.\end{aligned}$$

where the equality uses the definition of b and the last inequality uses (7.16). Therefore,

$$\frac{\Upsilon}{2} \leq \frac{1}{2}\left(\frac{1}{|F|}\sum_{i=1}^{|F|}\sigma_i\right)^2 \leq \frac{8b}{\alpha|F|^2}\sum_{i=b}^{|F|}\sigma_i^2,$$

where the first inequality uses another application of (7.16). This proves the claim for $\epsilon = 1$ and $C_1(\epsilon) \leq 1/256$.

Now, we prove the claim for $\epsilon < 1$. Let $b_0 \geq \frac{\alpha|F|}{C_1(\epsilon)}$ be an integer that we fix later. Let $z := \max_{b \geq b_0} \frac{b}{16|F|}\sum_{i=b}^{|F|}\sigma_i^2$. To prove the claim, it is enough to lower bound z. First, by the definition of z, for all $b \geq b_0$,

$$\frac{z}{|F|^\epsilon \cdot b^{1-\epsilon}} \geq \frac{b^\epsilon}{16|F|^{1+\epsilon}}\sum_{i=b}^{|F|}\sigma_i^2. \tag{7.17}$$

On the other hand, by (7.16),

$$\frac{1}{|F|}\cdot\sum_{i=1}^{|F|}\sigma_i^2 \leq \Upsilon/\alpha. \tag{7.18}$$

Let $\beta > 0$ be a parameter that we fix later. Summing up (7.17) for all $b_0 \leq b \leq |F|$ and β times (7.18), we get

$$\sum_{i=1}^{|F|}\left(\beta + \frac{\int_{x=b_0-1}^{i}(x-1)^\epsilon dx}{16|F|^\epsilon}\right)\cdot\frac{\sigma_i^2}{|F|} \leq \frac{\beta\cdot\Upsilon}{\alpha} + \frac{z}{|F|^\epsilon}\int_{x=b_0}^{|F|}\frac{dx}{(x-1)^{1-\epsilon}}.$$

Note that the integral on the LHS lower-bounds $\sum_{b_0 \leq b \leq i-1} b^\epsilon$ and the integral on the RHS upper-bounds $\sum_{b_0 \leq b < |F|} 1/b^{1-\epsilon}$. So,

$$\sum_{i=1}^{|F|}\left(\beta + \frac{[(i-1)^{1+\epsilon} - (b_0-1)^{1+\epsilon}]_+}{32|F|^\epsilon}\right)\cdot\frac{\sigma_i^2}{|F|} \leq \frac{\beta\cdot\Upsilon}{\alpha} + \frac{z}{|F|^\epsilon}\cdot\frac{(|F|-1)^\epsilon}{\epsilon} \leq \frac{\beta\cdot\Upsilon}{\alpha} + \frac{z}{\epsilon}. \tag{7.19}$$

where for $x \in \mathbb{R}$, $[x]_+ = \max\{x, 0\}$.

Therefore, by (7.16) and Cauchy-Schwarz,

$$\begin{aligned}\Upsilon \leq \left(\frac{1}{|F|}\cdot\sum_{i=1}^{|F|}\sigma_i\right)^2 &\leq \left(\sum_{i=1}^{|F|}\left(\beta + \frac{[(i-1)^{1+\epsilon} - (b_0-1)^{1+\epsilon}]_+}{32|F|^\epsilon}\right)\frac{\sigma_i^2}{|F|}\right)\cdot\left(\sum_{i=1}^{|F|}\frac{1/|F|}{\beta + \frac{[(i-1)^{1+\epsilon}-(b_0-1)^{1+\epsilon}]_+}{32|F|^\epsilon}}\right) \\ &\leq \left(\frac{\beta\cdot\Upsilon}{\alpha} + \frac{z}{\epsilon}\right)\cdot\frac{32(3+1/\epsilon)}{\beta^{\frac{\epsilon}{1+\epsilon}}|F|^{\frac{1}{1+\epsilon}}}.\end{aligned} \tag{7.20}$$

To see the last inequality we need to do some algebra. The first term on the RHS follows from (7.19). We obtain the second term in the last inequality by choosing $b_0 = 1 + \beta^{\frac{1}{1+\epsilon}}|F|^{\frac{\epsilon}{1+\epsilon}}$; later we will choose $\beta, C_1(\epsilon)$ making sure that $b_0 \geq \alpha|F|/C_1(\epsilon)$. In particular,

$$\sum_{j=1}^{|F|} \frac{1/|F|}{\beta + \frac{[(j-1)^{1+\epsilon}-(b_0-1)^{1+\epsilon}]_+}{32|F|^{\epsilon}}} \leq \frac{b_0-1}{\beta|F|} + \sum_{i=1}^{\infty} \sum_{j=(b_0-1)i^{1/(1+\epsilon)}+1}^{(b_0-1)(i+1)^{1/(1+\epsilon)}} \frac{32}{i \cdot \beta \cdot |F|}$$
$$\leq \frac{b_0-1}{\beta \cdot |F|}\left(1 + \sum_{i=1}^{\infty} \frac{32}{i^{\frac{1+2\epsilon}{1+\epsilon}}}\right) \leq \frac{32(3+1/\epsilon)(b_0-1)}{\beta|F|} \leq \frac{32(3+1/\epsilon)}{\beta^{\frac{\epsilon}{1+\epsilon}}|F|^{\frac{1}{1+\epsilon}}},$$

where in the first inequality we used $b_0 \geq 1 + \beta^{\frac{1}{1+\epsilon}}|F|^{\frac{\epsilon}{1+\epsilon}}$, in second inequality we used $(i+1)^{\frac{1}{1+\epsilon}} - i^{\frac{1}{1+\epsilon}} = i^{\frac{1}{1+\epsilon}}\big((1+1/i)^{\frac{1}{1+\epsilon}} - 1\big) \leq i^{\frac{-\epsilon}{1+\epsilon}}$, and in the last inequality we used $b_0 \leq 1 + \beta^{\frac{1}{1+\epsilon}}|F|^{\frac{\epsilon}{1+\epsilon}}$.

Now, the claim follows directly from (7.20). Letting $\beta = \frac{\alpha^{1+\epsilon}|F|}{(192+64/\epsilon)^{1+\epsilon}}$, we obtain,

$$z \geq \frac{\epsilon \cdot \beta^{\frac{\epsilon}{1+\epsilon}} \cdot |F|^{\frac{1}{1+\epsilon}} \cdot \Upsilon}{32(3+1/\epsilon)} - \frac{\epsilon \cdot \beta \cdot \Upsilon}{\alpha} \geq \frac{\alpha^{\epsilon} \cdot \Upsilon \cdot |F|}{(192/\epsilon + 64/\epsilon^2)^{1+\epsilon}}.$$

The claim follows by letting $C_1(\epsilon) = (192/\epsilon + 64/\epsilon^2)^{1+\epsilon}$, and noting $b_0 = 1 + \beta^{\frac{1}{1+\epsilon}}|F|^{\frac{\epsilon}{1+\epsilon}} \geq \alpha|F|/C_1(\epsilon)$. □

Observe that the above claim implies lemma 7.7. It is sufficient to run the greedy algorithm with the infimum value of r such that the greedy algorithm returns at most b balls. □

Construction of Bags of Balls

In this subsection we will state the main result of this section, proposition 7.18, and we give a high-level overview of the proof of theorem 7.4. Before that we need to define several combinatorial objects called bags of balls.

To prove theorem 7.4 we would like to follow a path similar to the proof of proposition 7.6, i.e., we would like to construct disjoint L_1 balls $B_1, B_2, \ldots$ centered at the vertices of G of radius $r_1, r_2, \ldots$ such that

$$\sum_i r_i \gtrsim \sum_{t \in T} \frac{1}{|\mathcal{O}(t)|}\left(\sum_{e \in \mathcal{O}(t)} \langle U^e, X_e \rangle\right)^2, \tag{7.21}$$

and then use a variant of Fact 7.5. This approach completely fails for the example of fig. 3.4 as we will show next.

Example 7.10. *Let G be a modification of the graph of fig. 3.4 with $n = 2^h + 1$ vertices where we remove all long edges of length $2^i \leq h$, and we shift all edges of length $2^i > h$ by i to the right, i.e., we replace an edge $\{j \cdot 2^i, (j+1) \cdot 2^i\}$ with $\{i + j \cdot 2^i, i + (j+1) \cdot 2^i\}$. It is easy to see that in this new graph the degree of every vertex is at most $O(k)$.*

Let $X_0, X_1, \ldots, X_{2^h}$ be the embedding of G where $X_i = \mathbf{1}_{[i]}$ and U be the semiorthogonal matrix we constructed in theorem 7.3. Suppose T has all vertices of G, i.e., we are minimizizing the average effective resistance of all degree cuts. If we follow the approach in the previous section, to prove (7.14)*, we need to construct disjoint L_1 balls $B_1, B_2, \ldots,$ with radii $r_1, r_2, \ldots$ such that*

$$\sum_i r_i \gtrsim \sum_{v \in V} \frac{1}{d(v)} \left(\sum_{e \in E(v, V \setminus \{v\})} \langle U^e, \mathsf{X}_e \rangle \right)^2 .$$

It follows that for the particular choice of X, U the RHS is about

$$\frac{n \log n}{\max_v d(v)} \gtrsim \frac{n \log n}{k} \gg n,$$

for $k \ll \log n$. Unfortunately, for any set of disjoint L_1 balls centered at vertices of G we have $\sum_i r_i \leq n$. So, it is impossible to prove (7.21) *for $k \ll \log n$ using disjoint balls.*

We will deviate from the approach of the previous section in two ways. First, the balls that we construct have different radii, in fact the radii of the balls form a geometrically decreasing sequence with a sufficiently large poly(k) decreasing factor; secondly, only the balls of the same radii are disjoint, but a small ball can completely lie inside a bigger ball.

To construct these balls we will group the edges of G based on their lengths into $\log(n)$ buckets and we apply lemma 7.7 to each bucket separately; we actually have a more complicated bucketing because we want to make sure that any two edges e, f in one bucket satisfy $\langle U^e, \mathsf{X}_e \rangle \approx \langle U^f, \mathsf{X}_f \rangle$ and $\|\mathsf{X}_e\| \approx \|\mathsf{X}_f\|$.

Since the balls that we construct are not disjoint we can no longer use the simple charging argument of Fact 7.5. Instead, we partition the set of balls of each radii into bags. The balls of a bag must satisfy certain properties that we describe next. These properties of bags of balls will be crucially used in section 7.3 to lower bound $\sum_e \|\mathsf{X}_e\|_1$.

Definition 7.11 (Bag of Balls)**.** *A bag of balls,* Bag*, is a set of disjoint L_1 balls of equal radii such that the center of each ball is a point X_v for some $v \in V$. A bag of balls is of type* (δ) *if each ball in the bag has radius δ. A bag of balls is of type* (δ, Δ) *if in addition to above, the maximum L_1 distance between the centers of the balls in the bag is at most Δ,*

$$\max_{B(X_v,\delta), B(X_u,\delta) \in \mathrm{Bag}} \|X_v - X_u\|_1 \leq \Delta. \tag{7.22}$$

We write $|\,\mathrm{Bag}\,|$ to denote the number of balls in Bag.

Definition 7.12 (Compact Bag of Balls)**.** *For $\beta > 0$, a bag of balls,* Bag*, with type* (δ, Δ) *is* β-compact *if $|\,\mathrm{Bag}\,| \geq 2$ and*

$$\beta \cdot \Delta \leq |\,\mathrm{Bag}\,| \cdot \delta. \tag{7.23}$$

It follows from the definition that for any compact bag of balls of type (δ, Δ), $\Delta \geq 2\delta$.

Definition 7.13 (Assigned Bag of Balls). *For a locally connected hierarchy $\mathcal{T}$ and $\beta > 0$, a bag of balls,* Bag, *with type (δ) is β-assigned to a node $t \in \mathcal{T}$, if*

$$\beta \cdot |\mathcal{O}(t)| \leq |\,\mathrm{Bag}\,|, \tag{7.24}$$

and for each ball $B(X_u, \delta) \in$ Bag, $u \in V(t)$ and there is an edge $\{u, v\} \in \mathcal{O}(t)$ such that $\left\|X_u - X_v\right\|_1 < \delta$.

We use the convention of writing Bag_t for a bag of balls assigned to a node t.

In section 7.3 we will show that β-compact bags of balls with $\beta \geq C$ and β-assigned bags of balls with $\beta \geq C'/k$ for some universal constants C, C' are enough to lower-bound the denominator of (7.14).

In general, compact bags of balls are significantly easier to handle than assigned bags of balls. Roughly speaking, given a number of compact bag of balls we can use the compactness property to "carve" them into disjoint *hollowed* balls such that the sum of the widths of the hollowed balls is at least a constant fraction of the sum of the original radii; then we use an argument analogous to Fact 7.5 to lower bound the denominator (see section 7.3 for the details).

On the other hand, it is impossible to construct disjoint (hollowed) balls out of a given number of assigned bags of balls without losing too much on the sum of the widths. Instead of restricting the (hollowed) balls to be disjoint, in the technical proof presented in section 7.3, we label the balls of each bag with the node of the locally connected hierarchy to which it is assigned. We construct a conflict set, $\mathcal{C}(B)$, by looking at the subtree rooted at the label of B, and pruning some of its subtrees. We carve the balls and modify the labels in such a way that, in the end, for any two intersecting balls, the conflict set s are disjoint. Then, to charge the denominator, we use the fact that for any node t, $G(t)$ is k-edge-connected and thus in any of the remaining balls, we can find $\Omega(k)$ edge-disjoint paths contained in $G(t)$; this argument does not overcharge the edges, because throughout the construction we make sure that these edge-disjoint paths are routed through $\mathcal{C}(B)$.

Definition 7.14 (Family of Bags of Balls). *A family of bags of balls,* FBag, *is a set of bags of balls of the same type such that all balls in all bags are disjoint. We say a family of compact bags of balls has type (δ, Δ) if all bags in the family have type (δ, Δ). For a locally connected hierarchy, $\mathcal{T}$, and $T \subseteq \mathcal{T}$, we say a family of assigned bags of balls has type (δ, T) if the bags in the family are assigned to distinct nodes of T.*

We abuse notation and write a ball $B \in$ FBag if there is a Bag $\in$ FBag such that $B \in$ Bag. Note that two distinct bags in FBag may have unequal numbers of balls.

To upper-bound the value of the dual we need to find a sequence of families of bags of balls with geometrically decreasing radii.

Definition 7.15 (Geometric Sequence of Assigned Bags of Balls). *For a locally connected hierarchy, $\mathcal{T}$, a λ-geometric sequence of families of assigned bags of balls is a sequence* $\mathrm{FBag}_1, \mathrm{FBag}_2, \ldots$ *such that* FBag_i *has type* (δ_i, T_i) *where* $T_1, T_2, \ldots$ *are disjoint subsets of nodes of* $\mathcal{T}$ *and for all* $i \geq 1$,

$$\delta_i \cdot \lambda > \delta_{i+1}.$$

Definition 7.16 (Geometric Sequence of Compact Bags of Balls). *A λ-geometric sequence of families of compact bags of balls is defined respectively as a sequence* $\mathrm{FBag}_1, \mathrm{FBag}_2, \ldots,$ *such that* FBag_i *has type* (δ_i, Δ_i), *and for all* $i \geq 1$,

$$\delta_i \cdot \lambda > \Delta_{i+1}.$$

Now, we are ready to describe the main result of this section. Let us justify the assumption of the proposition 7.18.

Definition 7.17 (α-bad Nodes). *We say a node $t \in \mathcal{T}$ is* α-bad *if*

$$\left(\mathbb{E}_{e\sim\mathcal{O}(t)}\langle U^e, \mathsf{X}_e\rangle\right)^2 \geq \alpha \cdot \mathbb{E}_{e\sim\mathcal{O}(t)} \left\|\mathsf{X}_e\right\|^2. \tag{7.25}$$

First, observe that if there is no $\tilde{O}(1/k)$-bad node in T, then we are done with theorem 7.4. So, to prove theorem 7.4 the only thing that we need to upper bound is the contribution of the bad nodes to the numerator. In the following proposition, we construct a λ-geometric sequence of bags of balls such that the sum of the radii of all balls in the sequence is at least

$$\frac{1}{\mathrm{poly}\log(k, 1/\lambda)} \sum_{t \text{ is } \alpha\text{-bad}} \frac{1}{|\mathcal{O}(t)|} \left(\sum_{e\in\mathcal{O}(t)} \langle U^e, \mathsf{X}_e\rangle\right)^2.$$

Note that, we construct a geometric sequence of families of either compact or assigned bags of balls.

Proposition 7.18. *Given a locally connected hierarchy $\mathcal{T}$ of G and a set $T \subseteq \mathcal{T}$ of* α-bad *nodes, for any $\beta > 1$, $\epsilon < 1/3$, and $\lambda < 1$, if α is sufficiently small such that $(\alpha/C_2(\alpha))^\epsilon \lesssim \frac{1}{\beta\cdot C_1(\epsilon)}$, then one of the following holds:*

1. *There is a λ-geometric sequence of families of β-compact bags of balls* $\mathrm{FBag}_1, \mathrm{FBag}_2, \ldots,$ *where* FBag_i *has type* (δ_i, Δ_i) *such that*
$$\frac{(\alpha/C_2(\alpha))^\epsilon}{\beta C_1(\epsilon) C_2(\alpha)\cdot|\log(\lambda \mathrm{poly}(\alpha))|}\cdot\sum_{t\in T}\frac{1}{|\mathcal{O}(t)|}\Big(\sum_{e\in\mathcal{O}'(t)}\langle U^e, \mathsf{X}_e\rangle\Big)^2 \leq \sum_i \sum_{\mathrm{Bag}\in\mathrm{FBag}_i} \delta_i\cdot|\mathrm{Bag}|. \tag{7.26}$$

2. *There is a λ-geometric sequence of families of $(\alpha/C_2(\alpha))^{1+2\epsilon}$-assigned bags of balls* $\mathrm{FBag}_1, \mathrm{FBag}_2, \ldots,$ *where* FBag_i *has type* (δ_i, S_i) *such that*
$$\frac{(\alpha/C_2(\alpha))^\epsilon}{\beta C_1(\epsilon) C_2(\alpha)|\log(\lambda \mathrm{poly}(\alpha))|}\cdot\sum_{t\in T}\frac{1}{|\mathcal{O}(t)|}\Big(\sum_{e\in\mathcal{O}'(t)}\langle U^e, \mathsf{X}_e\rangle\Big)^2 \leq \sum_i \sum_{\mathrm{Bag}\in\mathrm{FBag}_i} \delta_i\cdot|\mathrm{Bag}|. \tag{7.27}$$

Here, C_1 is the polynomial function that we defined in lemma 7.7 and C_2 is a polylogarithmic function that we will define in lemma 7.22.

In the proof of theorem 7.4, we invoke the above proposition for $\alpha = \operatorname{poly}\log(k)/k$, $\epsilon = \Theta(\log k/\log\log k)$, and λ be $1/\operatorname{poly}(k)$ fraction of the λ given in the statement of the theorem.

In the rest of this section, we prove the above proposition using lemma 7.7. We do this in two intermediate steps. In the first step we extract a $1/\operatorname{poly}\log(\alpha)$-*dominating* 2-*homogeneous* set $\mathcal{O}'(t)$ of edges in each $\mathcal{O}(t)$ for any bad node t according to the following definitions.

Definition 7.19 (Homogeneous Edges). *For $c > 1$, we say a set $F \subseteq E$ of edges is c-homogeneous if for any two edges $e, f \in F$,*

$$\frac{\langle U^e, \mathsf{X}_e\rangle^2}{\langle U^f, \mathsf{X}_f\rangle^2} < c \text{ and } \frac{\|\mathsf{X}_e\|_2^2}{\|\mathsf{X}_f\|_2^2} < c.$$

Definition 7.20 (Dominating Subset). *For a node $t \in \mathcal{T}$ a set $\mathcal{O}'(t) \subseteq \mathcal{O}(t)$ is called γ-dominating if*

$$\Big(\sum_{e\in\mathcal{O}'(t)} \langle U^e, \mathsf{X}_e\rangle\Big)^2 \geq \gamma \cdot \Big(\sum_{e\in\mathcal{O}(t)} \langle U^e, \mathsf{X}_e\rangle\Big)^2.$$

The term dominating refers to the fact that the set $\mathcal{O}'(t)$ essentially captures the contribution of the edges of $\mathcal{O}(t)$ to the numerator.

Then, we group the bad nodes into sets T_i such that the set $\cup_{t\in T_i}\mathcal{O}'(t)$ is homogeneous for all i. In the second step, we use lemma 7.7 to construct bags of balls for a give group of homogeneous edges. We postpone the first step to the next subsection.

Lemma 7.21. *Given a locally connected hierarchy $\mathcal{T}$ of G, a set $T \subseteq \mathcal{T}$ of α-bad nodes, and γ-dominating sets $\mathcal{O}'(t) \subseteq \mathcal{O}(t)$ for each $t \in T$ such that $\cup_{t\in T}\mathcal{O}'(t)$ is 4-homogeneous, for any $0 < \epsilon < 1/2$ and $\beta > 1$, if α, γ are sufficiently small such that $(\alpha\cdot\gamma)^\epsilon \lesssim \frac{1}{\beta\cdot C_1(\epsilon)}$, then one of the following holds:*

1. *There is a family of β-compact bags of balls with type (δ, Δ),* FBag, *such that*

$$\sum_{t\in T} \frac{1}{|\mathcal{O}(t)|}\Big(\sum_{e\in\mathcal{O}'(t)} \langle U^e, \mathsf{X}_e\rangle\Big)^2 \lesssim \frac{C_1(\epsilon)}{(\alpha\cdot\gamma)^\epsilon} \sum_{\mathrm{Bag}\in\mathrm{FBag}} \delta\cdot|\mathrm{Bag}|. \tag{7.28}$$

2. *There is a family of $(\alpha\cdot\gamma)^{1+2\epsilon}$-assigned bags of balls with type (δ, S),* FBag, *and $S \subseteq T$ such that*

$$\sum_{t\in T} \frac{1}{|\mathcal{O}(t)|}\cdot\Big(\sum_{e\in\mathcal{O}'(t)} \langle U^e, \mathsf{X}_e\rangle\Big)^2 \lesssim \frac{\beta C_1(\epsilon)}{(\alpha\cdot\gamma)^\epsilon} \sum_{\mathrm{Bag}\in\mathrm{FBag}} \delta\cdot|\mathrm{Bag}|. \tag{7.29}$$

where in both cases $\delta, \Delta = \min_{e \in \mathcal{O}'(t), t \in T} \langle U^e, X_e \rangle^2$ *up to an* $O(\alpha \cdot \gamma)$ *factor.*

Proof. Let,

$$
\begin{aligned}
F &:= \cup_{t \in T} \mathcal{O}'(t), \\
c_1 &:= \min_{e \in F} \langle U^e, \mathsf{X}_e \rangle^2, \\
c_2 &:= \max_{e \in F} \left\| \mathsf{X}_e \right\|_2^2, \\
N &:= \left| \cup_{t \in T} \mathcal{O}(t) \right|, \\
N' &:= \left| \cup_{t \in T} \mathcal{O}'(t) \right| = |F|.
\end{aligned}
$$

Note that $N \geq N'$ by definition. First, we show that the edges in F satisfy the assumption of lemma 7.7 with α replaced by $\asymp \alpha \gamma N / N'$. Then, we invoke lemma 7.7 and we obtain many disjoint balls $\mathcal{A}$ such that the sum of their radii is comparable to LHS of (7.28) or (7.29) (see (7.34)). Then, we greedily construct a new set $\mathcal{B}$ of disjoint large balls of radii $\Delta \geq c_2$. If $|\mathcal{B}|$ is small, we can partition the balls of $\mathcal{A}$ into compact bags of balls; otherwise, we use balls of $\mathcal{B}$ to construct assigned bags of balls.

First, observe that,

$$
\begin{aligned}
c_1 \cdot N' \gtrsim \sum_{t \in T} \frac{1}{|\mathcal{O}(t)|} \Big(\sum_{e \in \mathcal{O}'(t)} \langle U^e, \mathsf{X}_e \rangle \Big)^2 &\geq \sum_{t \in T} \frac{\gamma}{|\mathcal{O}'(t)|} \Big(\sum_{e \in \mathcal{O}(t)} \langle U^e, \mathsf{X}_e \rangle \Big)^2 \\
&\geq \sum_{t \in T} \gamma \cdot \alpha \cdot \sum_{e \in \mathcal{O}(t)} \left\| \mathsf{X}_e \right\|^2 \\
&\geq \alpha \cdot \gamma \cdot N \cdot c_2. \qquad (7.30)
\end{aligned}
$$

where the first inequality follows by 4-homogeneity of F, the second inequality uses the fact that each $\mathcal{O}'(t)$ is γ-dominating, the third inequality uses that each node t is α-bad, and the last inequality again uses the 4-homogeneity of F. This is the only place in the proof that we use $t \in T$ is α-bad and $\mathcal{O}'(t)$ is γ-dominating. By the above equation we can choose $\tilde{\alpha} \asymp \alpha \gamma N / N'$ such that

$$
\Big(\mathop{\mathbb{E}}_{e \sim F} \langle U^e, \mathsf{X}_e \rangle \Big)^2 \geq \tilde{\alpha} \cdot \mathop{\mathbb{E}}_{e \sim F} \left\| \mathsf{X}_e \right\|_2^2 .
$$

Throughout the proof we use that $\tilde{\alpha} \gtrsim \alpha\gamma$. Let $Y_v := U X_v$ for all $v \in V$. Since U is semiorthogonal, for each pair u, v

$$
\left\| Y_u - Y_v \right\|_2^2 \leq \left\| X_u - X_v \right\|_2^2 = \left\| X_u - X_v \right\|_1 .
$$

Applying lemma 7.7 to Y and F, we obtain a family $\mathcal{A}$ of b disjoint L_2^2 balls with radius δ such that

$$
b \geq \frac{\tilde{\alpha} N'}{C_1(\epsilon)}, \qquad (7.31)
$$

and

$$\delta \cdot b \geq \frac{\tilde{\alpha}^{\epsilon} \cdot N' \cdot c_1}{C_1(\epsilon)}. \tag{7.32}$$

Now, we extract *disjoint* L_1 balls in the space of $\{X_v\}_{v \in V}$ with radius δ out of balls in $\mathcal{A}$. Balls in $\mathcal{A}$ correspond to L_2^2 balls in the X embedding. Since U is a contraction operator, these L_2^2 balls are disjoint in the X embedding. Now, L_2^2 balls with radius δ are L_2 balls with radius $\sqrt{\delta}$, so the L_2^2 distance between the centers of any two balls is at least 4δ. Since X is a cut metric, the L_2^2 distance between centers is the same as their L_1 distance, so L_1 balls with radius δ around the same centers are disjoint (in fact radius 2δ works as well). So, by abusing notation we let $\mathcal{A}$ be the L_1 balls in the X embedding.

Next, we construct the large balls. Let

$$V'(t) = \{u \in V(t) : \exists \{u, v\} \in \mathcal{O}'(t)\}$$

be the endpoints of edges of $\mathcal{O}'(t)$ that are in $V(t)$. Also, let $V' = \cup_{t \in T} V'(t)$. Let $\mathcal{B}$ be a maximal family of disjoint L_1 balls of radius Δ on the points in V' for $\Delta := \max\{\delta, c_2\}$. To construct $\mathcal{B}$, we scan the points in V' in an arbitrary order; for each point X_u if the ball $B(X_u, \Delta)$ does not touch any of the balls already added to $\mathcal{B}$ we add B to $\mathcal{B}$. We will consider two cases depending on the size of $\mathcal{B}$; if $|\mathcal{B}|$ is small we construct compact bags of balls and we conclude with case (1); otherwise we construct assigned bags of balls and we conclude with (2).

Before getting into the details of the two cases, we prove two facts that are useful for both cases. First, without loss of generality, perhaps by decreasing δ, we assume $\delta \cdot b \asymp \frac{c_1 \tilde{\alpha}^{\epsilon} N'}{C_1(\epsilon)}$. We can bound δ as follows

$$\gamma \cdot \alpha \cdot c_1 \lesssim \frac{c_1 \tilde{\alpha}^{\epsilon} N' / C_1(\epsilon)}{N'} \lesssim \frac{\delta \cdot b}{b} = \delta = \frac{\delta \cdot b}{b} \lesssim \frac{c_1 \tilde{\alpha}^{\epsilon} N'}{\tilde{\alpha} N'} \leq \frac{c_1}{\gamma \cdot \alpha}, \tag{7.33}$$

where the first inequality uses the lemma's assumption that $(\gamma\alpha)^{1-\epsilon} \leq (\gamma\alpha)^{\epsilon} \lesssim 1/C_1(\alpha)$, the second inequality uses $b \leq 2N'$, the third inuequality uses $b \gtrsim \frac{\tilde{\alpha} N'}{C_1(\epsilon)}$ and the last inequality uses $\tilde{\alpha} \geq \gamma \cdot \alpha$.

Secondly, it follows from (7.32) that

$$\frac{\sum_{t \in T} \frac{1}{|\mathcal{O}(t)|} (\sum_{e \in \mathcal{O}'(t)} \langle U^e, \mathbf{X}_e \rangle)^2}{b \cdot \delta} \lesssim \frac{\sum_{t \in T} |\mathcal{O}'(t)| \cdot c_1}{c_1 \tilde{\alpha}^{\epsilon} N' / C_1(\epsilon)} \leq \frac{C_1(\epsilon)}{\tilde{\alpha}^{\epsilon}}. \tag{7.34}$$

In the above we used $|\mathcal{O}'(t)| \leq |\mathcal{O}(t)|$ for all t. To prove the lemma, in the first case we construct a family of compact bags of balls with at least $b/2$ balls of $\mathcal{A}$, and in the second case we construct a family of assigned bags of balls with at least $|\mathcal{B}|/2$ balls of $\mathcal{B}$.

Case 1. $|\mathcal{B}| < \frac{b \cdot \delta}{12\beta \cdot \Delta}$. We construct a family of compact bags of balls. For each ball $B = B(X_u, \Delta) \in \mathcal{B}$ let

$$f(B) := \left\{ B(X_v, \delta) \in \mathcal{A} : \left\| X_u - X_v \right\|_1 = \min_{B(X_{u'}, \Delta) \in \mathcal{B}} \left\| X_{u'} - X_v \right\|_1 \right\},$$

be the balls of $\mathcal{A}$ that are closer to B than any other ball of $\mathcal{B}$. We break ties arbitrarily, making sure that $f(B) \cap f(B') = \emptyset$ for any two distinct balls of $\mathcal{B}$.

First, we show any set $f(B)$ is a bag of balls of type $(\delta, 6\Delta)$; then we add those that are β-compact to FBag. It is sufficient to show that for any $B(X_u, \Delta) \in \mathcal{B}$, the L_1 distance between the centers of balls of $f(B)$ is at most 6Δ. Fix a ball $B = B(X_u, \Delta) \in \mathcal{B}$. For any ball $B(X_{v_1}, \delta) \in f(B)$ we show that $\left\|X_u - X_{v_1}\right\|_1 \leq 3\Delta$. Since for all $e \in F$, $\left\|\mathbf{X}_e\right\|_1 \leq c_2$, there is a vertex $u_1 \in V'$ such that $\left\|X_{v_1} - X_{u_1}\right\| \leq c_2$. Furthermore, by construction of $\mathcal{B}$, there is a ball $B(X_{u_2}, \Delta) \in \mathcal{B}$ such that $\left\|X_{u_1} - X_{u_2}\right\|_1 \leq 2\Delta$. Putting these together,

$$\left\|X_{v_1} - X_u\right\|_1 \leq \left\|X_{v_1} - X_{u_2}\right\|_1 \leq \left\|X_{v_1} - X_{u_1}\right\|_1 + \left\|X_{u_1} - X_{u_2}\right\|_1 \leq c_2 + 2\Delta \leq 3\Delta.$$

So, the L_1 distance between the centers of balls of $f(B)$ is at most 6Δ.

So, we just need to add those bags that are β-compact to FBag. For each $B \in \mathcal{B}$ if $|f(B)| \geq \beta \cdot (6\Delta)/\delta$, then $f(B)$ is β-compact, as $|f(B)| \geq 2$ and

$$\beta \cdot (6\Delta) \leq \delta \cdot |f(B)|.$$

So, we add $f(B)$ to FBag. Observe that all balls of FBag are disjoint because all balls of $\mathcal{A}$ are disjoint.

It remains to verify that FBag satisfies conclusion (1). First, by (7.33) and the fact that $\Delta = \max\{\delta, c_2\}$, $6\Delta \gtrsim \alpha \cdot \gamma \cdot c_1$. On the other hand, by (7.30), $c_2 \leq c_1/\alpha$ as shown in ,

$$6\Delta \lesssim \max\{\delta, c_2\} \lesssim \{c_1/\alpha\gamma, c_1/\alpha\gamma\}.$$

So we just need to verify (7.28). It is easy to see that the number of balls in FBag is at least $b/2$. This is because,

$$\sum_{\mathrm{Bag} \in \mathrm{FBag}} |\mathrm{Bag}| \geq b - \sum_{B \in \mathcal{B}} \mathbb{I}\left[|f(B)| < \frac{\beta \cdot (6\Delta)}{\delta}\right] \cdot |f(B)| \geq b - |\mathcal{B}| \cdot \frac{\beta \cdot (6\Delta)}{\delta} \geq b/2.$$

The last inequality uses the assumption of case 1, $|\mathcal{B}| \leq \frac{b \cdot \delta}{12\beta \cdot \Delta}$. So, (7.28) follows by (7.34).

Case 2. $|\mathcal{B}| \geq \frac{b \cdot \delta}{12\beta \cdot \Delta}$. We construct an assigned family of bags of balls. For any node $t \in T$, let Bag_t be the set of balls in $\mathcal{B}$ such that their centers are in $V'(t)$. If the center of a ball B in $\mathcal{B}$ belongs to multiple $V'(t)$'s we include B in exactly one of those sets arbitrarily. Note that each Bag_t is a bag of balls with type (Δ). For each $t \in T$, if

$$\frac{|\mathrm{Bag}_t|}{|\mathcal{B}|} \geq \frac{|\mathcal{O}(t)|}{4N}, \tag{7.35}$$

then we add Bag_t to FBag and we add t to S. Next, we argue that FBag is a family of $(\alpha \cdot \gamma)^{1+2\epsilon}$-assigned bag of balls. First, balls in FBag are disjoint because they are a subset of balls of $\mathcal{B}$ and each ball of $\mathcal{B}$ is in at most one bag of FBag.

Fix a node $t \in S$. We show Bag_t is $(\alpha\cdot\gamma)^{1+2\epsilon}$-assigned. Since for any ball $B(X_u,\Delta) \in \mathrm{Bag}_t$, $u \in V'(t)$, there is an edge $\{u,v\} \in \mathcal{O}'(t)$ such that $\left\|X_u - X_v\right\|_1 \le c_2 \le \Delta$. So, we just need to verify (7.24) with β replaced by $(\alpha\cdot\gamma)^{1+2\epsilon}$. If $\Delta = \delta$, by (7.35),

$$|\mathrm{Bag}_t| \ge \frac{|\mathcal{B}|\cdot|\mathcal{O}(t)|}{4N} \ge \frac{|\mathcal{O}(t)|\cdot b\cdot\delta}{48\beta\cdot\delta\cdot N} \gtrsim \frac{\tilde{\alpha}\cdot|\mathcal{O}(t)|\cdot N'}{\beta\cdot C_1(\epsilon)\cdot N} \ge (\alpha\cdot\gamma)^{1+\epsilon}\cdot|\mathcal{O}(t)|,$$

where the second inequality uses the assumption $|\mathcal{B}| \ge \frac{b\cdot\delta}{12\beta\cdot\Delta}$, the third inequality uses (7.31) and the last inequality uses $(\alpha\cdot\gamma)^{\epsilon} \lesssim \frac{1}{\beta\cdot C_1(\epsilon)}$. Otherwise, $\Delta = c_2$, by (7.35),

$$\begin{aligned}|\mathrm{Bag}_t| \ge \frac{|\mathcal{B}|\cdot|\mathcal{O}(t)|}{4N} \ge \frac{b\cdot\delta\cdot|\mathcal{O}(t)|}{48\beta\cdot\Delta\cdot N} &\gtrsim \frac{\tilde{\alpha}^{\epsilon}|\mathcal{O}(t)|}{C_1(\epsilon)\beta}\cdot\frac{N'\cdot c_1}{N\cdot c_2}.\\ &\gtrsim \frac{\tilde{\alpha}^{\epsilon}|\mathcal{O}(t)|}{C_1(\epsilon)\beta}\cdot\alpha\cdot\gamma \ge \alpha^{1+2\epsilon}\cdot|\mathcal{O}(t)|. \end{aligned} \tag{7.36}$$

The third inequality follows by (7.32), the fourth inequality uses (7.30), and the last inequality uses the assumption that $(\alpha\cdot\gamma)^{\epsilon} \lesssim \frac{1}{\beta\cdot C_1(\epsilon)}$. Therefore, FBag is a family of $(\alpha\cdot\gamma)^{1+2\epsilon}$ assigned bags of balls with type (Δ, S).

Finally, it remains to verify (7.29) where δ is replaced by Δ. First, we show that $\sum_{t\in S}|\mathrm{Bag}_t| \ge |\mathcal{B}|/2$. This is because by (7.35),

$$\sum_{t\in\mathcal{T}\setminus S}|\mathrm{Bag}_t| \le \sum_{t\in\mathcal{T}}\frac{|\mathcal{O}(t)|\cdot|\mathcal{B}|}{4N} \le |\mathcal{B}|/2.$$

Equation (7.29) follows by (7.34) and the assumption that $|\mathcal{B}| \ge \frac{b\cdot\delta}{12\beta\cdot\Delta}$. □

Construction of a Geometric Sequence of Families of Bags of Balls

In this section we prove proposition 7.18. First, we prove a bucketing lemma. We show that for any α-bad node $t \in \mathcal{T}$, we can extract a $1/\mathrm{poly}\log(\alpha)$-dominating 2-homogeneous set $\mathcal{O}'(t)$ of edges from $\mathcal{O}(t)$.

Lemma 7.22. *For a locally connected hierarchy, $\mathcal{T}$, of G, and an α-bad node $t \in \mathcal{T}$, if α is sufficiently small, then there is a 2-homogeneous set $\mathcal{O}'(t) \subset \mathcal{O}(t)$ such that $\mathcal{O}'(t)$ is $1/C_2(\alpha)$-dominating where $C_2(.)$ is a universal polylogarithmic function.*

Proof. We fix t throughout the proof and use $\mathcal{O}$ instead of $\mathcal{O}(t)$ for brevity. Throughout the proof all probabilities are measured under the uniform distribution on $\mathcal{O}$. Let

$$\begin{aligned} a_e &:= \langle U^e, \mathsf{X}_e\rangle,\\ b_e &:= \left\|\mathsf{X}_e\right\|,\\ \mu &:= \mathbb{E}_{e\sim\mathcal{O}}[a_e]. \end{aligned}$$

Note that since $\left\|U^e\right\| = 1$, $a_e \leq b_e$ for any e. To prove the claim it is enough to find a 2-homogeneous set $\mathcal{O}'$ such that

$$\mathbb{P}[e \in \mathcal{O}']^2 \cdot \min_{e\in\mathcal{O}'} a_e^2 \geq \frac{\mu^2}{C_2(\alpha)}. \tag{7.37}$$

Then, the lemma follows by

$$\Big(\sum_{e\in\mathcal{O}'} a_e\Big)^2 \geq |\mathcal{O}|^2 \cdot \mathbb{P}[e \in \mathcal{O}']^2 \min_{e\in\mathcal{O}'} a_e^2 \geq \frac{|\mathcal{O}|^2 \cdot \mu^2}{C_2(\alpha)} = \frac{1}{C_2(\alpha)} \cdot \big(\sum_{e\in\mathcal{O}} a_e\big)^2.$$

We prove (7.37) as follows: First, we partition the edges into sets $\mathcal{O}_1, \mathcal{O}_2, \dots$ such that for any $e, f \in \mathcal{O}_i$, $a_e \approx a_f$. Then, we show that there is an index i, such that $\mathbb{P}[e \in \mathcal{O}_i] \cdot \min_{e\in\mathcal{O}_i} a_e \gtrsim \frac{\mu}{\log(\alpha)}$ (see (7.40)). Then, we partition $\mathcal{O}_i$ into sets $\mathcal{O}_{i,1}, \mathcal{O}_{i,2}, \dots$ such that any $\mathcal{O}_{i,j}$ is 2-similar. Finally, we show that there is an index j such that $\mathcal{O}_{i,j}$ satisfies (7.37).

For $i \in \mathbb{Z}$ and $c := \sqrt{2}$, define,

$$\mathcal{O}_i := \{e \in \mathcal{O}(t) : c^i \leq a_e/\mu < c^{i+1}\}.$$

We write $\mathcal{O}_{\geq j} = \cup_{i=j}^{\infty} \mathcal{O}_i$. Also, for any i let $a_{\wedge i} = \min_{e\in\mathcal{O}_i} a_e$.

Next, we show that there exists $-4 \leq i < 2(2 + \log(1/\alpha))$ such that $\mathbb{P}[e \in \mathcal{O}_i]\, a_{\wedge i} \gtrsim \mu/\log(1/\alpha)$. First, observe that,

$$\sum_{i=-\infty}^{-6} a_{\wedge i} \cdot \mathbb{P}[e \in \mathcal{O}_i] \leq \sum_{i=-\infty}^{-6} c^{-5}\mu \cdot \mathbb{P}[e \in \mathcal{O}_i] \leq \mu/c^5. \tag{7.38}$$

Let $q = \Theta(\log(1/\alpha))$ be chosen such that $c^q = c^5/\alpha$. Then,

$$\begin{aligned}\frac{c^5\mu}{\alpha} \cdot \sum_{i=q}^{\infty} a_{\wedge i} \cdot \mathbb{P}[e \in \mathcal{O}_i] &\leq \sum_{i=q}^{\infty} a_{\wedge i}^2 \cdot \mathbb{P}[e \in \mathcal{O}_i] \\ &\leq \mathbb{E}_{e\sim\mathcal{O}_{\geq q}}[b_e^2] \cdot \mathbb{P}[e \in \mathcal{O}_{\geq q}] \\ &\leq \mathbb{E}_{e\sim\mathcal{O}}[b_e^2] \leq \frac{\mu^2}{\alpha}. \end{aligned} \tag{7.39}$$

The second inequality uses $a_e \leq b_e$ and the last inequality uses that t is α-bad. Summing up (7.38) and $\alpha/c^5\mu$ of (7.39) we get

$$\sum_{i\geq q \text{ or } i\leq -6} a_{\wedge i} \cdot \mathbb{P}[e \in \mathcal{O}_i] \leq \mu/c^3 \Rightarrow \sum_{i\geq q \text{ or } i\leq -6} a_e \mathbb{P}[e \in \mathcal{O}_i] \leq \mu/2,$$

where we used that for any edge $e \in \mathcal{O}_i$, $a_{\wedge i} \geq a_e/c$. Therefore,

$$\max_{-5\leq i<q} \mathbb{P}[e \in \mathcal{O}_i] \cdot a_{\wedge i} \geq \frac{1}{5+q} \sum_{i=-5}^{q} \mathbb{P}[e \in \mathcal{O}_i]\, a_{\wedge i} \geq \frac{1}{c(5+q)} \sum_{i=-5}^{q} a_e \mathbb{P}[e \in \mathcal{O}_i] \geq \frac{1}{c(5+q)} \cdot \frac{\mu}{2}. \tag{7.40}$$

Let i be the maximizer of the LHS of the above equation. It remains to choose a subset of $\mathcal{O}_i$ such that $b_e^2/b_f^2 < 2$ for all e, f in that subset.

For any integer $j \geq 0$, we define

$$\mathcal{O}_{i,j} := \{e \in \mathcal{O}_i : c^j \leq b_e/a_{\wedge i} < c^{j+1}\}.$$

Note that any set $\mathcal{O}_{i,j}$ is 2-similar. We show that there is an index $j < q$ such that $\mathcal{O}_{i,j}$ satisfies (7.37). Let $\mathcal{O}_{i,\geq q} = \cup_{j=q}^{\infty} \mathcal{O}_{i,j}$. Similar to (7.39),

$$c^{2q} \cdot \mathbb{P}[e \in \mathcal{O}_{i,\geq q}]\, a_{\wedge i}^2 \leq \mathbb{E}_{e \sim \mathcal{O}}[b_e^2] \leq \frac{\mu^2}{\alpha} \leq \frac{1}{\alpha} \cdot 8a_{\wedge i}^2 \cdot (5+q)^2 \cdot \mathbb{P}[e \in \mathcal{O}_i]^2,$$

where the last inequality uses (7.40). Using $c^q = c^5/\alpha$, we obtain

$$\mathbb{P}[e \in \mathcal{O}_{i,\geq q}] \leq \frac{\alpha}{4} \cdot (5+q)^2 \cdot \mathbb{P}[e \in \mathcal{O}_i]^2 \leq \frac{1}{2} \cdot \mathbb{P}[e \in \mathcal{O}_i]^2,$$

for a sufficiently small α. Now, let $j = \operatorname{argmax}_{0 \leq j < q} \mathbb{P}\left[e \in \mathcal{O}_{i,j}\right]$. Then,

$$\mathbb{P}\left[e \in \mathcal{O}_{i,j}\right]^2 \cdot a_{\wedge i}^2 \geq \frac{a_{\wedge i}^2}{q^2} \cdot (\mathbb{P}[e \in \mathcal{O}_i] - \mathbb{P}[e \in \mathcal{O}_{i,\geq q}])^2 \geq \frac{\mathbb{P}[e \in \mathcal{O}_i]^2 \cdot a_{\wedge i}^2}{4q^2} \geq \frac{\mu^2}{32q^2(5+q)^2}.$$

The last inequality uses (7.40). Now, (7.37) follows by the above inequality and $C_2(\alpha) = 32q^2(5+q)^2$ and $\mathcal{O}'(t) = \mathcal{O}_{i,j}$; □

Now, we are ready to prove proposition 7.18. First, by lemma 7.22 for each α-bad node $t \in T$, there is a 2-homogeneous γ-dominating set $\mathcal{O}'(t) \subseteq \mathcal{O}(t)$ where $\gamma = 1/C_2(\alpha)$. For each $t \in T$, let

$$a_t = \min_{e \in \mathcal{O}'(t)} \langle U^e, \mathsf{X}_e \rangle^2 \text{ and } b_t = \min_{e \in \mathcal{O}'(t)} \left\|\mathsf{X}_e\right\|_2^2.$$

Let $\tilde{\lambda} < 1$ be a function of λ that we fix later. For any integer $i \in \mathbb{Z}$, let

$$T_i := \{t \in T : \tilde{\lambda}^{i+1/2} \leq a_t < \tilde{\lambda}^{i-1/2}\}$$

Note that, by definition, for all $i \neq j$, $T_i \cap T_j = \emptyset$.

Next, we partition the bad nodes of each T_i into sets T_{i,j_a,j_b} such that each set $\cup_{t \in T_{i,j_a,j_b}} \mathcal{O}'(t)$ is 4-homogeneous. We will apply lemma 7.21 to the T_{i,j_a,j_b} with the largest contribution in the numerator. This will give us a family of either compact or assigned bags of balls. Then, we will drop the bags for odd (or even) i randomly. Since for any $t \in T_i$, $t' \in T_{i+2}$, $a_{t'} < \tilde{\lambda} a_t$ we will obtain a $\tilde{\lambda}$-geometric sequence of bags of balls.

First, we partition the nodes of each T_i into sets T_{i,j_a,j_b}; for all integers $0 \leq j_a$ and $0 \leq j_b$ let

$$T_{i,j_a,j_b} := \{t \in T_i : 2^{j_a} \leq \frac{a_t}{\tilde{\lambda}^{i+1/2}} < 2^{j_a+1}, 2^{j_b} \leq \frac{b_t}{a_t} < 2^{j_b+1}\}.$$

Observe that for all i, j_a, j_b, $\cup_{t\in T_{i,j_a,j_b}}\mathcal{O}'(t)$ is 4-homogeneous. Note that by the definition of T_i, for $j_a > \log(1/\tilde{\lambda})$, $T_{i,j_a,\cdot} = \emptyset$. On the other hand, since t is α-bad and $\mathcal{O}'(t)$ is γ-dominating, $a_t \gtrsim \alpha\gamma b_t$ (see (7.30)); so for $j_b > \log(1/\alpha\gamma) + O(1)$, $T_{i,\cdot,j_b} = \emptyset$. Therefore, for any i, the number of nonempty sets T_{i,j_a,j_b} is at most $O(\log(1/\tilde{\lambda}\alpha\gamma))$.

For a set $S \subseteq T$, let

$$\Pi(S) := \sum_{t\in S}\frac{1}{|\mathcal{O}(t)|}\Big(\sum_{e\in\mathcal{O}'(t)}\langle U^e, \mathbf{X}_e\rangle\Big)^2.$$

For each T_i let

$$T_i^* = \mathrm{argmax}_{T_{i,j_a,j_b}}\Pi(T_{i,j_a,j_b}).$$

Since any $t \in T_i^*$ is α-bad and $\mathcal{O}'(t)$ is γ-dominating, and $\cup_{t\in T_i^*}\mathcal{O}'(t)$ is 4-homogeneous, and by the lemma's assumption

$$(\gamma\alpha)^\epsilon = \frac{\alpha^\epsilon}{C_2(\alpha)^\epsilon} \lesssim \frac{1}{\beta\cdot C_1(\epsilon)},$$

we may invoke lemma 7.21 for each set T_i^*. This gives us either a family of β-compact bags of balls FBag_i with type (δ_i, Δ_i), or a family of $(\alpha\gamma)^{1+2\epsilon}$-assigned bags of balls, FBag_i of type (δ_i, S_i^*) where $S_i^* \subseteq T_i^*$. These families satisfy two additional constraints: Firstly, $\delta_i, \Delta_i = \min_{t\in T_i^*} a_t$ up to an $O(\alpha\gamma)$ factor, secondly, the sum of the radii of all balls in the family is at least $\frac{(\alpha\gamma)^\epsilon}{\beta C_1(\epsilon)}\Pi(T_i^*)$.

We remove half of the families to obtain a geometric sequence. First, by the definition of T_i,

$$\tilde{\lambda}\cdot\min_{t\in T_i^*} a_t \geq \min_{t\in T_{i+2}^*} a_t.$$

This means that if we remove families for either odd or even i's, then the decaying rate of $\min_{t\in T_i^*} a_t$ is at least $\tilde{\lambda}$. Therefore by the properties guaranteed by lemma 7.21, and the above fact, any subsequence of odd or even compact or assigned families of bags of balls is $O(\tilde{\lambda}/(\alpha\cdot\gamma)^2)$-geometric. Setting $\tilde{\lambda} \asymp \lambda\cdot(\alpha\cdot\gamma)^2$ produces λ-geometric sequences.

Without loss of generality we assume that $\Pi(\cup_i T_{2i}) \geq \Pi(\cup_i T_{2i+1})$. Drop the families for odd i; consider the sum of radii of balls in the remaining compact families and in the remaining assigned families; one of them is greater. We let this be our λ-geometric family.

It remains to verify (7.26) and (7.27). By lemma 7.21, the sum of the radii in the constructed geometric sequence is at least $\gtrsim \frac{(\alpha\cdot\gamma)^\epsilon}{\beta C_1(\epsilon)}\sum_i \Pi(T_{2i}^*)$. By the definition of T_i^*,

$$\sum_i \Pi(T_{2i}^*) \gtrsim \frac{1}{|\log(\tilde{\lambda}\alpha\gamma)|}\sum_i \Pi(T_{2i}) \geq \frac{\Pi(T)}{|\log(\lambda\mathrm{poly}(\alpha))|}.$$

Now, since each $\mathcal{O}'(t)$ is $\gamma = 1/C_2(\alpha)$-dominating,

$$\Pi(T) \geq \frac{1}{C_2(\alpha)}\cdot\sum_{t\in T}\frac{1}{|\mathcal{O}(t)|}\cdot\Big(\sum_{e\in\mathcal{O}(t)}\langle U^e, \mathbf{X}_e\rangle\Big)^2.$$

7.3 Lower-bounding the Denominator of the Dual

In this part we upper-bound the sum of radii of balls in a geometric sequence. Throughout this section we use $C_3, C_4 > 0$ as large universal constants. The following two propositions are the main statements that we prove in this section.

Proposition 7.23. *Given a k-edge-connected graph G, and a λ-geometric sequence of families of C_3-compact bags of balls* $\mathrm{FBag}_1, \mathrm{FBag}_2, \ldots$ *where* FBag_i *has type (δ_i, Δ_i), if $\lambda \le 1/12$ and $C_3 \ge 36$, then*

$$\frac{k}{4} \cdot \sum_i \delta_i \sum_{\mathrm{Bag} \in \mathrm{FBag}_i} |\,\mathrm{Bag}\,| \le \sum_{\{u,v\} \in E} \left\| X_u - X_v \right\|_1 .$$

Proposition 7.24. *Given a $(k, k \cdot \lambda, T)$-LCH, $\mathcal{T}$, of G and a λ-geometric sequence of families of $24C_3/k$-assigned bags of balls,* $\mathrm{FBag}_1, \mathrm{FBag}_2, \ldots$ *such that each* FBag_i *is of type (δ_i, T_i) where $T_i \subseteq T$, if $C_4 \ge 3$, $\lambda \le 1/6C_4$ and $C_3 \ge 2((C_4+1) + 4(C_4+2)^2)$, then*

$$\frac{k}{8} \cdot \frac{C_4}{12C_3} \cdot \sum_i \delta_i \sum_{t \in T_i} |\,\mathrm{Bag}_t\,| \le \sum_{\{u,v\} \in E} \left\| X_u - X_v \right\|_1 .$$

Note that in the above proposition, the assumption $\lambda \le 1/6C_4$ follows from $k \cdot \lambda < 1$.

First, we use the above propositions to finish the proof of theorem 7.4. Recall theorem 7.4:

Theorem 7.4. *For any k-edge-connected graph $G = (V, E)$ and any (k, λ, T)-LCH, of G, and for $h > 0$, any cut metric $X \in \{0,1\}^{h \times V}$, and any semiorthogonal matrix $U \in \mathbb{R}^{E \times h}$,*

$$\frac{\sum_{t \in T} \frac{1}{|\mathcal{O}(t)|} \left(\sum_{e \in \mathcal{O}(t)} \langle U^e, \mathsf{X}_e \rangle \right)^2}{\sum_{e \in E} \left\| \mathsf{X}_e \right\|^2} \le \frac{f_1(k, \lambda)}{k}. \tag{7.14}$$

Proof. Let $T_{\alpha\text{-bad}} \subseteq T$ be the set of α-bad nodes for a parameter α that we set below. It follows that,

$$\alpha \ge \frac{\sum_{t \in T \setminus T_{\alpha\text{-bad}}} \frac{1}{|\mathcal{O}(t)|} \cdot \left(\sum_{e \in \mathcal{O}(t)} \langle U^e, \mathsf{X}_e \rangle \right)^2}{\sum_{t \in T \setminus T_{\alpha\text{-bad}}} \sum_{e \in \mathcal{O}(t)} \left\| \mathsf{X}_e \right\|_1} \ge \frac{\sum_{t \in T \setminus T_{\alpha\text{-bad}}} \frac{1}{|\mathcal{O}(t)|} \cdot \left(\sum_{e \in \mathcal{O}(t)} \langle U^e, \mathsf{X}_e \rangle \right)^2}{2 \sum_{e \in E} \left\| \mathsf{X}_e \right\|_1}. \tag{7.41}$$

The second inequality uses the fact that each edge is in at most two sets $\mathcal{O}(t)$.

We apply proposition 7.18 to $T_{\alpha\text{-bad}}$. Let $C_4 = 3$, $\beta = 36$ and $C_3 = 104$. We choose $\alpha = \Theta(\mathrm{poly}\log(k)/k)$, $\epsilon = \Theta(\log\log(k)/\log(k))$ such that the following conditions are satisfied

$$\left(\frac{\alpha}{C_2(\alpha)} \right)^{\epsilon} \lesssim \frac{1}{\beta \cdot C_1(\epsilon)},$$

$$\left(\frac{\alpha}{C_2(\alpha)} \right)^{1+2\epsilon} \ge \frac{24C_3}{k}.$$

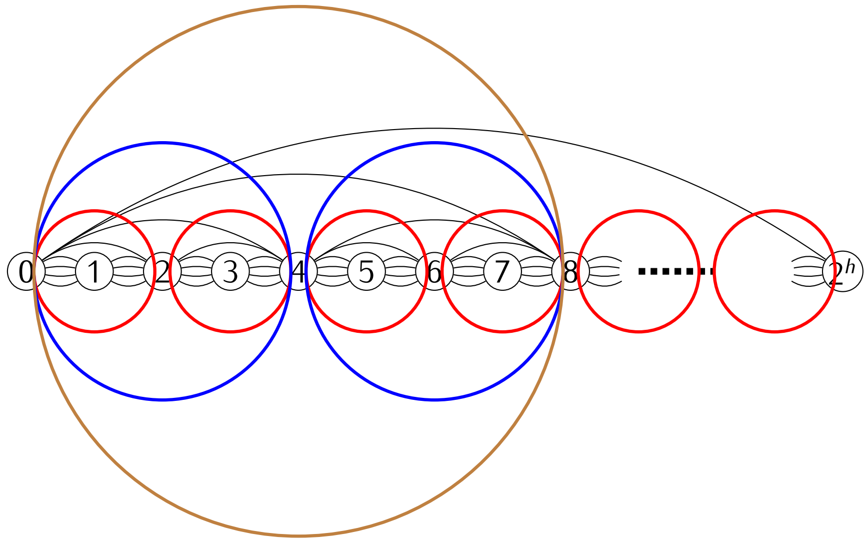

Figure 7.2: Consider the natural L_1 mapping of the graph of fig. 3.4 where vertex i is mapped to the number i. Consider h layers of L_1 balls as shown above where the radii of all balls in layer i is 2^i and they are disjoint. Although the sum of the radii of all balls in this family is $\Theta(n \cdot h)$, the sum of the L_1 lengths of the edges of G is $n \cdot (h + k)$.

Recall that $C_1(\epsilon)$ is an inverse polynomial of ϵ and $C_2(\alpha)$ is a polylogarithmic function of α so the above assignment is feasible. Also let $\tilde{\lambda} < \lambda/k$ be such that $\tilde{\lambda} < 1/6C_4$.

Now, by proposition 7.18 either there is a $\tilde{\lambda}$-geometric sequence of 36-compact bags of balls FBag_1, FBag_2, ..., that satisfies (7.26), or there is a $\tilde{\lambda}$-geometric sequence of $24C_3/k$-assigned bags of balls FBag_1, FBag, ..., that satisfies (7.27). Now, by proposition 7.23 and proposition 7.24 we get

$$\frac{\sum_{t \in T_{\alpha\text{-bad}}} \frac{1}{|\mathcal{O}(t)|} \cdot \left(\sum_{e \in \mathcal{O}(t)} \langle U^e, \mathsf{X}_e \rangle \right)^2}{\sum_{e \in E} \left\| \mathsf{X}_e \right\|_1} \lesssim \frac{C_1(\epsilon) C_2(\alpha) \cdot |\log(\tilde{\lambda}\text{poly}(\alpha))|}{k \cdot (\alpha / C_2(\alpha))^\epsilon}$$

The theorem follows from the above equation together with (7.41). □

In the rest of this section we prove above propositions. Before getting into the proofs, we give a simple example to show that, in order to bound the denominator, it is necessary to use that the given λ-geometric sequence of bags of balls is either compact or assigned. The following example is designed based on the dual solution that we constructed in theorem 7.3.

Example 7.25. *Let G be the graph illustrated in fig. 3.4, and let $X_0, X_1, \ldots, X_{2^h}$ be an embedding of G where $X_i = \mathbf{1}_{[i]}$. Now, for any $1 \leq j \leq h-1$, let Bag_j be the union of balls*

$$B(X_{2^j}, 2^j), B(X_{3 \cdot 2^j}, 2^j), B(X_{5 \cdot 2^j}, 2^j), \ldots, B(X_{2^h - 2^j}, 2^j).$$

1. We process bags of balls in phases; we assume that phase ℓ starts at time $\tau_{\ell-1}+1$ and ends at τ_ℓ. In phase ℓ we process the bags in FBag_ℓ; in other words, we process larger balls earlier than smaller ones. In each time step (except the last one) of phase ℓ we process exactly one bag of FBag_ℓ.

2. In addition to adding new balls, in each phase we may shrink or delete some of the already inserted (hollowed) balls but when we insert a ball of FBag_ℓ we never alter it until after the end of phase ℓ.

3. We keep the invariant that for any τ, all (hollowed) balls in $\mathcal{Z}_\tau$ are disjoint. This crucial property will not hold in our construction of the assigned bags of balls in the next section and it is the main reason that our second construction is more technical.

4. For any hollowed ball $B(x, r_1||r_2) \in \mathcal{Z}_\tau$, there are vertices $u, v \in V$ such that $\left\|x - X_u\right\|_1 \leq r_1$ and $\left\|x - X_v\right\|_1 \geq r_2$.

Figure 7.3: Properties of the inductive charging argument for compact bags of balls.

Note that the center of each of these balls is a vertex of G and that for any j, all balls of Bag_j *have equal radius and are disjoint (see fig. 7.2). So we get a* $1/2$*-geometric sequence of bags of balls (and similarly we can obtain a λ-geometric sequence by letting j be multiples of* $\log(1/\lambda)$*). As alluded to in the proof of theorem 7.3, the sum of the radii of balls in the given sequence is $h \cdot 2^h$ while the sum of the L_1 lengths of edges of G is only $(h+k)\cdot 2^h$.*

The above example serves as a crucial barrier to both of our proofs. In the proof of proposition 7.23 we bypass this barrier using the compactness of bags of balls. Note that in the above example Bag_j is not compact, and indeed the diameter of centers of balls of Bag_j is 2^h which is the same as the sum of the radii of balls in Bag_j. In the proof of proposition 7.24 we bypass the above barrier using the properties of the locally connected hierarchy.

Charging Argument for Compact Bags of Balls

In this section we prove proposition 7.23. We construct a set of *disjoint* L_1 hollowed balls inductively from the given compact bags of balls. For any integer $\tau \geq 0$, we use $\mathcal{Z}_\tau$ to denote the set of hollowed balls in the construction at time τ. Initially, we have $\mathcal{Z}_0 = \emptyset$ and $\mathcal{Z}_\infty$ is the final construction. We describe the main properties of our construction in fig. 7.3.

Inductive Charging. Before explaining our construction, we describe our inductive charging argument. First, by the following lemma, in our construction, we only need to lower-bound the sum of the widths of all hollowed balls of $\mathcal{Z}_\infty$ by (a constant multiple of) the sum of radii of all balls in the given sequence of compact bags of balls.

Lemma 7.26. *For any $\tau \geq 0$,*

$$k \cdot \sum_{B(x,r_1||r_2)\in\mathcal{Z}_\tau} (r_2 - r_1) \leq \sum_{\{u,v\}\in E} \left\|X_u - X_v\right\|_1 .$$

Proof. We simply use the k-edge-connectivity of G. First, by property 4 of fig. 7.3 for each hollowed ball $B = B(x, r_1||r_2) \in \mathcal{Z}_\tau$ there are vertices $u, v \in V$ such that $\left\|x - X_u\right\|_1 \leq r_1$ and $\left\|x - X_v\right\|_1 \geq r_2$. Since G is k-edge-connected, there are at least k edge-disjoint paths between u, v. Each of these paths must cross B and, by the triangle inequality, the length of the intersection with B is at least $r_2 - r_1$. Finally, since by property 3 of fig. 7.3, balls of $\mathcal{Z}_\tau$ are disjoint, this argument does not overcount the L_1-length of any edge of G. □

Suppose at the end of our construction, we allocate $r_2 - r_1$ tokens to any hollowed ball $B(x, r_1||r_2) \in \mathcal{Z}_\infty$. Our goal is to distribute these tokens between all bags of balls such that each bag, Bag, of type (δ_i, Δ_i) receives at least $|\text{Bag}| \cdot \delta_i/4$ tokens. We prove this by an induction on τ. Suppose $\tau_{\ell-1} < \tau \leq \tau_\ell$; for a hollowed ball $B(x, r_1||r_2) \in \mathcal{Z}_\tau$, define

$$\text{token}_\tau(B) := \begin{cases} \delta_\ell - 6\Delta_{\ell+1} & \text{if } B \in \text{FBag}_\ell \\ [(r_2 - r_1) - 6\Delta_\ell]^+ & \text{otherwise.} \end{cases} \tag{7.42}$$

Instead of allocating $r_2 - r_1$ tokens to a ball at time τ, we allocate $\text{token}_\tau(B)$. The term $6\Delta_\ell$ takes into account the fact that we shrink balls in $\mathcal{Z}_\tau$ later in the post processing phase. We prove the following lemma inductively.

Lemma 7.27. *At any time $\tau_{\ell-1} + 1 \leq \tau \leq \tau_\ell$, if we allocate* $\text{token}_\tau(B)$ *tokens to any hollowed ball $B \in \mathcal{Z}_\tau$, then we can distribute these tokens among the bags of balls that we processed by time τ such that each* Bag *of type (δ_i, Δ_i) receives at least $\delta_i \cdot |\text{Bag}|/4$ tokens.*

It is easy to see that proposition 7.23 follows by applying the above lemma to the final set of hollowed balls $\mathcal{Z}_\infty$ and using lemma 7.26, since

$$\frac{1}{4}\sum_i \sum_{\text{Bag}\in\text{FBag}_i} \delta_i \cdot |\text{Bag}| \leq \sum_{B(x,r_1||r_2)\in\mathcal{Z}_\tau} r_2 - r_1 \leq \frac{1}{k} \cdot \sum_{\{u,v\}\in E} \left\|X_u - X_v\right\|_1 .$$

Construction. It remains to prove lemma 7.27. First, we need some definitions. We say a ball $B = B(X_u, \delta_\ell) \in \text{FBag}_\ell$ is in the *interior* of a hollowed ball $B' = B(x, r_1||r_2) \in \mathcal{Z}_\tau$ if

$$r_1 + \delta_\ell + \Delta_\ell \leq \left\|X_u - x\right\|_1 \leq r_2 - \delta_\ell - \Delta_\ell .$$

Note that B is inside B' when $r_1 + \delta_\ell \leq \left\|X_u - x\right\|_1 \leq r_2 - \delta_\ell$; so a ball B may be inside B' but not in the interior of B'. If such a B' exists, we call B an interior ball. If B is not an interior ball, we call it a border ball. Since hollowed balls in $\mathcal{Z}_\tau$ are disjoint, B can be in the interior of at most one hollowed ball of $\mathcal{Z}_\tau$.

Fact 7.28. *Any ball $B \in \mathrm{FBag}_\ell$ is in the interior of at most one hollowed ball of $\mathcal{Z}_\tau$.*

Suppose lemma 7.27 holds at time $\tau > \tau_{\ell-1}$; we show it also holds at time $\tau+1$. At time τ, we process a bag of balls in FBag_ℓ that has at least one interior ball (and is not processed yet); if there is no such bag then we run the post processing algorithm that we will describe later. Suppose at time τ we are processing $\mathrm{Bag}^* = \{B_1 = B(X_{u_1}, \delta_\ell), \dots, B_b = B(X_{u_b}, \delta_\ell)\}$ of FBag_ℓ and assume that one of these balls, say B_1, is in the interior of a hollowed ball $B(x, r_1 || r_2) \in \mathcal{Z}_\tau$.

First, we show that all balls of Bag^* are inside of B. Let

$$r_1' = \min_{1 \le i \le b} \left\|x - X_{u_i}\right\|_1 \text{ and } r_2' = \max_{1 \le i \le b} \left\|x - X_{u_i}\right\|_1$$

It follows that

$$r_2' \le \left\|x - X_{u_1}\right\|_1 + \Delta_\ell \le (r_2 - \delta_\ell - \Delta_\ell) + \Delta_\ell \le r_2 - \delta_\ell,$$

where we used (7.22); similarly, $r_1' \ge r_1 + \delta_\ell$. Therefore, all balls of Bag^* are inside of B and by property 3 of fig. 7.3 they do not touch any other (hollowed) ball of $\mathcal{Z}_\tau$.

Now, we construct $\mathcal{Z}_{\tau+1}$. We remove B and we add two new hollowed balls $B_1' = B(x, r_1 || r_1' - \delta_\ell)$ and $B_2' = B(x, r_2' + \delta_\ell || r_2)$. In addition, we add all of the balls of Bag^* (see fig. 7.4). It is easy to see that balls in $\mathcal{Z}_{\tau+1}$ are disjoint. We send $\delta_\ell/4$ tokens of each of $B_1, \dots, B_b$ to Bag^*. We send the rest of their tokens and all of the tokens of B_1', B_2' to B and we re-distribute them by the induction hypothesis. It follows that Bag^* receives exactly $b \cdot \delta_\ell/4$ tokens and B receives $\mathrm{token}_\tau(B)$.

$$\begin{aligned}
\mathrm{token}_{\tau+1}(B_1') + \mathrm{token}_{\tau+1}(B_2') + \sum_{i=1}^{b} \mathrm{token}_{\tau+1}(B_i)& \\
&\ge r_2 - r_1 - (r_2' - r_1') - 2\delta_\ell - 12\Delta_\ell + b \cdot (\delta_\ell - 6\Delta_{\ell+1}) \\
&\ge \mathrm{token}_\tau(B) + b \cdot \delta_\ell(1 - 6\lambda) - 7\Delta_\ell \\
&\ge \mathrm{token}_\tau(B) + b \cdot \delta_\ell/2 - C_3\Delta_\ell/4 \\
&\ge \mathrm{token}_\tau(B) + b \cdot \delta_\ell/4.
\end{aligned}$$

where the first inequality uses (7.42), the second inequality uses $\Delta_{\ell+1} \le \lambda \cdot \delta_\ell$ and $\Delta_\ell \ge 2\delta_\ell$, the third inequality uses that $\lambda < 1/12$ and $C_3 \ge 28$. The last inequality uses that Bag^* is C_3-compact, i.e., (7.23); this is the only place that we use the compactness of Bag^*. Therefore, lemma 7.27 holds at time $\tau + 1$.

Post Processing. Let τ_ℓ be the time by which we have processed all bags of FBag_ℓ with at least one interior ball, and let FBag_ℓ' be the set of bags that we have not processed yet, i.e., all balls of FBag_ℓ' are border balls with respect to $\mathcal{Z}_{\tau_\ell}$. As alluded to, at the end of phase ℓ, i.e., at time τ_ℓ, we shrink all (hollowed) balls of $\mathcal{Z}_\tau$ except those that were in FBag_ℓ. Given a

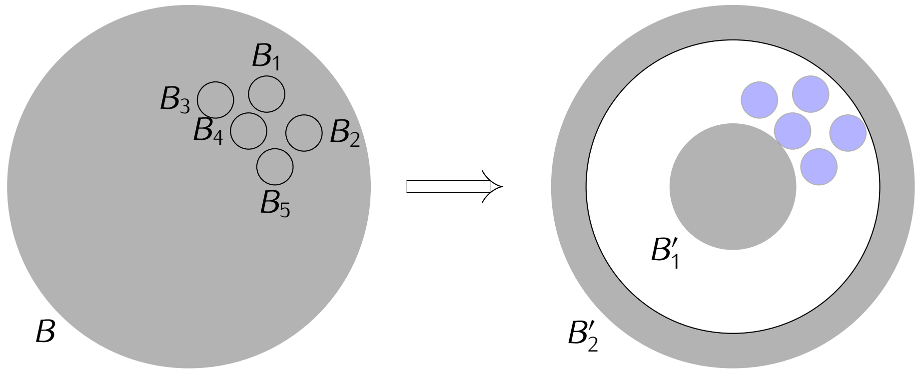

Figure 7.4: Balls $B_1, \dots, B_5$ represent the balls of Bag^*; B_1 is in the interior of a ball $B \in \mathcal{Z}_\tau$. We decompose B into two hollowed balls, B_1', B_2' that do not intersect any of the balls in the given compact set as shown on the right.

hollowed ball $B = B(x, r_1 || r_2) \in \mathcal{Z}_{\tau_\ell}$, the shrink_ℓ operator is defined as follows:

$$\mathrm{shrink}_\ell(B) := \begin{cases} B & \text{if } B \in \mathrm{FBag}_\ell \\ B(x, r_1 + 2\delta_\ell + \Delta_\ell || r_2 - 2\delta_\ell - \Delta_\ell) & \text{if } B \notin \mathrm{FBag}_\ell \text{ and } r_2 - r_1 > 2\Delta_\ell + 4\delta_\ell \\ B(x, 0) = \emptyset & \text{otherwise.} \end{cases} \tag{7.43}$$

At time τ_ℓ, for any hollowed ball $B \in \mathcal{Z}_{\tau_\ell}$ we add $\mathrm{shrink}_\ell(B)$ to $\mathcal{Z}_{\tau_\ell+1}$. In addition, we add all balls of all bags of FBag_ℓ' to $\mathcal{Z}_{\tau_\ell+1}$. This is the end of phase ℓ and we consider $\mathcal{Z}_{\tau+1}$ as our construction in the beginning of phase $\ell + 1$.

Let us verify that balls of $\mathcal{Z}_{\tau+1}$ are disjoint, i.e., $\mathcal{Z}_{\tau+1}$ satisfies property 3 of fig. 7.3. For any hollowed ball $B = B(x, r_1 || r_2) \in \mathcal{Z}_{\tau_\ell}$ and ball $B' = B(X_{u'}, \delta_\ell) \in \mathrm{FBag}_\ell'$, we show that $\mathrm{shrink}_\ell(B)$ and B' do not intersect. First, if $B \in \mathrm{FBag}_\ell$, then $\mathrm{shrink}_\ell(B) = B$, by definition 7.14 any two balls of FBag_ℓ do not intersect, so $\mathrm{shrink}_\ell(B), B'$ do not intersect. Now, suppose $B \notin \mathrm{FBag}_\ell$. Since $B' \in \mathrm{FBag}_\ell'$, B' is not in the interior of B, i.e., either $\left\|x - X_{u'}\right\|_1 < r_1 + \delta_\ell + \Delta_\ell$ or $\left\|x - X_{u'}\right\|_1 > r_2 - \delta_\ell - \Delta_\ell$. In both cases, B' does not intersect $\mathrm{shrink}_\ell(B)$.

It remains to distribute the tokens. We send all tokens of all balls of all bags of FBag_ℓ' to their corresponding bag. Therefore, any $\mathrm{Bag} \in \mathrm{FBag}_\ell'$, receives at least

$$b \cdot (\delta_\ell - 6\Delta_{\ell+1}) \geq b \cdot \delta_\ell(1 - 6\lambda) \geq b \cdot \ \delta_\ell/2$$

tokens. In addition, for every hollowed ball $B \in \mathcal{Z}_{\tau_\ell}$, we send all tokens of $\mathrm{shrink}_\ell(B)$ to B and we redistribute by induction. Since

$$\mathrm{token}_{\tau_\ell}(B) \leq \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B)),$$

B receives at least the same number of tokens. This completes the proof of proposition 7.23.

Charging Argument for Assigned Bags of Balls

In this part we prove proposition 7.24. Before getting into the details of the proof we illustrate the ideas we use to bypass the barrier of example 7.25. The first observation is that, unlike the previous section, we cannot construct a family of *disjoint* hollowed balls in $\mathcal{Z}_\infty$ in such a way that the sum of widths of hollowed balls of $\mathcal{Z}_\infty$ is a constant fraction of the sum of radii of all balls in the given geometric sequence. Instead, we let hollowed balls of $\mathcal{Z}_\infty$ intersect and we employ a ball labeling technique that uses the locally connected hierarchy, $\mathcal{T}$.

Let us give a simple example to show the crux of our analysis. Suppose a node $t_1 \in \mathcal{T}$ has exactly two children, t_2, t_3. Say at time $\tau_{\ell-1} < \tau \leq \tau_\ell$ we are processing Bag_{t_2}. Suppose $\mathcal{Z}_\tau$ has a large ball $B = B(x, r) \in \mathrm{Bag}_t$ as shown on the left side of fig. 7.5 such that t is an ancestor of t_1. Say Bag_{t_2} has four balls $B_1, \ldots, B_4$. Because Bag_{t_2} is not compact, if we remove the part of B that intersects with balls of Bag_{t_2} and add $B_1, \ldots, B_4$, the sum of the widths of hollowed balls in $\mathcal{Z}_{\tau+1}$ is the same as that sum in $\mathcal{Z}_\tau$, and therefore we gain nothing from adding balls of Bag_t. Instead, we add a new ball that intersects $B_1, \ldots, B_4$ as shown on the right side of fig. 7.5.

Say the center of each B_i is X_{u_i} for $u_i \in V(t_2)$; each X_{u_i} corresponds to a blue dot in fig. 7.5. By the definition of assigned bags of balls, definition 7.13, for each i there is a vertex $v_i \in V(t_1) \setminus V(t_2) = V(t_3)$ such that $\left\|X_{u_i} - X_{v_i}\right\|_1 \leq \delta_\ell$ (each X_{v_i} corresponds to a red dot in fig. 7.5). We add all balls of Bag_{t_2} and a new hollowed ball centered at x, the center of B, ranging from the closest red vertex to x to the farthest one. We also break B into two hollowed balls and remove the part of it that intersects either of these 5 new (hollowed) balls.

Observe that, the sum of the widths of hollowed balls of $\mathcal{Z}_{\tau+1}$ is $\Omega(\delta_\ell \cdot |\mathrm{Bag}_{t_2}|)$ more than this sum in $\mathcal{Z}_\tau$. The only problem is that, the balls of $\mathcal{Z}_{\tau+1}$ are intersecting. So, it is not clear if analogous to lemma 7.26, we can charge the sum of the widths of hollowed balls of $\mathcal{Z}_{\tau+1}$ to the sum of L_1 lengths of edges of G. Our idea is to label hollowed balls with different subsets of edges of G. Although the red hollowed ball and the blue balls intersect, we charge their widths to disjoint subsets of edges of G; we charge the width of the red ball with k edge-disjoint paths supported on $G[V(t_1) \setminus V(t_2)]$ going across this hollowed ball and we charge the radius of each blue ball with k edge-disjoint paths supported on $G(t_2)$ going across that ball.

We remark that the above idea is essentially the main new operation we need for the charging argument, compared to the argument for the compact bags of balls. One of the main obstacles in using this idea is that t_1 can have more than two children. In that case $G[V(t_1) \setminus V(t_2)]$ is not necessarily k-edge-connected. To overcome this, we find a natural decomposition of $G[V(t_1) \setminus V(t_2)]$ into $k/4$-edge-connected components; since each assigned bag of balls, Bag_t has $\gg \mathcal{O}(t)/k$ balls, the centers of a large number of balls of Bag_t are neighbors of one of these components; so we can charge the red ball in the above argument by $k/4$ edge-disjoint paths in that component.

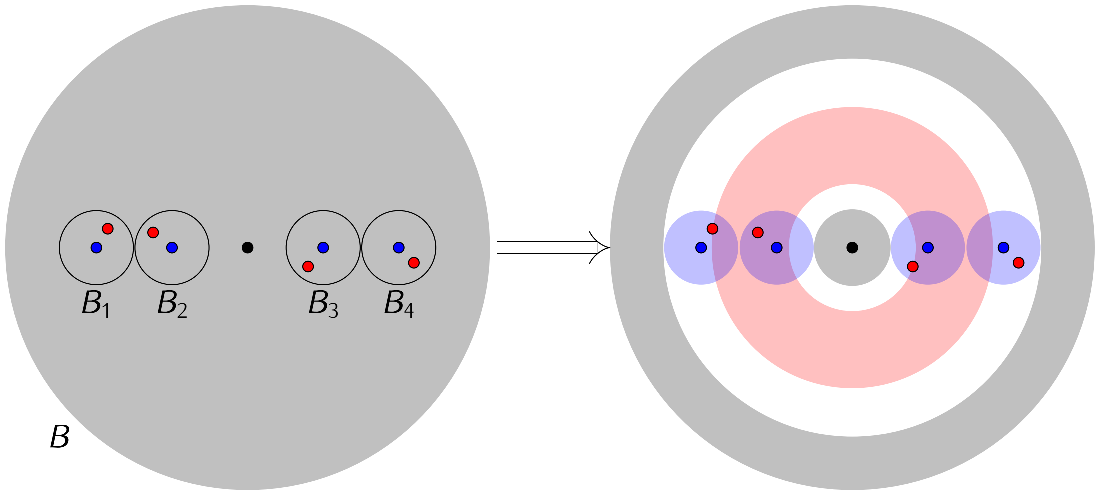

Figure 7.5: A simple example of the ball labeling technique. The grey (hollowed) ball B on the left is one of the hollowed balls of $\mathcal{Z}_\tau$. Small L_1 balls with blue vertices as their centers represent balls of Bag_{t_2} that we are processing at time τ. Each red vertex together with the closest blue vertex are the endpoints of an edge of $\mathcal{O}(t_2)$. The right figure shows new balls added to $\mathcal{Z}_{\tau+1}$. In particular, each blue vertex is in $V(t_2)$ and each red vertex is in $V(t_3)$ where t_2, t_3 are the only children of t_1.

Ball Labeling

In this part we define a valid labeling of hollowed balls in our construction (see fig. 7.7). In the proof of proposition 7.23, we used the disjointness property of balls in the construction in two places; namely in the proofs of lemma 7.26 and Fact 7.28. We address both of these issues by our ball labeling technique.

Basic Label. In the proof of lemma 7.26 we used the disjointness property to charge the sum of the widths of hollowed balls of a set $\mathcal{Z}_\tau$ to the sum of the L_1 lengths of edges of G with no overcounting. Let us give a simple example to show the difficulty in extending this argument to the new setting where balls may intersect. Suppose $\mathcal{Z}_\tau$ is a union of 10 identical copies of $B(x, r)$ with the guarantee that there is a vertex of G at x and one at distance r of x. Then, the sum of the L_1 lengths of edges of G can be as small as $k \cdot r$, as G may just be k-edge-disjoint paths from a vertex at x to a vertex at distance r of x.

A hollowed ball $B = B(x, r_1||r_2)$, can be labeled with $t \in \mathcal{T}$, denoted by $t(B) = t$, if there are vertices $u, v \in V(t)$ such that $\left\|x - X_u\right\|_1 \leq r_1$ and $\left\|x - X_v\right\|_1 \geq r_2$. Recall that, by the definition of $\mathcal{T}$, for any node $t \in \mathcal{T}$, $G(t)$ is k-edge connected. Therefore, if B is labeled with t, then k edge-disjoint paths supported on $E(t)$ cross B. For any ball $B \in \text{Bag}_t$ we let $t(B) = t$. Furthermore, when we shrink or divide a ball into smaller ones the label of the shrunk ball or the new subdivisions remain unchanged.

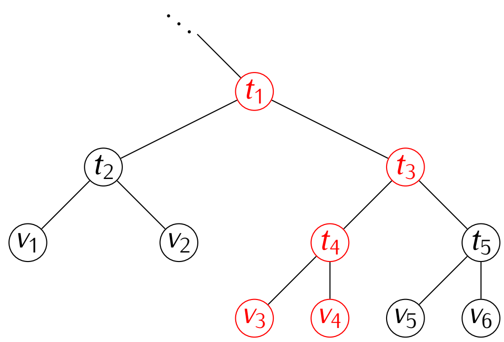

Figure 7.6: The red nodes represent the conflict set of a ball B with $t(B) = t_1$, i.e., $\mathcal{C}(B) = \{t_1, t_3, t_4, v_3, v_4\}$. The edge-disjoint paths of B can be routed in the induced subgraph $G[\{v_3, v_4\}]$.

The simplest definition of the validity of the ball labeling is to make sure that for any two intersecting balls B and B', $t(B)$ and $t(B')$ are not ancestor-descendant. Unfortunately, this simple definition is not enough for our inductive argument, and as we elaborate next, we will enrich the label of some of the balls B by "disallowing routing through some of the descendants of $t(B)$". Recall that $t, t' \in \mathcal{T}$ are *ancestor-descendant* if either t is a weak ancestor of t' or t' is a weak ancestor of t. Recall that t is a weak ancestor of t' if either t is an ancestor of t' or $t = t'$.

To this end, we define a conflict set, $\mathcal{C}(B)$ to be a *connected* subset of the nodes of $\mathcal{T}$ rooted at $t(B)$ (see fig. 7.6). In a valid ball labeling, we make sure that for any two intersecting hollowed balls B and B', $\mathcal{C}(B) \cap \mathcal{C}(B') = \emptyset$. For example, if $t(B), t(B')$ are not ancestor-descendant this condition is always satisfied. In the charging argument, we may only charge the width of B with edge-disjoint paths supported on the leaves of $\mathcal{T}$ which are in $\mathcal{C}(B)$ (see fig. 7.6). Recall that the leaves of $\mathcal{T}$ are identified with the vertices of G.

Avoiding Balls As alluded to in fig. 7.5, we may add new (hollowed) balls, called avoiding balls, to $\mathcal{Z}_\tau$ that do not exist in the given geometric sequence. An avoiding (hollowed) ball B, has an additional label, $t_d(B)$, where $t_d(B)$ is always a descendant of $t(B)$; the name avoiding stands for the fact that the edge-disjoint paths of $G(t(B))$ that are crossing B are avoiding the induced subgraph $G(t_d(B))$. Therefore, we exclude the subtree of $t_d(B)$ from $\mathcal{C}(B)$, i.e., $\mathcal{C}(B) \cap t_d(B) = \emptyset$.

We insert an avoiding hollowed ball only when we shrink or remove part of a nonavoiding (hollowed) ball that already exists in $\mathcal{Z}_\tau$. For example, if B' is the red ball on the right side of fig. 7.5, then $t(B') = t$, $t_d(B') = t_2$. Note that it is important that avoiding balls are replacing nonavoiding balls; if in the arrangement of fig. 7.5 the ball B were an avoiding ball, then the red ball would have to avoid two induced subgraphs; further escalation of this would lead to unmanageable labels. We get around this by never introducing an avoiding ball when the original B is avoiding. Also, for the charging argument to work we need to allocate a fraction of the number of tokens that would be normally allocated to a nonavoiding ball.

For any avoiding hollowed ball $B = B(x, r_1||r_2)$ there must be vertices $u, v \in V(t(B)) \setminus V(t_d(B))$ such that $\left\|X_u - x\right\|_1 \leq r_1$, $\left\|X_v - x\right\|_1 \geq r_2$ and that there are at least $k/4$ edge-disjoint paths from u to v in the induced graph $G[V(t(B)) \setminus V(t_d(B))]$. Note that if for such a ball, one defines $\mathcal{C}(B)$ to be the subtree rooted at $t(B)$ minus the subtree rooted at $t_d(B)$, then these $k/4$ edge-disjoint paths must be supported on the leaves of $\mathcal{T}$ that are in $\mathcal{C}(B)$.

Non-insertable Balls In Fact 7.28 we used the disjointness property to argue that any ball of FBag_ℓ is in the interior of at most one hollowed ball of $\mathcal{Z}_\tau$. Here, this fact may not necessarily hold: Suppose at time τ, a ball $B \in \text{Bag}_t$ is in the "interior" of two balls B_1, B_2, i.e., the center of B is far from the boundaries of B_1, B_2, and t is an ancestor-descendant of both $t(B_1), t(B_2)$. Then, B_1, B_2 intersect. Assuming that balls of $\mathcal{Z}_\tau$ have a "valid labeling", since B_1, B_2 are intersecting, $t(B_1), t(B_2)$ are not ancestor-descendant. One would hope that this configuration is impossible. But in fact, it could be the case that $t(B_1), t(B_2)$ are descendants of $t(B)$ that are not ancestor-descendants of each other. In this configuration, one cannot hope to add B with the label $t(B) = t$.

In general, the above scenario occurs only if the bags assigned to descendants of a node t appear earlier in the geometric sequence, i.e., if we process Bag_t after processing bags assigned to its descendants. In the first reading of the proof, one can assume that this scenario does not happen and avoid the notation $t_P(.)$ and (non-)insertable balls that we define below. To address this issue we will use the third property of the locally connected hierarchy. To any (hollowed) ball B in our construction with $t(B) = t$, we will assign $t_P(B) \subset \mathcal{T}$ to be a set of descendants of t with the guarantee that there are k edge-disjoint paths across B supported on $G[V(t) \setminus \cup_{t' \in t_P(B)} V(t')]$. In other words, we exclude the subtrees rooted at nodes of $t_P(B)$ from $\mathcal{C}(B)$. We will prune everything from Bag_t except the balls B such that $t_P(B)$ includes all descendants of t that are processed earlier than t. We use the third property of the locally connected hierarchy, $\mathcal{T}$, to show that the pruning step only removes a small fraction of balls.

Recall that FBag_ℓ has type (δ_ℓ, T_ℓ). For a node $t \in T_\ell$, we say a node t' is a *predecessor* of t, if t' is a descendant of t and $t' \in T_i$ for some $i < \ell$. For any node t and any ball $B = B(X_u, r) \in \text{Bag}_t$ we say B is *non-insertable* by t' if t' is a predecessor of t and an endpoint of an edge of $\mathcal{P}(t')$ is in B (see section 3.5 for the definition of $\mathcal{P}(t')$). We say B is *insertable* otherwise. For any insertable ball $B \in \text{Bag}_t$ we let $t_P(B)$ be the set of predecessors of t. In other words, a ball $B = B(X_u, r) \in \text{Bag}_t$ is insertable if and only if

i) For any $t' \in t_P(B)$, all endpoints of the edges of $\mathcal{P}(t')$ are outside of B, and

ii) For any $t' \in t_P(B)$, $u \notin V(t')$, i.e., u does not belong to any of the subtrees rooted at nodes of $t_P(B)$.

Observe that, by the definition of assigned bags of balls, (ii) follows from (i). In particular, since $B \in \text{Bag}_t$, there is an edge $\{u, v\} \in \mathcal{O}(t)$ for $v \notin V(t)$. Therefore, if $u \in V(t')$, $\{u, v\} \in \mathcal{P}(t')$ which is a contradiction.

Preprocessing

In this subsection, we delete all non-insertable balls and we show that they contribute only to a small fraction of the sum of the radii of the given geometric sequence. Then, we formally define a valid labeling and we show that we can lower bound the denominator by the sum of the widths of balls in a valid labeling. At the end of this subsection, we reduce proposition 7.24 to a "simpler" statement, that is the existence of an arrangement of a set of hollowed balls with a valid labeling such that the sum of the widths of all hollowed balls in the construction is a constant fraction of the sum of the radii of all balls in the given geometric sequence.

In the following lemma we show that for any node $t \in T_\ell$, the sum of radii of all balls that are non-insertable by t is $\ll \delta_\ell \cdot |\operatorname{Bag}_t|$.

Lemma 7.29. *For any node $t \in T_\ell$,*

$$\sum_i \sum_{B \in \mathrm{FBag}_i} \mathbb{I}[B \textit{ is non-insertable by } t] \cdot \delta_i \leq \frac{4\delta_\ell \cdot |\operatorname{Bag}_t|}{C_3}.$$

Proof. For any i let b_i be the number of balls in FBag_i that are non-insertable by t. By definition, $b_i = 0$ for $i \leq \ell$. We will show that for all $i > \ell$,

$$b_i \leq 2|\mathcal{P}(t)|. \tag{7.44}$$

Then,

$$\begin{aligned}
\sum_i \sum_{B \in \mathrm{FBag}_i} \mathbb{I}[B \text{ is non-insertable by } t] \cdot \delta_i &= \sum_{i>\ell} b_i \cdot \delta_i \\
&\leq 2|\mathcal{P}(t)| \sum_{i>\ell} \delta_i \\
&\leq 4\lambda \cdot |\mathcal{P}(t)| \cdot \delta_\ell \\
&\leq \frac{4|\mathcal{O}(t)|}{k} \cdot \delta_\ell \\
&\leq \frac{4|\operatorname{Bag}_t|\delta_\ell}{C_3}.
\end{aligned}$$

where the second to last inequality uses $\mathcal{T}$ is a $(k, k\lambda, T)$-LCH of G, i.e., that $t \in T$ and $\lambda \cdot k \cdot |\mathcal{P}(t)| \leq |\mathcal{O}(t)|$. The last inequality uses (7.24) and that Bag_t is a C_3/k-assigned bag of balls.

It remains to prove (7.44). Fix $i > \ell$. For any ball $B = B(X_u, \delta_i) \in \operatorname{Bag}_{t'}$ that is non-insertable by t, at least one endpoint of an edge of $\mathcal{P}(t)$ is in B. Since all balls of FBag_i are disjoint, $b_i \leq 2|\mathcal{P}(t)|$. □

By the above lemma it is sufficient to prove proposition 7.24 with the assumption that all balls in the given geometric sequence are insertable (see proposition 7.31 at the end of this part).

Any set of balls has a valid ball labeling if it satisfies the following properties.

1. For any nonavoiding ball B, $\mathcal{C}(B)$ is the connected subtree rooted at $t(B)$ excluding the subtrees rooted at nodes of $t_P(B)$. If B is avoiding, in addition to above, $\mathcal{C}(B)$ excludes the subtree rooted at $t_d(B)$. Note that we always have $t(B) \in \mathcal{C}(B)$.
2. For any hollowed ball $B = B(x, r_1||r_2)$, any $t' \in t_P(B)$, and $\{u, v\} \in \mathcal{P}(t')$, $\left\|x - X_u\right\|_1, \left\|x - X_v\right\|_1 \geq r_2$.
3. For any ball $B = (x, r_1||r_2)$, there is a vertex $u \in \mathcal{C}(B)$ such that $\left\|x - X_u\right\|_1 \leq r_1$ and there are at least $k/4$ edge-disjoint paths originating from u, crossing B, supported on $V(t(B)) \setminus V(t_d(B))$. In the proof of lemma 7.30 we show that this implies that we have $k/4$ edge-disjoint paths crossing B and supported on leaves of $\mathcal{T}$ which are in $\mathcal{C}(B)$.
4. For any two intersecting (hollowed) balls B_1 and B_2, $\mathcal{C}(B_1) \cap \mathcal{C}(B_2) = \emptyset$. Observe that $\mathcal{C}(B_1) \cap \mathcal{C}(B_2) \neq \emptyset$ if and only if either $t(B_1) \in \mathcal{C}(B_2)$ or $t(B_2) \in \mathcal{C}(B_1)$.

Figure 7.7: Properties of a valid ball labeling

In fig. 7.7 we define a valid labeling of balls. Later, in our inductive argument we will make sure that at any time τ, $\mathcal{Z}_\tau$ has a valid labeling.

The following lemma extends lemma 7.26 to the new setting where the balls of $\mathcal{Z}_\tau$ may intersect.

Lemma 7.30. *For any set of hollowed balls $\mathcal{Z}$ with a valid labeling we have,*

$$\frac{k}{4} \cdot \sum_{B(x,r_1||r_2) \in \mathcal{Z}} (r_2 - r_1) \leq \sum_{\{u,v\} \in E} \left\|X_u - X_v\right\|_1.$$

Proof. By property 3, for any ball $B = (x, r_1||r_2)$ there are $k/4$ edge-disjoint paths crossing B originating from a vertex $u \in \mathcal{C}(B)$ such that $\left\|X_u - x\right\|_1 \leq r_1$. We only keep the portion of each of these paths starting from u until the first vertex that lies outside of $B(x, r_2)$ (and we discard the rest). Next, we show that these paths remain inside $\mathcal{C}(B)$. This is because by property 3 these paths exclude the subtree rooted at $t_d(B)$. In addition, these paths start at a vertex that does not lie in any of the subtrees rooted at $t_P(B)$; by property 2 they can never enter such a vertex. Therefore, these paths avoid the subtrees rooted at $t_P(B)$ as well, or in other words they are completely supported on $\mathcal{C}(B)$.

We further trim each of these paths from both ends so that the resulting paths lie inside B. By the L_1 triangle inequality, the L_1 length of the trimmed paths is at least the width of B. Now, by property 4, no edge of G is charged by more than its L_1 length. $\square$

Proposition 7.31. *Given a $(k, k \cdot \lambda, T)$-LCH $\mathcal{T}$ of G and a λ-geometric sequence of families of $12C_3/k$-assigned bags of balls, $\text{FBag}_1, \text{FBag}_2, \ldots$, such that FBag_i has type (δ_i, T_i) and T_i's are disjoint subsets of T, if all balls of all bags in the sequence are insertable, $C_4 \geq 3$, $\lambda \leq 1/6C_4$, and $C_3 \geq 2((C_4+1) + 4(C_4+2)^2)$, then there is a set $\mathcal{Z}$ of hollowed balls with a valid labeling such that*

$$\frac{C_4}{12C_3} \cdot \sum_i \sum_{t \in T_i} \delta_i \cdot |\text{Bag}_t| \leq \sum_{B(x, r_1 || r_2) \in \mathcal{Z}} (r_2 - r_1).$$

It is easy to see that the above proposition together with lemma 7.30 implies proposition 7.24.

Proof of proposition 7.24. For any i and any $t \in T_i$ we remove all non-insertable balls in Bag_t. If at least half of the balls of Bag_t are insertable then we will have a $12C_3/k$-assigned bag of balls. Otherwise, we remove Bag_t from our geometric sequence and we remove t from T_i. The resulting geometric sequence satisfies the conditions of proposition 7.31.

By lemma 7.29, the sum of the radii of balls that we removed, which is at most twice the sum of the radii of all non-insertable balls, is at most half of the radii of all balls in the given geometric sequence,

$$\begin{aligned}\sum_j \sum_{B \in \text{FBag}_j} \mathbb{I}[B \text{ is non-insertable}] \cdot \delta_j &\leq \sum_i \sum_{t \in T_i} \sum_j \sum_{B \in \text{FBag}_j} \mathbb{I}[B \text{ is non-insertable by } t] \cdot \delta_j \\ &\leq \sum_i \sum_{t \in T_i} \frac{4|\text{Bag}_t| \cdot \delta_i}{C_3} \leq \sum_i \sum_{t \in T_i} \frac{|\text{Bag}_t| \cdot \delta_i}{4}\end{aligned}$$

where the last inequality uses $C_3 \geq 16$. Therefore, the proposition follows by lemma 7.30. □

We conclude this section with a simple fact. We show that any insertable ball $B \in \text{Bag}_t$ satisfies properties 2 and 3.

Fact 7.32. *Any insertable ball $B = B(X_u, \delta_\ell) \in \text{Bag}_t$ satisfies properties 2 and 3 of fig. 7.7.*

Proof. Property 2 follows by the definition of insertable balls. To see 3 note that all balls of Bag_t are nonavoiding; in addition, since B is insertable, u does not belong to any of the subtrees rooted at $t_P(B)$. Since by definition of $\mathcal{T}$, $G(t)$ is k-edge-connected, there are k edge-disjoint paths from u to a vertex of $V(t)$ outside of B (note that since $|\text{Bag}_t| > 1$ there is always a vertex of $V(t)$ outside of B). □

Order of Processing

In the rest of this section we prove proposition 7.31. So from now on, we assume all balls of all bags in the sequence are insertable and that every bag is $12C_3/k$-assigned.

Similar to section 7.3, we give an inductive proof. In this part we describe general properties of our construction and we use them to prove two essential lemmas. We process families of bags of balls in phases, and in phase ℓ we process FBag_ℓ. We need to use slightly larger (compared to the previous section) constants in the definition of interior balls.

Definition 7.33 (Interior ball). *We say a ball $B = B(X_u, \delta_\ell) \in \mathrm{Bag}_t$ is* in the interior of *a hollowed ball $B' = B(x, r_1 \| r_2)$ if $\mathcal{C}(B) \cap \mathcal{C}(B') \neq \emptyset$ and,*

$$r_1 + C_3 \cdot \delta_\ell < \left\| x - X_u \right\|_1 < r_2 - C_3 \cdot \delta_\ell.$$

We say B is an interior ball (with respect to $\mathcal{Z}$) if B is in the interior of a hollowed ball (of $\mathcal{Z}$). If B is not an interior ball, we call it a border ball. Similar to the previous section we insert all border balls of phase ℓ at time τ_ℓ.

See fig. 7.8 for the main properties of our inductive construction. In the rest of this part we use these properties to prove lemmas 7.35 and 7.37. The following fact follows simply by property 3.

Lemma 7.34. *Suppose we are processing $\mathrm{Bag}_t \in \mathrm{FBag}_\ell$ at time τ. For any $s \geq 0$ and any ball $B \in \mathrm{Bag}_t$ and $B' \in \mathcal{Z}_{\tau,s}$, if $\mathcal{C}(B) \cap \mathcal{C}(B') \neq \emptyset$, then $t(B')$ is a weak ancestor of t.*

Proof. Let $t' = t(B')$. If $\mathcal{C}(B) \cap \mathcal{C}(B') \neq \emptyset$, then by property 1 of fig. 7.7, t, t' are ancestor-descendant. So, we just need to show that t' is not a descendant of t.

First, by properties 2 and 3 of fig. 7.8, $t' \in T_i$ for some $i \leq \ell$. If $t' \in T_\ell$ either $t' = t$ or $\mathrm{Bag}_{t'}$ is processed by time τ. Therefore, by property 1 of fig. 7.8, t' is not a descendant of t and we are done. Otherwise, $t' \in T_i$ and $i < \ell$. If t' is a descendant of t, then it is a predecessor of t and since B is an insertable ball, $t' \in t_P(B)$. So $t' \notin \mathcal{C}(B)$ and $\mathcal{C}(B') \cap \mathcal{C}(B) = \emptyset$, which cannot be the case. □

In the following lemma we show that when we are processing Bag_t (at time τ) any ball in this bag is in the interior of at most one hollowed ball of $\mathcal{Z}_{\tau,s}$.

Lemma 7.35. *Say we process $\mathrm{Bag}_t \in \mathrm{FBag}_\ell$ at time τ. For any $s \geq 0$, and any ball $B \in \mathrm{Bag}_t$ and $B' \in \mathcal{Z}_{\tau,s}$, if $\mathcal{C}(B) \cap \mathcal{C}(B') \neq \emptyset$, then for any ball $B'' \neq B'$ in $\mathcal{Z}_{\tau,s}$ that intersects B', $\mathcal{C}(B) \cap \mathcal{C}(B'') = \emptyset$.*

Consequently, if B is in the interior of $B' \in \mathcal{Z}_{\tau,s}$, then $\mathcal{C}(B) \cap \mathcal{C}(B'') = \emptyset$ for any $B'' \in \mathcal{Z}_{\tau,s}$ that intersects B' and $B'' \neq B'$. So, B is in the interior of at most one ball of $\mathcal{Z}_{\tau,s}$.

Proof. Let $t' = t(B')$; fix a ball $B'' \in \mathcal{Z}_{\tau,s}$ and let $t'' = t(B'')$. Assume, for the sake of contradiciton, that $\mathcal{C}(B) \cap \mathcal{C}(B'') \neq \emptyset$. First, by lemma 7.34, t', t'' are weak ancestors of t. Since t' is a weak ancestor of t and $\mathcal{C}(B) \cap \mathcal{C}(B') \neq \emptyset$, we have $t \in \mathcal{C}(B')$. Similarly, $t \in \mathcal{C}(B'')$. Therefore, $\mathcal{C}(B') \cap \mathcal{C}(B'') \neq \emptyset$ which contradicts the validity of the labeling since B', B'' intersect. □

[1] Note that the root has depth 0.

1. Phase ℓ starts at $\tau_{\ell-1}+1$ and ends at τ_ℓ. In phase ℓ, we process assigned bags of balls in FBag_ℓ, in the increasing order of the depth[1] of the node to which they are assigned in $\mathcal{T}$. For example, if $\mathrm{Bag}_{t_1}, \mathrm{Bag}_{t_2} \in \mathrm{FBag}_\ell$ and t_1 is an ancestor of t_2, we process Bag_{t_1} before Bag_{t_2}.
2. Any ball of FBag_ℓ that we insert (in phase ℓ) remains unchanged till the end of phase ℓ. All other hollowed balls may be shrunk or be split into several balls but their labels (and their conflict set s) remain invariant.
3. Say at time $\tau_{\ell-1} < \tau < \tau_\ell$ we are processing Bag_t. We construct $\mathcal{Z}_{\tau+1}$ inductively by constructing $\mathcal{Z}_{\tau,0} = \mathcal{Z}_\tau$, $\mathcal{Z}_{\tau,1}$, ..., $\mathcal{Z}_{\tau,\infty} = \mathcal{Z}_{\tau+1}$. We make sure that each set $\mathcal{Z}_{\tau,s}$ has a valid labeling. When we are constructing $\mathcal{Z}_{\tau,s+1}$, we insert several new (hollowed) balls where only some of them are in Bag_t. Those not in Bag_t are inserted as a result of a conflict in labeling that would be introduced if we inserted a ball of Bag_t. In these cases, we split or shrink an already inserted nonavoiding ball B' and we insert new hollowed balls $B \subseteq B'$ such that $\mathcal{C}(B) \subseteq \mathcal{C}(B')$. We also let $t(B) = t(B')$ or $t(B) = t$ depending on whether B is avoiding or nonavoiding.
4. At time τ_ℓ we process the border balls of all bags of FBag_ℓ.

Figure 7.8: Properties of our Inductive Construction

Next, we show that once a ball of Bag_t becomes a border ball, it remains a border ball till the end of phase ℓ. In the proof we use the following simple fact.

Fact 7.36. *Suppose B, B' have a valid labeling; for any ball $B'' \subseteq B$ such that $\mathcal{C}(B'') \subseteq \mathcal{C}(B)$, $\{B', B''\}$'s labeling is valid as well.*

Lemma 7.37. *Suppose we are processing $\mathrm{Bag}_t \in \mathrm{FBag}_\ell$ at time τ. For any ball $B \in \mathrm{Bag}_t$ if B is an interior ball with respect to (some hollowed ball of) $\mathcal{Z}_{\tau',s}$ for some $\tau \leq \tau' < \tau_\ell$ and $s > 0$, then, it is also in the interior of a ball of $\mathcal{Z}_{\tau',s-1}$.*

This lets us backtrack through $\mathcal{Z}_{\tau',s}$'s until we reach $\mathcal{Z}_{\tau,0}$. So, if B is a border ball at the time we start processing Bag_t it remains a border ball until time τ_ℓ.

Proof. If B is in the interior of a newly inserted ball $B' \in \mathcal{Z}_{\tau',s}$, by property 3, the conflict set of B' is a subset of a conflict set of a ball $B'' \in \mathcal{Z}_{\tau'-1,s}$ containing B'. So, by Fact 7.36, B is also in the interior of B''. $\square$

The Construction

At any time $\tau_{\ell-1} < \tau \le \tau_\ell$ and $s \ge 0$, we allocate $\text{token}_{\tau,s}(B)$ tokens to any hollowed ball $B = B(x, r_1||r_2) \in \mathcal{Z}_{\tau,s}$, where

$$\text{token}_{\tau,s}(B) = \begin{cases} \delta_\ell - C_4 \cdot \delta_{\ell+1} & \text{if } B \in \text{FBag}_\ell \\ [r_2 - r_1 - C_4 \cdot \delta_\ell]^+ & \text{if } B \notin \text{FBag}_\ell \text{ is nonavoiding} \\ [\frac{r_2 - r_1 - C_4 \cdot \delta_\ell}{2(2+C_4)}]^+ & \text{otherwise.} \end{cases}$$

Note that we allocate significantly smaller number of tokens to the avoiding hollowed balls; roughly speaking we allocate $1/2C_4$ fraction of what we allocate for a same-sized nonavoiding ball.

Say we are processing Bag_t at time $\tau_{\ell-1} + 1 \le \tau < \tau_\ell$. We process Bag_t in several steps; we start with $\mathcal{Z} = \mathcal{Z}_\tau$ and in each iteration of the loop we may add/remove several (hollowed) balls to/from $\mathcal{Z}$. We use $\mathcal{Z}_{\tau,s}$ to denote the set $\mathcal{Z}$ after the s-th iteration of the loop, so, $\mathcal{Z} = \mathcal{Z}_{\tau,0} = \mathcal{Z}_\tau$ before entering the loop and $\mathcal{Z} = \mathcal{Z}_{\tau,\infty} = \mathcal{Z}_{\tau+1}$ after the loop. Before processing Bag_t, we let Bor_t be the set of border balls of Bag_t with respect to $\mathcal{Z}_{\tau,0}$ and relint_t be the set of interior balls. We update these sets in each iteration of the loop. We use $\text{Bor}_{t,s}, \text{relint}_{t,s}$ to denote the sets $\text{Bor}_t, \text{relint}_t$ after the s-th iteration of the loop, respectively. In addition, we use $\text{Bor}_{t,\infty}, \text{relint}_{t,\infty}$ to denote these sets after the execution of the loop. We will process the balls in $\text{Bor}_{t,\infty}$ at the end of phase ℓ. The details of our construction are described in algorithm 4.

The following is the main result of this part.

Lemma 7.38. *For any $\tau, s \ge 0$ the following holds. The set $\mathcal{Z}_{\tau,s}$'s labeling is valid. If we allocate $\text{token}_{\tau,s}(B)$ tokens to any hollowed ball $B(x, r_1||r_2) \in \mathcal{Z}_{\tau,s}$, then we can distribute these tokens among nodes whose bags we have processed by time τ such that for any $i < \ell$, any $t' \in T_i$ receives at least $\frac{C_4}{12C_3} \cdot |\text{Bag}_{t'}| \cdot \delta_i$ tokens, and any $t' \in T_\ell$ that is processed by time τ receives at least*

$$\frac{C_4}{6C_3} \cdot (|\text{Bag}_{t'}| - |\text{Bor}_{t',\infty}| - |\text{relint}_{t',\infty}|) \cdot \delta_\ell$$

tokens, and the node t that we are processing at time τ receives at least

$$\frac{C_4}{6C_3} \cdot (|\text{Bag}_t| - |\text{Bor}_{t,s}| - |\text{relint}_{t,s}|) \cdot \delta_\ell$$

tokens.

Later, in the post processing phase we show that any node t receives at least $\frac{C_4}{6C_3}|\text{Bor}_{t,\infty}| \cdot \delta_\ell$ new tokens. This implies proposition 7.31 as by the stopping condition of the main loop of algorithm 4, for any $t \in T_\ell$, $|\text{relint}_{t,\infty}| < |\text{Bag}_t|/2$.

We prove the above lemma by an induction on τ, s. From now on, we assume that all conclusions of the lemma hold for τ, s and we prove the same holds for $\tau, s+1$. We construct $\mathcal{Z}_{\tau,s+1}$ (from $\mathcal{Z}_{\tau,s}$) in one of the three steps of the loop, i.e., steps 5, 14, 16. We analyze these steps in the following three cases.

Algorithm 4 Construction of $\mathcal{Z}_{\tau+1}$ by processing Bag_t.

Input: $\mathcal{Z}_\tau$ and $\mathrm{Bag}_t \in \mathrm{FBag}_\ell$.

Output: $\mathcal{Z}_{\tau+1}$

1: Let $\mathcal{Z} = \mathcal{Z}_\tau$, t^* be parent of t and $\mathrm{Bor}_t, \mathrm{relint}_t$ be the border balls and interior balls of Bag_t respectively. Also, let $\mathcal{O}'(t) = \{\{u, v\} \in \mathcal{O}(t) : \left\|X_u - X_v\right\|_1 < \delta_\ell\}$.

2: **while** $|\,\mathrm{relint}_t\,| \geq |\,\mathrm{Bag}_t\,|/2$ **do**

3: **if** $\exists B' \in \mathrm{relint}_t$ s.t. B' is in the interior of an *avoiding* hollowed ball $B \in \mathcal{Z}$, **then**

4: Suppose $B' = B(X_u, \delta_\ell)$ and $B = B(x, r_1||r_2)$.

5: **Update** $\mathcal{Z}$**:** Remove B and add $B_1 = B(x, r_1||\left\|X_u - x\right\|_1 - \delta_\ell)$ and $B_2 = B(x, \left\|X_u - x\right\|_1 + \delta_\ell||r_2)$ with the same labels as B. Add B' (to $\mathcal{Z}$) and remove it from relint_t. **Goto** step 19.

6: **else**

7: Let $S_1, \ldots, S_j$ be a natural decomposition of $G[V(t^*) \setminus V(t)]$ into $k/4$-edge-connected subgraphs as defined in definition 2.13. ▷ In lemma 7.39 we will show that $j \leq 2|\mathcal{O}(t)|/k$.

8: Let $U \subseteq V(t)$ be the centers of balls of relint_t,

$$V_i := \{v \in S_i : \exists u \in U, \{u, v\} \in \mathcal{O}'(t)\},$$
$$U_i := \{u \in U : \exists v \in S_i, \{u, v\} \in \mathcal{O}'(t)\}$$

▷ By definition 7.13, every vertex of U is incident to an edge of $\mathcal{O}'(t)$, so $\cup_{i=1}^{j} U_i = U$. Also, since Bag_t is a $12C_3/k$-assigned bag, $|U| = |\,\mathrm{relint}_t\,| \geq |\,\mathrm{Bag}_t\,|/2 \geq \frac{6C_3|\mathcal{O}(t)|}{k}$.

9: Let $i = \mathrm{argmax}_{1 \leq i \leq j} |U_i|$. ▷ So, $|U_i| \geq |U|/j \geq 3C_3$.

10: Let $B = B(x, r_1||r_2) \in \mathcal{Z}$ be a nonavoiding ball such that a ball of relint_t with its center in U_i is in the interior of B. ▷ We will show that $t(B)$ is an ancestor of t.

11: We define $r_1' = \max\{r_1, \min_{v \in V_i} \left\|x - X_v\right\|_1\}$ and $r_2' := \min\{r_2, \max_{v \in V_i} \left\|x - X_v\right\|_1\}$.

12: Let $\mathrm{relint}_{B'}$ be the balls of relint_t whose centers are in the hollowed ball $B' = B(x, r_1' - \delta_\ell||r_2' + \delta_\ell)$ and $U_{B'}$ be the centers of balls of $\mathrm{relint}_{B'}$. ▷ We may have $U_i \nsubseteq U_{B'}$ as some vertices of U_i may not even be in B, but all vertices of $U_{B'}$ are in B.

13: **if** $|\,\mathrm{relint}_{B'}\,| \cdot \delta_\ell > 3(r_2' - r_1')$ **then** ▷ We treat $\mathrm{relint}_{B'}$ as if it was a 3-compact bag of balls.

14: **Update** $\mathcal{Z}$**:** Remove B and add $B_1 = B(x, r_1||r_1' - 2\delta_\ell)$ and $B_2 = B(x, r_2' + 2\delta_\ell||r_2)$ with the same labels as B. Add all balls of $\mathrm{relint}_{B'}$ to $\mathcal{Z}$ and remove them from relint_t.

15: **else**

16: **Update** $\mathcal{Z}$**:** Remove B and add $B_1 = B(x, r_1||r_1')$ and $B_2 = B(x, r_2'||r_2)$ to $\mathcal{Z}$, with the same labels as B. Add a new (nonavoiding) hollowed ball $B_3 = B(x, r_1' + \delta_\ell||r_2' - \delta_\ell)$ with $t(B_3) = t$ and $t_P(B_3)$ consisting of nodes $t' \in t_P(B)$ such that t' is a descendant of t. Add an avoiding hollowed ball $B_4 = B(x, r_1'||r_2')$ with $t(B_4) = t(B)$, $t_d(B_4) = t$ and $t_P(B_4) = t_P(B)$. Remove all balls of $\mathrm{relint}_{B'}$ from relint_t. See fig. 7.9 for an example. ▷ Note that no ball of $\mathrm{relint}_t \setminus \mathrm{relint}_{B'}$ is in the interior of B_1 or B_2.

17: **end if**

18: **end if**

19: Move all balls of relint_t that become border balls w.r.t. $\mathcal{Z}$ into Bor_t.

20: **end while**

return $\mathcal{Z}$.

Case 1: A ball $B' \in \text{relint}_{t,s}$ is in the interior of an avoiding hollowed ball $B = B(x, r_1||r_2) \in \mathcal{Z}_{\tau,s}$.
In this case by lemma 7.35, for any ball $B'' \in \mathcal{Z}_{\tau,s}$ such that $B \neq B''$, $\{B', B''\}$'s labeling is valid. Since, by definition, B' intersects neither of B_1, B_2, $\mathcal{Z}_{\tau,s+1}$'s labeling is valid. We send all tokens of B_1 and B_2 and $\delta_\ell/2$ of the tokens of B' to B and we redistribute them by the induction hypothesis. We send the rest of the tokens of B' to t. Then, B receives,

$$\text{token}_{\tau,s+1}(B_1) + \text{token}_{\tau,s+1}(B_2) + \frac{\delta_\ell}{2} \geq \frac{(r_1 - r_2 - 2\delta_\ell) - 2C_4 \cdot \delta_\ell + \delta_\ell(2 + C_4)}{2(2 + C_4)} = \text{token}_{\tau,s}(B).$$

In the above equation, we crucially use that, roughly speaking, $\text{token}_{\tau,s}(B)$ is a only a constant fraction of the width of B when B is an avoiding ball. This is not the case when we deal with nonavoiding balls in cases 2,3.

On the other hand, t receives

$$\text{token}_{\tau,s+1}(B') - \delta_\ell/2 \geq \delta_\ell - C_4 \cdot \delta_{\ell+1} - \delta_\ell/2 \geq \delta_\ell/4.$$

new tokens, where we used $\delta_{\ell+1} \leq \lambda \cdot \delta_\ell$ and $\lambda \leq 1/4C_4$. Since $|\,\text{Bor}_{t,s+1}\,| + |\,\text{relint}_{t,s+1}\,| = |\,\text{Bor}_{t,s}\,| + |\,\text{relint}_{t,s}\,| - 1$ we are done by induction.

Now suppose that the above does not happen. Consider the induced graph $G[V(t^*) - V(t)]$. Note that this graph may be disconnected. Let $S_1, S_2, \ldots, S_j$ be a natural decomposition of this graph as defined in definition 2.13. In the following lemma we show that $j \leq 2|\mathcal{O}(t)|/k$.

Lemma 7.39. $j \leq \frac{2|\mathcal{O}(t)|}{k}$.

Proof. By the definition of $\mathcal{T}$, $G(t^*)$ is k-edge-connected. Therefore, for any $1 \leq i \leq j$,

$$\partial_{G(t^*)}(S_i) \geq k.$$

Therefore,

$$j \cdot k \leq \sum_{i=1}^{j} \partial_{G(t^*)}(S_i) = \partial_{G(t^*)}(V(t)) + \sum_{i=1}^{j} \partial_{G[V(t^*)\setminus V(t)]}(S_i) = |\mathcal{O}(t)| + \sum_{i=1}^{j} \partial_{G[V(t^*)\setminus V(t)]}(S_i).$$

But, by lemma 2.14, the second term on the RHS is at most $2(j-1)(k/4-1)$. Therefore, $j \leq 2|\mathcal{O}(t)|/k$. □

As we mentioned in the comments of the algorithm, by the assumption that Bag_t is $12C_3/k$-assigned, the above lemma implies that

$$|U_i| \geq 3C_3. \tag{7.45}$$

Next, we prove a technical lemma which will be used in both of cases 2 and 3. In case 2 we use this lemma together with the above inequality to show that $|\,\text{relint}_{B'}\,| \geq 3(C_3 - 1)$; we will use this in our charging argument to compensate for the tokens lost by splitting B. In case 3, we use the following lemma to show that $r_2' - r_1' \geq (C_3 - 1) \cdot \delta_\ell$. Similarly, we use this inequality to compensate for the tokens lost by splitting B.

Lemma 7.40. *Let U, U_i, V_i be defined as in step 8. If $U_i \nsubseteq U_{B'}$, then $r'_2 - r'_1 \geq (C_3 - 1) \cdot \delta_\ell$.*

Proof. First, we show that there is a vertex $v \in V_i$ such that $X_v \notin B$. For the sake of contradiction assume $V_i \subset B$. We show that any vertex $u \in U_i$ is in $U_{B'}$ which is a contradiction. Fix a vertex $u \in U_i$. By definition 7.13, there is a vertex $v \in V_i$ such that $\{u, v\} \in \mathcal{O}'(t)$. Since $X_v \in B$, by the definition of r'_1, r'_2, we have $r'_1 \leq \left\|X_v - x\right\|_1 \leq r'_2$. So, $X_u \in B(x, r'_1 - \delta_\ell || r'_2 + \delta_\ell)$, i.e., $u \in U_{B'}$. This is a contradiction.

Now, let $v \in V_i$ be such that either $\left\|X_v - x\right\|_1 \geq r_2$ or $\left\|X_v - x\right\|_1 \leq r_1$. Here, we assume the former; the other case can be analyzed similarly. Then, we have $r'_2 = r_2$. But by definition of B, there is a ball $B(X_u, \delta_\ell) \in \text{relint}_{t,s}$ in the interior of B such that $u \in U_i$. Since $u \in U_i$, there is a vertex $w \in V_i$ such that $\left\|X_u - X_w\right\|_1 < \delta_\ell$. Therefore,

$$r'_1 \leq \left\|x - X_w\right\|_1 \leq \left\|x - X_u\right\|_1 + \delta_\ell \leq r_2 - C_3\delta_\ell + \delta_\ell.$$

where the last inequality uses that $B(X_u, \delta_\ell)$ is in the interior of B. So, $r'_2 - r'_1 \geq (C_3 - 1)\delta_\ell$. □

Case 2: $|\,\text{relint}_{B'}\,| \cdot \delta_\ell > 3(r'_2 - r'_1)$.
First, we show $\mathcal{Z}_{t,s+1}$'s labeling is valid. Then, we distribute the tokens. To show that $\mathcal{Z}_{t,s+1}$'s labeling is valid, first we argue that all balls of $\text{relint}_{B'}$ are in the interior of B. Fix a ball $A \in \text{relint}_{B'}$, we show A is in the interior of B. First, $\{A, B\}$'s labeling is *in*valid. Because i) A, B intersect by the definition of $\text{relint}_{B'}$ and ii) a ball of Bag_t is in the interior of B and all balls of Bag_t have the same labels. Secondly, since $\text{relint}_{B'} \subseteq \text{relint}_{t,s}$, A is an interior ball. Therefore, by lemma 7.35, A is in the interior of B. Now, by lemma 7.35, for any $B'' \in \mathcal{Z}_{\tau,s}$ where $B'' \neq B$, $\{A, B''\}$'s labeling is valid. Furthermore, by construction, B_1, B_2 do not intersect any balls of $\text{relint}_{B'}$. Hence, $\mathcal{Z}_{t,s+1}$'s labeling is valid.

Next, we describe the distribution of tokens allocated to the balls of $\mathcal{Z}_{\tau,s+1}$. Before that, we show that $|\,\text{relint}_{B'}\,| \geq 3(C_3 - 1)$. We consider two cases. If $U_i \subseteq U_{B'}$. Then, by (7.45),

$$|\,\text{relint}_{B'}\,| = |U_{B'}| \geq |U_i| \geq 3C_3.$$

Otherwise, $U_i \nsubseteq U_{B'}$. Then, by lemma 7.40,

$$|\,\text{relint}_{B'}\,| \geq \frac{3(r'_2 - r'_1)}{\delta_\ell} \geq \frac{3(C_3 - 1) \cdot \delta_\ell}{\delta_\ell} = 3(C_3 - 1).$$

Therefore, $|\,\text{relint}_{B'}\,| \geq 3(C_3 - 1)$.

Now, we send all tokens of B_1, B_2 and 3/4 of the tokens of each ball of $\text{relint}_{B'}$ to B and we redistribute them by the induction hypothesis. B receives,

$$\begin{aligned}
&\mathrm{token}_{\tau,s+1}(B_1) + \mathrm{token}_{\tau,s+1}(B_2) + \frac{3}{4}|\,\mathrm{relint}_{B'}\,|(\delta_\ell - C_4 \cdot \delta_{\ell+1}) \\
&\qquad \geq r_2 - r_1 - 4\delta_\ell - (r_2' - r_1') - 2C_4 \cdot \delta_\ell + \frac{3}{4} \cdot |\,\mathrm{relint}_{B'}\,| \cdot \frac{5}{6}\delta_\ell \\
&\qquad \geq \mathrm{token}_{\tau,s}(B) - (4 + C_4) \cdot \delta_\ell + \frac{7}{24}|\,\mathrm{relint}_{B'}\,| \cdot \delta_\ell \\
&\qquad \geq \mathrm{token}_{\tau,s}(B) - (4 + C_4) \cdot \delta_\ell + \frac{7}{8}(C_3 - 1) \cdot \delta_\ell \\
&\qquad \geq \mathrm{token}_{\tau,s}(B).
\end{aligned}$$

where the first inequality uses $\delta_{\ell+1} < \lambda \cdot \delta_\ell$ and $\lambda < 1/6C_4$, the second inequality uses the assumption $3(r_2' - r_1') < |\,\mathrm{relint}_{B'}\,| \cdot \delta_\ell$, the third inequality uses $|\,\mathrm{relint}_{B'}\,| \geq 3(C_3 - 1)$ and the last inequality uses $C_3 \geq 8(C_4 + 5)/7$. On the other hand, each ball $B' \in \mathrm{relint}_{B'}$ sends

$$\frac{1}{4}\,\mathrm{token}_\tau(B') \geq \frac{1}{4} \cdot \frac{5}{6}\delta_\ell$$

to t. So, t receives $|\,\mathrm{relint}_{B'}\,| \cdot \delta_\ell/5$ new tokens. Since

$$|\,\mathrm{Bor}_{t,s+1}\,| + |\,\mathrm{relint}_{t,s+1}\,| = |\,\mathrm{Bor}_{t,s}\,| + |\,\mathrm{relint}_{t,s}\,| - |\,\mathrm{relint}_{B'}\,|,$$

and we are done by induction.

Case 3: $|\,\mathrm{relint}_B\,| \cdot \delta_\ell \leq 3(r_2' - r_1')$.
As usual, first we verify the validity of the labeling, then we show that the tokens assigned to B_3, B_4 compensate the loss of B and the balls of $\mathrm{relint}_{B'}$ that we delete. We emphasize that verifying the validity of labeling is more involved in this case compared to cases 1, 2; this is because case 3 is the only one in which we insert new balls, i.e., B_3, B_4, that do not exist in the given geometric sequence of bags of balls.

First, we show that property 3 of fig. 7.8 is satisfied; then we verify properties 4, 2, 3 of fig. 7.7 in that order. Recall that the labels of B_3 and B_4 are defined as follows:

	$t(.)$	$t_d(.)$	$t_P(.)$	$\mathcal{C}(.)$
B_3	t	NA	$t_P(B) \cap \{\text{descendants of } t\}$	$\mathcal{C}(B) \cap$ subtree rooted at t
B_4	$t(B)$	t	$t_P(B)$	$\mathcal{C}(B) \setminus$ subtree rooted at t

Note that by lemma 7.34 and that a ball of Bag_t is in the interior of B, $t(B)$ is a weak ancestor of t. Therefore, $\mathcal{C}(B_3), \mathcal{C}(B_4) \subseteq \mathcal{C}(B)$ as required by property 3 of fig. 7.8. Let us now verify that $t_d(B_4) = t$ is a proper descendent of $t(B_4) = t(B)$, i.e., B_4 is a valid avoiding ball. Since we showed $t(B)$ is a weak ancestor of t, it is enough to show that $t(B) \neq t$. If $t(B) = t$, then B is constructed in an iteration $s' \leq s$ of the loop. This does not happen because whenever we construct a new ball in step 16 we delete all balls of relint_t that intersect with the new ball;

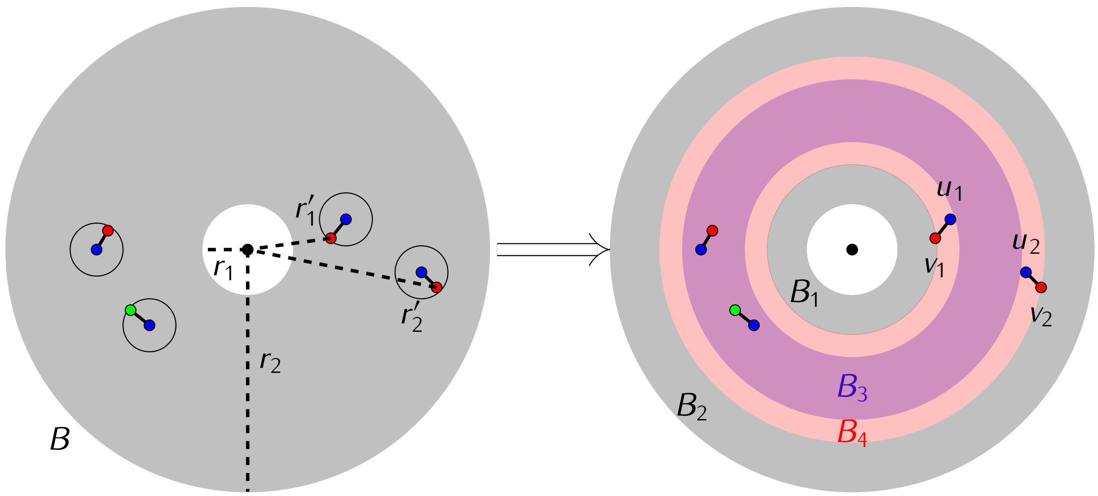

Figure 7.9: An illustration of Case 3. $U_{B'}$ is the blue vertices. V_i is the set of red vertices. The green vertex belongs to $V_{i'}$ for $i' \neq i$. The edges between the blue vertices and red/green vertices are in $\mathcal{O}'(t)$. We update $\mathcal{Z}_{\tau,s}$ as follows: We split B to balls B_1, B_2. We also add an avoiding B_4 from the closest red point (r'_1) to the farthest one (r'_2), and a nonavoiding ball, $B_3 = B(., r'_1 + \delta_\ell || r'_2 - \delta_\ell)$.

in addition, no new interior balls are added throughout the loop by lemma 7.37. Therefore $t(B) \neq t$.

Next, we verify property 4 of fig. 7.7. Since $\mathcal{C}(B_3), \mathcal{C}(B_4) \subseteq \mathcal{C}(B)$, by Fact 7.36, B_3, B_4 do not have a conflict with any ball of $\mathcal{Z}_{\tau,s} \setminus \{B\}$, i.e., for any ball $B'' \in \mathcal{Z}_{\tau,s} \setminus \{B\}$ that intersects one of them,

$$\mathcal{C}(B_3) \cap \mathcal{C}(B'') = \emptyset \text{ and } \mathcal{C}(B_4) \cap \mathcal{C}(B'') = \emptyset.$$

In addition, since $t_d(B_4) = t = t(B_3)$, $\mathcal{C}(B_3) \cap \mathcal{C}(B_4) = \emptyset$. Furthermore, B_3 and B_4 do not intersect B_1, B_2. So the labelings satisfy property 4 of fig. 7.7.

It remains to verify that B_3, B_4 satisfy properties 2 and 3 of fig. 7.7. B_3 and B_4 satisfy property 2 because $t_P(B_3), t_P(B_4) \subseteq t_P(B)$ and they are inside B. Finally, we need to verify property 3. First, we show B_3 satisfies property 3. By the definition of U_i there are vertices $u_1, u_2 \in U_i$ such that $\left\|x - X_{u_1}\right\| < r'_1 + \delta_\ell$ and $\left\|x - X_{u_2}\right\| > r'_2 - \delta_\ell$ (see fig. 7.9). Since $G(t)$ is k-edge-connected there are k edge-disjoint paths between u_1 and u_2 supported on $V(t)$. So, we just need to argue that $u_1 \in \mathcal{C}(B_3)$, i.e., for any $t' \in t_P(B_3)$, $u_1 \notin V(t')$. This is because, $u_1 \in U_i$ is incident to an edge e of $\mathcal{O}'(t)$. Since t' is a descendant of t, if $u_1 \in V(t')$ then $e \in \mathcal{P}(t')$ so an endpoint of an edge of $\mathcal{P}(t')$ has distance less than r_2 from the center of B which is contradictory with $t' \in t_P(B_3) \subseteq t_P(B)$.

Lastly, we show B_4 satisfies property 3. By the definition of V_i there are vertices $v_1, v_2 \in V_i$ such that $\left\|x - X_{v_1}\right\| \leq r'_1$ and $\left\|x - X_{v_2}\right\| \geq r'_2$ (see fig. 7.9). Since $V_i \subseteq S_i$ and S_i is $k/4$-edge-connected in $G[V(t^*) \setminus V(t)]$, there are $k/4$ edge-disjoint paths from v_1 to v_2 in $G[V(t(B)) \setminus V(t)]$. We need to argue that $v_1 \in \mathcal{C}(B_4)$, i.e., it is enough to show that for any $t' \in t_P(B_4)$, we have

$v_1 \notin V(t')$. This is similar to the argument in the previous paragraph. First, since $v_1 \in V_i$, v_1 is incident to an edge $e \in \mathcal{O}'(t)$. Since $t' \in t_P(B)$ and $\|X_{v_1} - x\|_1 \leq r_2$, we must have $e \notin \mathcal{P}(t')$. Therefore, if $v_1 \in V(t')$, t' must be a weak ancestor of t^*. But, since $t(B)$ is an ancestor of t and a ball of Bag_t is in the interior of B, we must have $t \in \mathcal{C}(B)$, i.e., $t_P(B)$ cannot not contain a weak ancestor of t. So, $v_1 \notin V(t')$.

It remains to distribute the tokens. First, we show that $r_2' - r_1' \geq (C_3 - 1) \cdot \delta_\ell$. If $U_i \nsubseteq U_{B'}$, then by lemma 7.40, $r_2' - r_1' \geq (C_3 - 1) \cdot \delta_\ell$. Otherwise, by the assumption of Case 3,

$$r_2' - r_1' \geq \frac{1}{3}|\,\mathrm{relint}_{B'}\,| \cdot \delta_\ell \geq \frac{1}{3}|U_i| \cdot \delta_\ell \geq C_3 \cdot \delta_\ell,$$

where the last inequality follows by (7.45). We send all tokens of B_1, B_2, B_3, and $(2C_4 + 2)\delta_\ell$ tokens of B_4 to B and we redistribute them by the induction hypothesis. We send the rest of the tokens of B_4 to t. Ball B receives

$$\sum_{i=1}^{3} \mathrm{token}_{\tau,s+1}(B_i) + (2C_4 + 2) \cdot \delta_\ell \geq r_2 - r_1 - 2\delta_\ell - 3C_4\delta_\ell + (2C_4 + 2)\delta_\ell = \mathrm{token}_{\tau,s}(B).$$

On the other hand, t receives,

$$\begin{aligned}
\mathrm{token}_{\tau,s+1}(B_4) - (2C_4 + 2)\delta_\ell &= \frac{r_2' - r_1' - C_4 \cdot \delta_\ell - 4(2 + C_4)^2 \cdot \delta_\ell}{2(2 + C_4)} \\
&\geq \frac{r_2' - r_1' - (C_3 - 1)\delta_\ell/2}{2(2 + C_4)} \\
&\geq \frac{r_2' - r_1'}{4(2 + C_4)} \\
&\geq \frac{|\,\mathrm{relint}_B\,| \cdot \delta_\ell}{12(2 + C_4)} \geq \frac{C_4|\,\mathrm{relint}_B\,| \cdot \delta_\ell}{6C_3},
\end{aligned}$$

new tokens. In the first inequality we used $(C_3 - 1) \geq 2(C_4 + 4(C_4 + 2)^2)$, the second inequality uses $r_2' - r_1' \geq (C_3 - 1) \cdot \delta_\ell$, and the third inequality uses the assumption $r_2' - r_1' \geq \frac{1}{3} \cdot |\,\mathrm{relint}_B\,| \cdot \delta_\ell$. This concludes the proof of lemma 7.38.

Post-processing

Say we have processed all $\mathrm{Bag}_t \in \mathrm{FBag}_\ell$ and we are the end of phase ℓ, i.e., time τ_ℓ. We need to make sure that each node $t \in T_\ell$ receives at least $\frac{C_4}{6C_3}|\,\mathrm{Bor}_{t,\infty}\,| \cdot \delta_\ell$ new tokens. Then, by lemma 7.38, each node t, altogether, receives at least

$$\frac{C_4}{6C_3}(|\,\mathrm{Bag}_t\,| - |\,\mathrm{relint}_{t,\infty}\,|) \cdot \delta_\ell \geq \frac{C_4}{12C_3}|\,\mathrm{Bag}_t\,| \cdot \delta_\ell$$

tokens. The above inequality uses that by the condition of the main loop of algorithm 4, for any $t \in T_\ell$, $|\,\mathrm{relint}_{t,\infty}\,| \leq |\,\mathrm{Bag}_t\,|/2$.

We define the shrink operator as follows: For any hollowed ball $B = B(x, r_1||r_2) \in \mathcal{Z}_{\tau_\ell}$,

$$\text{shrink}_\ell(B) = \begin{cases} B & \text{if } B \in \text{FBag}_\ell \\ B(x, r_1 + (C_3+1)\delta_\ell || r_2 - (C_3+1)\delta_\ell) & \text{if } B \notin \text{FBag}_\ell \text{ and } r_2 - r_1 > 2(C_3+1)\delta_\ell \\ B(x, 0) = \emptyset & \text{otherwise.} \end{cases} \tag{7.46}$$

Let

$$b := \sum_{t \in T_\ell} |\,\text{Bor}_{t,\infty}\,|,$$
$$\text{excess} := \sum_{B \in \mathcal{Z}_{\tau_\ell}} (\text{token}_{\tau_\ell+1}(B) - \text{token}_{\tau_\ell}(B)).$$

Think of excess as the additional number of tokens that we gain for all hollowed balls $B \in \mathcal{Z}_{\tau_\ell}$ when we go to the new phase $\ell+1$. Our idea is simple. If excess is very large then we do not add any of the border balls and we just distribute excess between all nodes of T_ℓ. Otherwise, we shrink balls of $\mathcal{Z}_{\tau_\ell}$ and we add the border balls.

Case 1: excess $\geq \frac{C_4}{6C_3} \cdot b \cdot \delta_\ell$.
In this case, we do not add any of the border balls and we simply let $\mathcal{Z}_{\tau_\ell+1} = \mathcal{Z}_{\tau_\ell}$.

Now, observe that for any hollowed ball $B \in \mathcal{Z}_{\tau_\ell}$, we have $\text{token}_{\tau_\ell+1}(B) - \text{token}_{\tau_\ell}(B)$ additional tokens that B has not used. We distribute these tokens between the nodes of T_ℓ proportional to their number of border balls. More precisely, for any ball $B \in \mathcal{Z}_{\tau_\ell}$ and $t \in T_\ell$, we send

$$\frac{|\,\text{Bor}_{t,\infty}\,|}{b} \cdot (\text{token}_{\tau_\ell+1}(B) - \text{token}_{\tau_\ell}(B))$$

tokens to t. Therefore, t receives

$$\begin{aligned} \sum_{B \in \mathcal{Z}_{\tau_\ell}} \frac{|\,\text{Bor}_{t,\infty}\,|}{b} \cdot (\text{token}_{\tau_\ell+1}(B) - \text{token}_{\tau_\ell}(B)) &= \frac{|\,\text{Bor}_{t,\infty}\,| \cdot \text{excess}}{b} \\ &\geq \frac{C_4}{6C_3} \cdot |\,\text{Bor}_{t,\infty}\,| \cdot \delta_\ell, \end{aligned}$$

and we are done.

Case 2: excess $< \frac{C_4}{6C_3} \cdot b \cdot \delta_\ell$.
For each hollowed ball $B \in \mathcal{Z}_{\tau_\ell}$ we replace B by $\text{shrink}_\ell(B)$ in $\mathcal{Z}_{\tau_\ell+1}$. We also add all balls of $\text{Bor}_{t,\infty}$ for all $t \in T_\ell$ to $\mathcal{Z}_{\tau_\ell+1}$. By lemma 7.37 any border ball $B \in \text{Bor}_{t,\infty}$ is not in the interior of any ball of $\mathcal{Z}_{\tau_\ell}$. By the definition of the shrink operator, and using the fact that balls of FBag_ℓ do not intersect, any ball of $\cup_{t \in T_\ell} \text{Bor}_{t,\infty}$ does not intersect any ball of $\mathcal{Z}_{\tau_\ell+1}$. So, $\mathcal{Z}_{\tau_\ell+1}$'s labeling is valid.

It remains to distribute the tokens. First, we prove a technical lemma.

Lemma 7.41. *If* excess $< \frac{C_4}{6C_3} \cdot b \cdot \delta_\ell$, *then*

$$b \cdot \delta_\ell \geq 2 \sum_{B \in \mathcal{Z}_{\tau_\ell}} (\mathrm{token}_{\tau_\ell}(B) - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B))).$$

Proof. It is sufficient to show that for any hollowed ball $B = B(x, r_1||r_2) \in \mathcal{Z}_{\tau_\ell}$

$$\mathrm{token}_{\tau_\ell+1}(B) - \mathrm{token}_{\tau_\ell}(B) \geq \frac{C_4}{3C_3} \cdot (\mathrm{token}_{\tau_\ell}(B) - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B))). \tag{7.47}$$

Because, then

$$\begin{aligned}\sum_{B \in \mathcal{Z}_{\tau_\ell}} \mathrm{token}_{\tau_\ell}(B) - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B)) &\leq \frac{3C_3}{C_4} \sum_{B \in \mathcal{Z}_{\tau_\ell}} \mathrm{token}_{\tau_\ell+1}(B) - \mathrm{token}_{\tau_\ell}(B) \\ &= \frac{3C_3}{C_4} \,\mathrm{excess} \leq \frac{b \cdot \delta}{2},\end{aligned}$$

as desired. The last inequality follows by the lemma's assumption.

It remains to prove (7.47). First, note that if $\mathrm{token}_{\tau_\ell}(B) = 0$ then the above holds trivially. So assume $\mathrm{token}_{\tau_\ell}(B) > 0$. We consider three cases. i) $B \in \mathrm{FBag}_\ell$. In this case both sides of the above inequality is zero. This is because $\mathrm{shrink}_\ell(B) = B$ and $\mathrm{token}_{\tau_\ell}(B) = \mathrm{token}_{\tau_\ell+1}(B)$. ii) B is a nonavoiding hollowed ball. Since $\mathrm{token}_{\tau_\ell}(B) > 0$, $r_2 - r_1 > C_4 \cdot \delta_\ell$. Therefore,

$$\begin{aligned}\mathrm{token}_{\tau_\ell+1}(B) - \mathrm{token}_{\tau_\ell}(B) &= C_4 \cdot (\delta_\ell - \delta_{\ell+1}) \geq \frac{2}{3} \cdot C_4 \cdot \delta_l \\ \mathrm{token}_{\tau_\ell}(B) - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B)) &\leq 2(C_3+1)\delta_\ell + C_4 \cdot (\delta_{\ell+1} - \delta_\ell) \leq 2C_3 \cdot \delta_\ell.\end{aligned}$$

using $\delta_{\ell+1} \leq \delta_\ell/3$ and $C_4 \geq 3$. So, (7.47) is correct. iii) B is an avoiding hollowed ball. Equation (7.47) is equivalent to case (ii) up to a $2(2 + C_4)$ factor in both sides of the inequality. ☐

For any ball $B \in \mathrm{Bor}_{t,\infty}$ and any ball $B' \in \mathcal{Z}_{\tau_\ell}$, we send

$$\frac{\delta_\ell}{2} \cdot \frac{\mathrm{token}_{\tau_\ell}(B') - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B'))}{\sum_{B'' \in \mathcal{Z}_{\tau_\ell}} \mathrm{token}_{\tau_\ell}(B'') - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B''))}$$

tokens to B' and we send the remaining tokens to t. For any ball $B \in \mathcal{Z}_{\tau_\ell}$, also send all of the tokens of $\mathrm{shrink}_\ell(B)$ to B.

Therefore, by lemma 7.41, any ball $B \in \mathcal{Z}_{\tau_\ell}$ receives at least

$$\begin{aligned}\mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B)) + b \cdot \frac{\delta_\ell}{2} \cdot \frac{\mathrm{token}_{\tau_\ell}(B) - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B))}{\sum_{B' \in \mathcal{Z}_{\tau_\ell}} \mathrm{token}_{\tau_\ell}(B') - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B'))} \\ \geq \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B)) + (\mathrm{token}_{\tau_\ell}(B) - \mathrm{token}_{\tau_\ell+1}(\mathrm{shrink}_\ell(B))) \\ = \mathrm{token}_{\tau_\ell}(B),\end{aligned}$$

that we redistribute by the induction hypothesis. On the other hand, any $t \in T_\ell$ receives

$$|\,\mathrm{Bor}_{t,\infty}| \cdot (\delta_\ell - \delta_\ell/2 - C_4 \cdot \delta_{\ell+1}) \geq |\,\mathrm{Bor}_{t,\infty}| \cdot \delta_\ell/4$$

new tokens, and we are done with the induction. This completes the proof of proposition 7.31.

Chapter 8

Finding an ATSP Tour

We have an algorithm for approximating the cost of the best ATSP tour. But as it has been mentioned many times, our techniques do not yield an algorithm for finding a tour with the same guarantee. There are two obstructions to constructing such an algorithm:

- Exponential size of Tree-CP: The convex programs we solve to find preconditioners for Kadison-Singer have exponential size. We will show how to get around this in the first part of this chapter.
- We do not know how to find the solution guaranteed in the Kadison-Singer problem in polynomial time. In the second part of this chapter we will show some progress towards making Kadison-Singer algorithmic.

8.1 Leveraging Directed Thinness

In this part we prove theorem 3.9. We emphasize that our algorithm does not necessarily find a thin tree. As alluded to in the introduction, the main barrier is that verifying the thinness is a variant of the sparsest cut problem for which the best known algorithm only gives an $O(\sqrt{\log n})$-approximation factor. Instead, we use the fact that "directed thinness", as defined in (3.3) of theorem 3.17, is polynomially testable and it is enough to solve ATSP. We refrain from giving the details and we refer interested readers to [Asa+10]. Our rough idea is as follows: We run the ellipsoid algorithm on the convex program Tree-CP by first discarding the 2^n constraints $\mathbf{1}_S^\intercal D \mathbf{1}_S \leq \mathbf{1}_S^\intercal L_G \mathbf{1}_S$ that verify D is a shortcut matrix. If the directed thinness of the output tree fails, the undirected thinness fails as well, so we get a set S for which $\mathbf{1}_S^\intercal D \mathbf{1}_S > \mathbf{1}_S^\intercal L_G \mathbf{1}_S$. That corresponds to a violating constraint of the convex program which the ellipsoid algorithm can use in the same way that it uses separation oracles. Repeating this procedure, either the ellipsoid algorithm converges, i.e., we find an actual undirected thin tree, or we find an ATSP tour along the way.

To complete the proof we need to make sure that we can construct the starting locally connected hierarchy in polynomial time; we will describe our algorithm later. Apart from that,

Algorithm 5 Expander extraction

Input: A $k \geq 7 \log n$-edge-connected graph $G = (V, E)$.
Output: A $k/20$-edge-connected, $1/4$-dense induced subgraph that is an $\Omega(1/k^2)$-expander.

```
 1: Let U ← V. We always let H be the induced subgraph on U.
 2: loop
 3:     if there is a vertex v ∈ U such that d_H(v) ≤ 7 d_G(v)/20 then
 4:         Let U ← U \ {v} and goto 2.
 5:     end if                      ▷ If this case does not happen, H is 7/20-dense.
 6:     Let S be the output of the spectral partitioning algorithm on H, and let T = U \ S.
 7:     if φ_G(S) ≤ φ_G(U) or φ_G(T) ≤ φ_G(U) then
 8:         Let U = S or U = T whichever has the smallest φ_G(.), and goto 2.
 9:     end if
10:     if max{φ_H(S), φ_H(T)} < 1/k then
11:         Let U = S or U = T whichever has fewer vertices, and goto 2.
12:     end if          ▷ If this case does not happen, by Cheeger's inequality, H is an
        Ω(1/k^2)-expander.
13:     If H is k/20-edge-connected, return H. Otherwise, let S ⊆ U be such that ∂_H(S) <
        k/20 and φ_H(S) ≥ φ_H(U \ S).                         ▷ So, φ_H(S) ≥ Ω(1/k^2).
14:     Let S_1, S_2, ... be a natural decomposition of G[S] into k/20-edge-connected compo-
        nents. By (6.3) there is S_i such that ∂_H(S_i) < k/10. Return G[S_i].
15: end loop
```

the main difficulty is that to obtain the shortcut matrix D promised in theorem 3.6 we need to solve $O(\log\log(n))$ many convex programs (Tree-CP($\mathcal{T}_i$)) and each one depends on the solution of the previous ones. In other words, we should be recursively calling $O(\log\log n)$ many ellipsoid algorithms. Therefore, if we find a separating hyperplane for one of the ellipsoids, we should restart the ellipsoid algorithms for all the proceeding convex programs. The resulting algorithm runs in time $n^{O(\log\log n)}$ and has an approximation factor of $\mathrm{poly}\log\log(n)$. We can also tradeoff the approximation factor with the running time of the algorithm by modifying algorithm 3 to have $O(\ell)$ number of iterations. For constant values of ℓ this gives a polynomial time approximation algorithm.

We will give an algorithm to construct an $(\Omega(1/k^2), .)$-expanding $(k/20, 1/4, \mathcal{T})$-LCH, $\mathcal{T}_0$ for some $\alpha \asymp 1/\log^2(n)$. Then, we run a modified version of algorithm 3 to obtain locally connected hierarchies $\mathcal{T}_1, \ldots, \mathcal{T}_{2\ell}$; in particular, we only run the loop for 2ℓ iterations; to make sure that $\mathcal{T}_{2\ell}$ is $(\Omega(1), .)$-expanding, we need to boost the expansion by $\left(\frac{1}{\alpha}\right)^{1/2\ell}$ in every iteration of the loop. To be more precise, for any $1 \leq i \leq 2\ell$, instead of (6.4), we let

$$F_i := \left\{ e \in E : \mathcal{R}\mathrm{eff}_{D_i}(e) \leq \frac{O((1/\alpha)^{1/2\ell}) f_1(k', \lambda')}{k'} \right\}.$$

The proof simply follows by a modification to the expansion boosting lemma. The resulting algorithm runs in time $n^{O(\ell)}$ and has an approximation factor of $\mathrm{poly}\log\log(n) \cdot \log^{1/\ell}(n)$.

It remains to find the starting locally connected hierarchy $\mathcal{T}_0$. Given a $k \geq 7\log n$-edge-connected graph $G = (V, E)$, all we need is to find a 1/4-dense $k/20$-edge-connected induced subgraph $G[S]$ whose expansion is $\Omega(1/k^2)$. We essentially make the proof of lemma 6.4 constructive using the spectral partitioning algorithm [AM85; Alo86] at the cost of obtaining an $\Omega(1/k^2)$-expander instead of a $1/k$-expander. This is because, by Cheeger's inequality, the spectral partitioning algorithm gives a square-root approximation to the problem of approximating $\phi(G)$. The details of the algorithm are described in algorithm 5.

8.2 Towards an Algorithm for Kadison-Singer

The main difficulty in making Kadison-Singer algorithmic is that we do not know how to computer mixed characteristic polynomials as defined in section 4.7. It is in fact enough to compute the maximum roots of these polynomials, however that does not seem to be any easier. Our idea is to approximately compute the maximum root.

The first observation is that computing top coefficients of the mixed characteristic polynomials is easier than the rest.

Theorem 8.1. *Given a determinantal distribution μ, along with vectors defining it, along with vectors $v_1, \ldots, v_n$, the k-th top coefficient of the mixed characteristic polynomial*

$$\mathbb{E}_{S\sim\mu}\chi[\sum_{i\in S} v_i v_i^\intercal](x),$$

can be computed in time $n^{O(k)}$.

Proof. It is enough to note that by lemma 2.8, the coefficients of $\chi[\sum_{i\in S} v_i v_i^\intercal]$ are simply $\sigma_k(\sum_{i\in S} v_i v_i^\intercal)$ for different values of k. By lemma 2.8 again we have

$$\sigma_k(\sum_{i\in S} v_i v_i^\intercal) = \sum_{T\in\binom{S}{k}} \sigma_k(\sum_{i\in T} v_i v_i^\intercal).$$

Therefore we can write the k-th coefficient of the mixed characteristic polynomial as

$$\sum_{T\in\binom{[n]}{k}} \mathbb{P}_{S\sim\mu}[T \subseteq S] \cdot \sigma_k(\sum_{i\in T} v_i v_i^\intercal).$$

It is an easy exercise to see that both factors inside the sum can be computed in polynomial time and therefore the sum can be computer in time $n^{O(k)}$. □

The next observation is that the top coefficients are enough to compute an approximation of the maximum root.

Theorem 8.2. *Given a real-rooted degree n polynomial $p \in \mathbb{R}[x]$ with nonnegative roots, one can compute a $\sqrt[k]{n}$-approximation of the maximum root by only accessing the top k coefficients.*

Proof. Let $\alpha_1 \leq \cdots \leq \alpha_n$ be the roots of p. The top k coefficients of p are the first k elementary symmetric polynomials of $\alpha_1, \ldots, \alpha_n$. Using the Newton identities, one can compute the first k moments of $\alpha_1, \ldots, \alpha_k$ using these elementary symmetric polynomials. Therefore we can compute

$$\sqrt[k]{\sum_{i=1}^{n} \alpha_i^k}.$$

It is easy to see that this quantity is at least α_n and at most $\sqrt[k]{n} \cdot \alpha_n$. $\square$

Based on these two observations one can see that the maximum root can be $(1+\epsilon)$-approximated in time $n^{O(\log n/\epsilon)}$. However $(1+\epsilon)$-approximation is not enough, since rounding Kadison-Singer requires n steps and $(1+\epsilon)^n$ is beyond our tolerance of error.

We can be more celever though. Take a batch of vectors such as $v_1, \ldots, v_{\sqrt{n}}$, and look at μ conditioned on all $2^{\sqrt{n}}$ combinations of them being in or out of the set distributed according to μ. The mixed characteristic polynomial for these conditional distributions interlace. So we simply approximate their maximum roots with error $1 + 1/\sqrt{n}$. It takes $2^{\tilde{O}(\sqrt{n})}$ time to approximate the maximum roots for all of these polynomials. We simply pick the conditioning that gave us the lowest approximation of the maximum root. Then we pick another batch of $\sqrt{n}$ vectors and repeat. This procedure takes $\sqrt{n}$ iterations and each time we could be losing a factor of $1 + 1/\sqrt{n}$ in the approximation. So in the end we only have lost a factor of $O(1)$ in the approximation and the whole procedure has taken $2^{\tilde{O}(\sqrt{n})}$ time.

It would be very interesting if computing the top coefficients of the mixed characteristic polynomial leads to algorithms with better running times.

Bibliography

[Alo86] N. Alon. "Eigenvalues and expanders". In: *Combinatorica* 6 (2 Jan. 1986), pp. 83–96 (cit. on p. 124).

[AM85] N. Alon and V. Milman. "Isoperimetric inequalities for graphs, and superconcentrators". In: *Journal of Combinatorial Theory, Series B* 38.1 (Feb. 1985), pp. 73–88 (cit. on p. 124).

[And79a] J. Anderson. "Extensions, restrictions, and representations of states on C*-algebras". In: *Transactions of the American Mathematical Society* 249.2 (1979), pp. 303–329 (cit. on p. 5).

[And79b] J. Anderson. "Extreme points in sets of positive linear maps on B(H)". In: *Journal of Functional Analysis* 31.2 (1979), pp. 195–217 (cit. on p. 5).

[And81] J. Anderson. "A conjecture concerning the pure states of B(H) and a related theorem". In: *Topics in modern operator theory*. Springer, 1981, pp. 27–43 (cit. on p. 5).

[AO14] N. Anari and S. Oveis Gharan. "The Kadison-Singer Problem for Strongly Rayleigh Measures and Applications to Asymmetric TSP". In: *arXiv preprint arXiv:1412.1143* (2014) (cit. on pp. 7, 27).

[AO15] N. Anari and S. Oveis Gharan. "Effective-Resistance-Reducing Flows, Spectrally Thin Trees, and Asymmetric TSP". In: *2015 IEEE 53rd Annual Symposium on Foundations of Computer Science — FOCS 2015*. 2015, pp. 20–39 (cit. on p. 7).

[App+11] D. L. Applegate et al. *The Traveling Salesman Problem: A Computational Study: A Computational Study*. Princeton university press, 2011 (cit. on p. 1).

[ARV09] S. Arora, S. Rao, and U. Vazirani. "Expander flows, geometric embeddings and graph partitioning". In: *Journal of the ACM (JACM)* 56.2 (2009), p. 5 (cit. on p. 60).

[Asa+10] A. Asadpour et al. "An O (log n/log log n)-approximation algorithm for the asymmetric traveling salesman problem". In: *SODA* (2010), pp. 379–389 (cit. on pp. 2–5, 18, 36, 42, 122).

[AW02] R. Ahlswede and A. Winter. "Strong converse for identification via quantum channels". In: *Information Theory, IEEE Transactions on* 48.3 (2002), pp. 569–579 (cit. on p. 19).

[AW13] C. Akemann and N. Weaver. “A Lyapunov type theorem from Kadison-Singer”. 2013. URL: `http://arxiv.org/abs/1308.5276` (cit. on pp. 23, 25).

[BB08] J. Borcea and P. Brändén. “Applications of stable polynomials to mixed determinants: Johnson’s conjectures, unimodality, and symmetrized Fischer products”. In: *Duke Math. Journal* 143.2 (2008), pp. 205–223 (cit. on p. 15).

[BB10] J. Borcea and P. Brädén. “Multivariate Pólya-Schur classification problems in the Weyl algebra”. In: *Proceedings of the London Mathematical Society* 101.3 (1 2010), pp. 73–104 (cit. on p. 55).

[BBL09] J. Borcea, P. Brändén, and T. M. Liggett. “Negative dependence and the geometry of polynomials.” In: *Journal of American Mathematical Society* 22 (2009), pp. 521–567 (cit. on pp. 39, 42).

[Blä02] M. Bläser. “A new approximation algorithm for the asymmetric TSP with triangle inequality”. In: *SODA*. 2002, pp. 638–645 (cit. on p. 3).

[BSS14] J. D. Batson, D. A. Spielman, and N. Srivastava. “Twice-Ramanujan Sparsifiers”. In: *SIAM Review* 56.2 (2014), pp. 315–334 (cit. on pp. 21, 47).

[BV06] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2006 (cit. on pp. 11, 73).

[Cas+07] P. G. Casazza et al. “Projections and the Kadison-Singer problem”. In: *arXiv preprint math/0701450* (2007) (cit. on p. 5).

[CGK06] M. Charikar, M. X. Goemans, and H. Karloff. “On the Integrality Ratio for the Asymmetric Traveling Salesman Problem”. In: *MOR* 31.2 (2006), pp. 245–252 (cit. on p. 3).

[Chr76] N. Christofides. *Worst-case analysis of a new heuristic for the travelling salesman problem*. Tech. rep. DTIC Document, 1976 (cit. on pp. 1, 3).

[Ded92] J. P. Dedieu. “Obreschkoff’s theorem revisited: what convex sets are contained in the set of hyperbolic polynomials?” In: *Journal of Pure and Applied Algebra* 81.3 (1992), pp. 269–278 (cit. on p. 49).

[EL75] P. Erdős and L. Lovász. “Problems and results on 3-chromatic hypergraphs and some related questions”. In: *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II*. North-Holland, Amsterdam, 1975, 609–627. Colloq. Math. Soc. János Bolyai, Vol. 10 (cit. on p. 5).

[ES14] J. Erickson and A. Sidiropoulos. “A near-optimal approximation algorithm for Asymmetric TSP on embedded graphs”. In: *SOCG*. 2014, p. 130 (cit. on pp. 3, 19).

[FGM82] A. M. Frieze, G. Galbiati, and F. Maffioli. “On the worst-case performance of some algorithms for the asymmetric traveling salesman problem”. In: *Networks* 12 (1982), pp. 23–39 (cit. on p. 3).

[FJ15] U. Feige and S. Jozeph. "Separation between estimation and approximation". In: *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*. ACM. 2015, pp. 271–276 (cit. on p. 1).

[FS07] U. Feige and M. Singh. "Improved approximation ratios for traveling salesman tours and path sin directed graphs". In: *APPROX*. 2007, pp. 104–118 (cit. on p. 3).

[Fun+11] W. S. Fung et al. "A general framework for graph sparsification". In: *STOC*. 2011, pp. 71–80 (cit. on p. 21).

[GBS08] A. Ghosh, S. Boyd, and A. Saberi. "Minimizing Effective Resistance of a Graph". In: *SIAM Rev.* 50.1 (Feb. 2008), pp. 37–66 (cit. on p. 27).

[God04] L. A. Goddyn. "Some Open Problems I Like". 2004. URL: `http://www.math.sfu.ca/~goddyn/Problems/problems.html` (cit. on pp. 4, 19).

[Goe+09] M. X. Goemans et al. "A Randomized Rounding Algorithm for the Asymmetric Traveling Salesman Problem". 2009. URL: `http://arxiv.org/abs/0909.0941` (cit. on p. 18).

[Goe06] M. X. Goemans. "Minimum Bounded Degree Spanning Trees". In: *FOCS*. 2006, pp. 273–282 (cit. on p. 61).

[HK70] M. Held and R. M. Karp. "The traveling-salesman problem and minimum spanning trees". In: *Operations Research* 18.6 (1970), pp. 1138–1162 (cit. on pp. 1, 2, 17).

[HO14] N. J. A. Harvey and N. Olver. "Pipage Rounding, Pessimistic Estimators and Matrix Concentration". In: *SODA*. 2014, pp. 926–945 (cit. on pp. 5, 19, 21–23, 44).

[Jae84] F. Jaeger. "On circular flows in graphs, Finite and Infinite Sets". In: *Colloquia Mathematica Societatis Janos Bolyai 37* (1984), pp. 391–402 (cit. on p. 19).

[Kap+05] H. Kaplan et al. "Approximation algorithms for asymmetric TSP by decomposing directed regular multigraphs". In: *J. ACM* (2005), pp. 602–626 (cit. on p. 3).

[Kar72] R. M. Karp. *Reducibility among combinatorial problems*. Springer, 1972 (cit. on p. 1).

[Kar99] D. R. Karger. "Random sampling in cut, flow, and network design problems". In: *MOR* 24.2 (May 1999), pp. 383–413 (cit. on p. 37).

[KS59] R. V. Kadison and I. M. Singer. "Extensions of pure states". In: *American journal of mathematics* (1959), pp. 383–400 (cit. on pp. 5, 38).

[Lov+13] L. M. Lovász et al. "Nowhere-zero 3-flows and modulo k-orientations". In: *J. Comb. Theory, Ser. B* 103.5 (2013), pp. 587–598 (cit. on p. 19).

[LP13] R. Lyons and Y. Peres. *Probability on Trees and Networks*. To appear. Cambridge University Press, 2013 (cit. on pp. 21, 44).

[LPR05] A. Lewis, P. Parrilo, and M. Ramana. "The Lax conjecture is true". In: *Proceedings of the American Mathematical Society* 133.9 (2005) (cit. on p. 55).

[Mir75] L. Mirsky. "A trace inequality of John von Neumann". English. In: *Monatshefte für Mathematik* 79.4 (1975), pp. 303–306 (cit. on p. 12).

[MSS13a] A. Marcus, D. A. Spielman, and N. Srivastava. "Interlacing Families I: Bipartite Ramanujan Graphs of All Degrees". In: *FOCS*. 2013, pp. 529–537 (cit. on pp. 5, 47, 48).

[MSS13b] A. Marcus, D. A. Spielman, and N. Srivastava. "Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem". In: (June 2013). eprint: `1306.3969`. URL: `http://arxiv.org/abs/1306.3969` (cit. on pp. 2, 5, 13, 15, 19, 22, 23, 25, 26, 38–40, 44, 47, 48, 52–55, 59).

[MSS15] A. W. Marcus, N. Srivastava, and D. A. Spielman. "Interlacing Families IV: Bipartite Ramanujan Graphs of All Sizes". In: *arXiv preprint arXiv:1505.08010* (2015) (cit. on p. 5).

[Nas61] C. S. J. A. Nash-Williams. "Edge disjoint spanning trees of finite graphs". In: *J. London Math. Soc.* 36 (1961), pp. 445–45 (cit. on pp. 14, 45).

[OS11] S. Oveis Gharan and A. Saberi. "The Asymmetric Traveling Salesman Problem on Graphs with Bounded Genus". In: *SODA*. 2011, pp. 967–975 (cit. on pp. 3, 18, 19).

[OSS11] S. Oveis Gharan, A. Saberi, and M. Singh. "A randomized rounding approach to the traveling salesman problem". In: *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*. IEEE. 2011, pp. 550–559 (cit. on pp. 2, 39).

[PP14] R. Pemantle and Y. Peres. "Concentration of Lipschitz Functionals of Determinantal and Other Strong Rayleigh Measures". In: *Combinatorics, Probability and Computing* 23 (01 Jan. 2014), pp. 140–160 (cit. on p. 39).

[RT87] P. Raghavan and C. D. Tompson. "Randomized rounding: a technique for provably good algorithms and algorithmic proofs". In: *Combinatorica* 7.4 (Dec. 1987), pp. 365–374 (cit. on p. 18).

[Rud99] M. Rudelson. "Random vectors in the isotropic position". In: *Journal of Functional Analysis* 164.1 (1999), pp. 60–72 (cit. on p. 19).

[SL07] M. Singh and L. C. Lau. "Approximating minimum bounded degree spanning trees to within one of optimal". In: *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. ACM. 2007, pp. 661–670 (cit. on p. 61).

[SLB74] G. L. Shurbet, T. O. Lewis, and T. L. Boullion. "Quadratic Matrix Equations". In: *Ohio Journal of Science* 74 (5 1974), pp. 273–277 (cit. on p. 74).

[SS11] D. A. Spielman and N. Srivastava. "Graph Sparsification by Effective Resistances". In: *SIAM J. Comput.* 40.6 (2011), pp. 1913–1926 (cit. on p. 21).

[ST04] D. A. Spielman and S.-H. Teng. "Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems". In: *STOC*. 2004, pp. 81–90 (cit. on p. 21).

[SV14] M. Singh and N. K. Vishnoi. "Entropy, optimization and counting". In: *STOC*. 2014, pp. 50–59 (cit. on p. 42).

[Sve15] O. Svensson. "Approximating ATSP by Relaxing Connectivity". to appear in FOCS. 2015. URL: `http://arxiv.org/abs/1502.02051` (cit. on p. 3).

[Sze13] M. Szegedy. "The Lovász Local Lemma–A Survey". In: *Computer Science–Theory and Applications*. Springer, 2013, pp. 1–11 (cit. on p. 5).

[Tho12] C. Thomassen. "The weak 3-flow conjecture and the weak circular flow conjecture". In: *Journal of Combinatorial Theory, Series B* 102.2 (2012), pp. 521–529 (cit. on p. 19).

[Wag11] D. G. Wagner. "Multivariate stable polynomials: theory and applications". In: *Bulletin of the American Mathematical Society* 48.1 (Jan. 2011), pp. 53–53 (cit. on p. 50).

[Wea04] N. Weaver. "The Kadison–Singer problem in discrepancy theory". In: *Discrete mathematics* 278.1 (2004), pp. 227–239 (cit. on pp. 5, 19, 38, 59).