Advancements in Combinatorial Optimization and Spectral Graph Theory: Approaches to Permanents, Metric Embeddings, and Geometric Graphs

Farzam Ebrahimnejad

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:
Shayan Oveis Gharan, Chair
James R. Lee, Chair
Simon S. Du

Program Authorized to Offer Degree:
Computer Science & Engineering

University of Washington

Abstract

Advancements in Combinatorial Optimization and Spectral Graph Theory: Approaches to Permanents, Metric Embeddings, and Geometric Graphs

Farzam Ebrahimnejad

Chairs of the Supervisory Committee:
Shayan Oveis Gharan and James R. Lee
Computer Science & Engineering

In this thesis I will cover the five research papers that I co-authored during my Ph.D. studies. First, I will cover my results on planar and geometric graphs. Namely, in the papers "On planar graphs of uniform polynomial growth" and "Non-existence of annular separators in geometric graphs" we resolve several conjectures and open problems about spectral properties of a certain family of geometric graphs with uniform polynomial growth. Next, I will cover the paper "Multiscale entropic regularization for MTS on general metric spaces" which positively answers an open question in online algorithms by providing an $O(\log^n)$-competitive algorithm for the metrical task systems problem which is purely based on a mirror descent framework. Next, I will focus on our result on "Counting and sampling perfect matchings in regular expanding non-bipartite graphs" in which we provide algorithms to efficiently sample perfect matchings in those families of graphs. Finally, I will discuss our latest result, "On approximability of the permanent of PSD matrices", in which we improve both the lower-bound and upper-bounds for approximating the permanent of PSD matrices.

Contents

Acknowledgements

First and foremost, I would like to dedicate this thesis to my parents for being so loving and supportive for my entire life, but especially during my PhD studies. I hope I will be able to reunite with them after all these years soon.
Next, I would like to thank my advisors, Shayan and James, from both of whom I have learned a lot during the past years, and not just academically. I would like to especially thank Shayan for being so caring and supportive even in the personal aspects of my life. He has been like family to me these past years.
I would also like to thank my family and all my friends who have been there for me these past few years, in my great and not-so-great days: Keivan, Sobhan, Siavash, Hadi, Farzam, Artin, Sina, Reza, Moein, Negar, Alireza, Alireza, Behrooz, Roozbeh, Rajab, Abtin, Mahbod, Ehsan, Farnaz, Abbas, Swati, Hoda, Kuikui, Farzad, Shiva, Rustin, Parsoa, Tarun, Ansh, Siva, Iman, Ali, Arash, Majid, Mobin, Zahra, Mojdeh, Hessam, Reihaneh, and Amin. I want to also thank the rest of the theory group and the CSE department for being so helpful during my PhD studies.
Finally, I wish to remember my two very dear friends, Ali and Hooman, whom I have lost in the past few years.

1 Introduction

In this thesis I will cover the five research papers that I co-authored during my Ph.D. studies. During the past years I have worked on several research projects covering a wide range of topics in theoretical computer science. But the unifying theme of my research projects was that it has been focused on solving combinatorial problems with algebraic and analytical techniques. More specifically, my research has touched on problems in approximation algorithms, online algorithms, spectral graph theory, geometry of polynomials, probability, metric geometry, and convex optimization.

In Sections 2 and 3, I will cover my results on planar and geometric graphs. Namely, in the papers "On planar graphs of uniform polynomial growth" and "Non-existence of annular separators in geometric graphs" we resolve several conjectures and open problems about spectral properties of a certain family of geometric graphs with uniform polynomial growth. In these projects we present a novel construction of planar and sphere-packed graphs of uniform polynomial growth, and show that these constructions resolve several open problems about the spectral features of such graphs. The construction is based on recursive tilings of the d-dimensional cube.

Next, in Section 4 I will cover the paper "Multiscale entropic regularization for MTS on general metric spaces" which positively answers an open question in online algorithms by providing an $O(\log^n)$-competitive algorithm for the metrical task systems (MTS) problem which is purely based on a mirror descent framework. The main idea of this project is that given an arbitrary metric space $\mathcal{M}$, we construct a Directed Acycling Graph (DAG) on top of $\mathcal{M}$ capturing its geometry, and then we show that running a generalization of the mirror descent framework previously developed for MTS would result in a state-of-the-art $O(\log^2 n)$ bound of [CL19], which was also later shown to be tight by [BCR23].

In Section 5 I will focus on our result on "Counting and sampling perfect matchings in regular expanding non-bipartite graphs" in which we provide algorithms to efficiently sample perfect matchings in those families of graphs.

Finally, in Section 6 I will discuss our latest result, "On approximability of the permanent of PSD matrices", in which we improve both the lower-bound and upper-bounds for approximating the permanent of PSD matrices. Previously the best known approximation algorithm for this problem was by a factor of $e^{(\gamma+1)n}$ due to [AGOS17], and no truly exponential lower-bound for the problem was known. In this project we provide a $e^{\gamma n}$ lower-bound (given P $\neq$ NP) and further exponentially improve the upper-bound by a factor of $e^{10^{-4}n}$. Our lower-bound technique is built upon the work of [BGG+23] on the hardness of approximation of matrix $p \to q$ norms. Our upper-bound method also builds upon the work of [AGOS17], first computing the same semi-definite program as theirs, but possibly outputting a smaller number if the given matrix has a small "smooth" rank, which we define formally in Section 6.

2 On planar graphs of uniform polynomial growth

2.1 Introduction

Say that a graph G has *uniform polynomial growth of degree d* if the cardinality of all balls of radius r in the graph metric lie between cr^d and Cr^d for two absolute constants $C > c > 0$, for every $r > 0$. Say that a graph has *nearly-uniform polyomial growth of degree d* if the cardinality of balls is trapped between $(\log r)^{-C} r^d$ and $(\log r)^C r^d$ for some universal constant $C \geqslant 1$.

Planar graphs of uniform (or nearly-uniform) polynomial volume growth of degree $d > 2$ arise in a number of contexts. In particular, they appear in the study of random triangulations in 2D quantum gravity [ADJ97] and as combinatorial approximations to the boundaries of 3-dimensional hyperbolic groups in geometric group theory (see, e.g., [BK02]).

When the dimension of volume growth disagrees with the topological dimension, one sometimes witnesses certain geometrically or spectrally degenerate behaviors. For instance, it is known that random planar triangulations of the 2-sphere have nearly-uniform polynomial volume growth of degree 4 (in an appropriate statistical, asymptotic sense) [Ang03]. The distributional limit (see Section 2.1.1) of such graphs is called the uniform infinite planar triangulation (UIPT). But this 4-dimensional volume growth does not come with 4-dimensional isoperimetry: With high probability, a ball in UIPT of radius r about a vertex v can be separated from the complement of a $2r$ ball about v by removing a set of size $O(r)$. And, indeed, Benjamini and Papasoglu [BP11] showed that this phenomenon holds generally: Such annular separators of size $O(r)$ exist in all planar graphs with uniform polynomial volume growth.

Similarly, it is known that diffusion on UIPT is *anomalous*. Specifically, the random walk on UIPT is almost surely subdiffusive. In other words, if $\{X_t\}$ is the random walk and d_G denotes the graph metric, then $\mathbb{E}\, d_G(X_0, X_t) \leqslant t^{1/2-\varepsilon}$ for some $\varepsilon > 0$. This was established by Benjamini and Curien [BC13]. In [Lee17], it is shown that on *any* unimodular random planar graph with nearly-uniform polynomial growth of degree $d > 3$ (in a suitable statistical sense), the random walk is subdiffusive. So again, a disagreement between the dimension of volume growth and the topological dimension results in a degeneracy typical in the geometry of fractals (see, e.g., [Bar98]).

Finally, consider a seminal result of Benjamini and Schramm [BS01]: If (G, ρ) is the local distributional limit of a sequence of finite planar graphs with uniformly bounded degrees, then (G, ρ) is almost surely recurrent. In this sense, any such limit is spectrally (at most) two-dimensional. This was extended by Gurel-Gurevich and Nachmias [GN13] to unimodular random graphs with an exponential tail on the degree of the root, making it applicable to UIPT. Benjamini [Ben13] has conjectured that this holds for every planar graph with uniform polynomial volume. We construct a family of counterexamples. Our focus on rational degrees of growth is largely for simplicity; suitable variants of our construction should yield similar results for all real $d > 2$.

Theorem 2.1. *For every rational $d > 2$, there is a* transient *planar graph with uniform polynomial growth of degree d.*

Conversely, it is well-known that *any graph* with growth rate $d \leqslant 2$ is recurrent. The examples underlying Theorem 2.1 cannot be unimodular. Nevertheless, we construct unimodular examples addressing some of the issues raised above. Angel and Nachmias (unpublished) showed the existence, for every $\varepsilon > 0$ sufficiently small, of a unimodular random planar graph (G, ρ) on which the random walk is almost surely diffusive, and which almost surely satisfies

$$\lim_{r\to\infty} \frac{\log |B_G(\rho, r)|}{\log r} = 3 - \varepsilon.$$

Here, $B_G(\rho, r)$ is the graph ball around ρ of radius r. In other words, r-balls have an asymptotic growth rate of $r^{3-\varepsilon}$ as $r \to \infty$.

The authors of [BP11] asked whether in planar graphs with uniform growth of degree $d \geqslant 2$, the speed of the walk should be at most $t^{1/d+o(1)}$. We recall the following weaker theorem.

Theorem 2.2 ([Lee17]). *Suppose (G, ρ) is a unimodular random planar graph and G almost surely has uniform polynomial growth of degree d. Then:*

$$\mathbb{E}\left[d_G(X_0, X_t) \mid X_0 = \rho\right] \lesssim t^{1/\max(2,d-1)}.$$

We construct examples where this dependence is nearly tight.

Theorem 2.3. *For every rational $d \geqslant 2$ and $\varepsilon > 0$, there is a constant $c(\varepsilon) > 0$ and a unimodular random planar graph (G, ρ) such that G almost surely has uniform polynomial growth of degree d, and*

$$\mathbb{E}\left[d_G(X_0, X_t) \mid X_0 = \rho\right] \geqslant c(\varepsilon) t^{1/(\max(2,d-1)+\varepsilon)}.$$

Finally, let us address another question from [BP11]. In conjunction with the existence of small annular separators, the authors asked whether a planar graph with uniform polynomial growth of degree $d > 2$ can be such that the complement of every ball is connected. For example, in UIPT, there are "baby universes" connected to the graph via a thin neck that can be cut off by removing a small graph ball.

Theorem 2.4. *For every rational $d \geqslant 2$, there is a unimodular random planar graph (G, ρ) such that almost surely:*

1. *G has uniform polynomial growth of degree d.*
2. *The complement of every graph ball in G is connected.*

Strongly recurrent graphs and annular resistances. Our unimodular constructions have the property that the "Einstein relations" (see, e.g., [Bar98]) for various dimensional exponents do not hold. In particular, this implies that the graphs we construct are not strongly recurrent (see, e.g., [KM08]). Indeed, the effective resistance across annuli can be made very small (see Section 2.2.3 for the definition of effective resistance).

Theorem 2.5. *For every $\varepsilon > 0$ and $d \geqslant 3$, there is a unimodular random planar graph (G, ρ) that almost surely has uniform polynomial volume growth of degree d and, moreover, almost surely satisfies*

$$R_{\mathrm{eff}}^G\left(B_G(\rho, R) \leftrightarrow V(G) \setminus B_G(\rho, 2R)\right) \leqslant C(\varepsilon) R^{-(1-\varepsilon)}, \quad \forall R \geqslant 1, \tag{2.1}$$

where $C(\varepsilon) \geqslant 1$ is a constant depending only on ε.

Note that the existence of annular separators of size $O(R)$ mentioned previously gives $R_{\mathrm{eff}}^G(B_G(\rho, R) \leftrightarrow V(G) \setminus B_G(\rho, 2R)) \gtrsim R^{-1}$ by the Nash-Williams inequality. Moreover, recall that since the graph (G, ρ) from Theorem 2.5 is unimodular and planar, it must be almost surely recurrent (cf. [BS01]) Therefore the electrical flow witnessing (2.1) cannot spread out "isotropically" from $B_G(\rho, R)$ to $B_G(\rho, 2R)$. Indeed, if one were able to send a flow roughly uniformly from $B_G(\rho, 2^i)$ to $B_G(\rho, 2^{i+1})$, then these electrical flows would chain to to give

$$R_{\mathrm{eff}}^G\left(\rho \leftrightarrow V(G) \setminus B_G(\rho, 2^i)\right) \lesssim \sum_{j \leqslant i} 2^{-(1-\varepsilon)j},$$

and taking $j \to \infty$ would show that G is transient.

We remark on one other interesting feature of Theorem 2.5. Suppose that Γ is a Gromov hyperbolic group whose visual boundary $\partial_\infty\Gamma$ is homeomorphic to the 2-sphere $\mathbb{S}^2$. The authors of [BK02] construct a family $\{G_n : n \geqslant 1\}$ of discrete approximations to $\partial_\infty\Gamma$ such that each G_n is a planar graph and the family $\{G_n\}$ has uniform polynomial volume growth.[1] They show that if there is a constant $c > 0$ so that the annuli in G_n satisfy uniform effective resistance estimates of the form

$$R_{\mathrm{eff}}^{G_n}(B_{G_n}(x,R) \leftrightarrow V(G_n) \setminus B_{G_n}(x,2R)) \geqslant c, \quad \forall 1 \leqslant R \leqslant \mathrm{diam}(G_n)/10,\ x \in V(G_n),\ \forall n \geqslant 1,$$

then $\partial_\infty\Gamma$ is quasisymmetric to $\mathbb{S}^2$ (cf. [BK02, Thm 11.1].)

In particular, if it were to hold that for any (infinite) planar graph G with uniform polynomial growth we have

$$R_{\mathrm{eff}}^{G}(B_G(x,R) \leftrightarrow V(G) \setminus B_G(x,2R)) \geqslant c > 0, \quad \forall R \geqslant 1, x \in V(G),$$

then it would confirm positively Cannon's conjecture from geometric group theory. Theorem 2.5 exhibits graphs for which this fails in essentially the strongest way possible.

2.1.1 Preliminaries

We will consider primarily connected, undirected graphs $G = (V, E)$, which we equip with the associated path metric d_G. We will sometimes write $V(G)$ and $E(G)$, respectively, for the vertex and edge sets of G. If $U \subseteq V(G)$, we write $G[U]$ for the subgraph induced on U.

For $v \in V$, let $\deg_G(v)$ denote the degree of v in G. Let $\mathrm{diam}(G) := \sup_{x,y\in V} d_G(x,y)$ denote the diameter (which is only finite for G finite). For $v \in V$ and $r \geqslant 0$, we use $B_G(v,r) = \{u \in V : d_G(u,v) \leqslant r\}$ to denote the closed ball in G. For subsets $S, T \subseteq V$, we write $d_G(S,T) := \inf\{d_G(s,t) : s \in S, t \in T\}$.

Say that an infinite graph G has *uniform volume growth of rate $f(r)$* if there exist constants $C, c > 0$ such that

$$c f(r) \leqslant |B_G(v,r)| \leqslant C f(r) \qquad \forall v \in V, r \geqslant 1.$$

A graph has *uniform polynomial growth of degree d* if it has uniform volume growth of rate $f(r) = r^d$, and has *uniform polynomial growth* if this holds for some $d > 0$.

For two expressions A and B, we use the notation $A \lesssim B$ to denote that $A \leqslant CB$ for some *universal* constant C. The notation $A \lesssim_\gamma B$ denotes that $A \leqslant C(\gamma)B$ where $C(\gamma)$ is a number depending only on the parameter γ. We write $A \asymp B$ for the conjunction $A \lesssim B \wedge B \lesssim A$.

Distributional limits of graphs

We briefly review the weak local topology on random rooted graphs. One may consult the extensive reference of Aldous and Lyons [AL07], and [BC12] for the corresponding theory of reversible random graphs. The paper [BS01] offers a concise introduction to distributional limits of finite planar graphs. We briefly review some relevant points.

[1] More precisely, for the boundary of a hyperbolic group as above, one can choose a sequence of approximations with this property.

Let $\mathcal{G}$ denote the set of isomorphism classes of connected, locally finite graphs; let $\mathcal{G}_\bullet$ denote the set of *rooted* isomorphism classes of *rooted*, connected, locally finite graphs. Define a metric on $\mathcal{G}_\bullet$ as follows: $\mathbb{d}_{\text{loc}}((G_1, \rho_1), (G_2, \rho_2)) = 1/(1 + \alpha)$, where

$$\alpha = \sup\left\{r > 0 : B_{G_1}(\rho_1, r) \cong_\rho B_{G_2}(\rho_2, r)\right\} ,$$

and we use $\cong_\rho$ to denote rooted isomorphism of graphs. $(\mathcal{G}_\bullet, \mathbb{d}_{\text{loc}})$ is a separable, complete metric space. For probability measures $\{\mu_n\}, \mu$ on $\mathcal{G}_\bullet$, write $\{\mu_n\} \Rightarrow \mu$ when μ_n converges weakly to μ with respect to $\mathbb{d}_{\text{loc}}$.

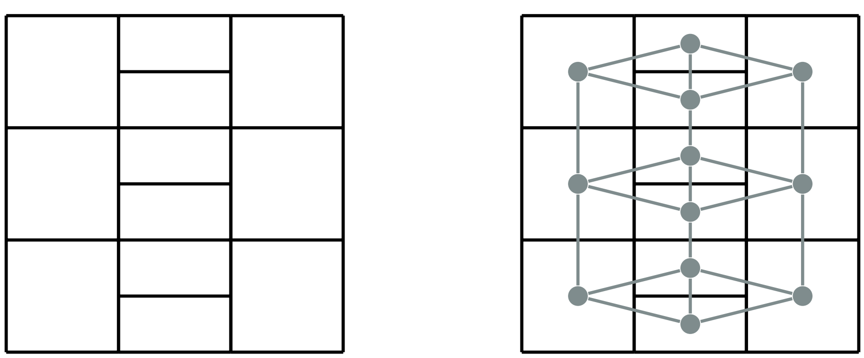
A random rooted graph (G, X_0) is said to be *reversible* if (G, X_0, X_1) and (G, X_1, X_0) have the same law, were X_1 is a uniformly random neighbor of X_0 in G. A random rooted graph (G, ρ) is said to be *unimodular* if it satisfies the Mass Transport Principle (see, e.g., [AL07]). For our purposes, it suffices to note that if $\mathbb{E}[\deg_G(\rho)] < \infty$, then (G, ρ) is reversible if and only if the random rooted graph $(\tilde{G}, \tilde{\rho})$ is unimodular, where $(\tilde{G}, \tilde{\rho})$ has the law of (G, ρ) biased by $\deg_G(\rho)$ [BC12, Prop. 2.5].

If $\{(G_n, \rho_n)\} \Rightarrow (G, \rho)$, we say that (G, ρ) is the *distributional limit* of the sequence $\{(G_n, \rho_n)\}$, where we have conflated random variables with their laws in the obvious way. Consider a sequence $\{G_n\} \subseteq \mathcal{G}$ of finite graphs, and let ρ_n denote a uniformly random element of $V(G_n)$. Then $\{(G_n, \rho_n)\}$ is a sequence of $\mathcal{G}_\bullet$-valued random variables, and one has the following: If $\{(G_n, \rho_n)\} \Rightarrow (G, \rho)$, then (G, ρ) is unimodular. Equivalently, if $\{(G_n, \rho_n)\}$ is a sequence of connected finite graphs and $\rho_n \in V(G_n)$ is chosen according to the stationary measure of G_n, then if $\{(G_n, \rho_n)\} \implies (G, \rho)$, it holds that (G, ρ) is a reversible random graph.

2.2 A transient planar graph of uniform polynomial growth

We begin by constructing a transient planar graph with uniform polynomial growth of degree $d > 2$. In Section 2.3, this construction is generalized to any rational $d > 2$.

2.2.1 Tilings and dual graphs



(a) A tiling of the unit square

(b) The associated dual graph

Figure 1: Tilings and their dual graph

Our constructions are based on planar tilings by rectangles. A *tile* is an axis-parallel closed rectangle $A \subseteq \mathbb{R}^2$. We will encode such a tile as a triple $(p(A), \ell_1(A), \ell_2(A))$, where $p(A) \in \mathbb{R}^2$

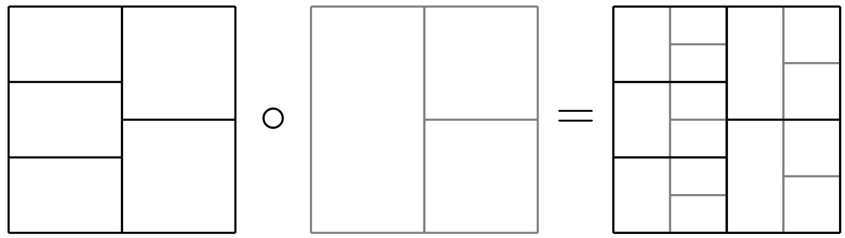


Figure 2: An example of the tiling product $S \circ T$

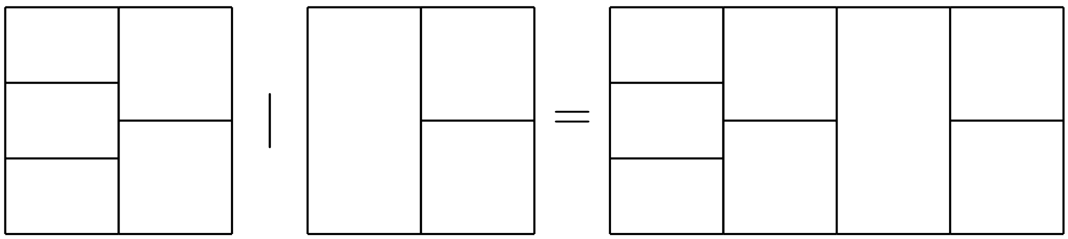


Figure 3: An example of the tiling concatenation $S \mid T$

denotes its bottom-left corner, $\ell_1(A)$ its width (length of its projection onto the x-axis), and $\ell_2(A)$ its height (length of its projection onto the y-axis). A *tiling* $\boldsymbol{T}$ is a finite collection of interior-disjoint tiles. Denote $[\![\boldsymbol{T}]\!] := \bigcup_{A \in \boldsymbol{T}} A$. If $R \subseteq \mathbb{R}^2$, we say that $\boldsymbol{T}$ *is a tiling of* R if $[\![\boldsymbol{T}]\!] = R$. See Figure 7(a) for a tiling of the unit square.

We associate to a tiling its *dual graph* $G(\boldsymbol{T})$ with vertex set $\boldsymbol{T}$ and with an edge between two tiles $A, B \in \boldsymbol{T}$ whenever $A \cap B$ has Hausdorff dimension one; in other words, A, B are tangent, but not only at a corner. Denote by $\mathcal{T}$ the set of all tilings of the unit square. See Figure 1(b). For the remainder of the paper, we will consider only tilings $\boldsymbol{T}$ for which $G(\boldsymbol{T})$ is connected.

Definition 2.6 (Tiling product)**.** For $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{T}$, define the product $\boldsymbol{S} \circ \boldsymbol{T} \in \mathcal{T}$ as the tiling formed by replacing every tile in $\boldsymbol{S}$ by an (appropriately scaled) copy of $\boldsymbol{T}$. More precisely: For every $A \in \boldsymbol{S}$ and $B \in \boldsymbol{T}$, there is a tile $R \in \boldsymbol{S} \circ \boldsymbol{T}$ with $\ell_i(R) := \ell_i(A)\ell_i(B)$, and

$$p_i(R) := p_i(A) + p_i(B)\ell_i(A),$$

for each $i \in \{1, 2\}$. See Figure 2.

If $\boldsymbol{T} \in \mathcal{T}$ and $n \geqslant 0$, we will use $\boldsymbol{T}^n := \boldsymbol{T} \circ \cdots \circ \boldsymbol{T}$ to denote the n-fold tile product of $\boldsymbol{T}$ with itself. The following observation shows that this is well-defined.

Observation 2.7. The tiling product is associative: $(\boldsymbol{S} \circ \boldsymbol{T}) \circ \boldsymbol{U} = \boldsymbol{S} \circ (\boldsymbol{T} \circ \boldsymbol{U})$ for all $S, T, U \in \mathcal{T}$. Moreover, if $\boldsymbol{I} \in \mathcal{T}$ consists of the single tile $[0, 1]^2$, then $\boldsymbol{T} \circ \boldsymbol{I} = \boldsymbol{I} \circ \boldsymbol{T}$ for all $\boldsymbol{T} \in \mathcal{T}$.

Definition 2.8 (Tiling concatenation)**.** Suppose that $\boldsymbol{S}$ is a tiling of a rectangle R and $\boldsymbol{T}$ is a tiling of a rectangle R' and the heights of R and R' coincide. Let R'' denote the translation of R' for which the left edge of R'' coincides with the right edge of R, and denote by $\boldsymbol{S} \mid \boldsymbol{T}$ the induced tiling of the rectangle $R \cup R''$. See Figure 3.

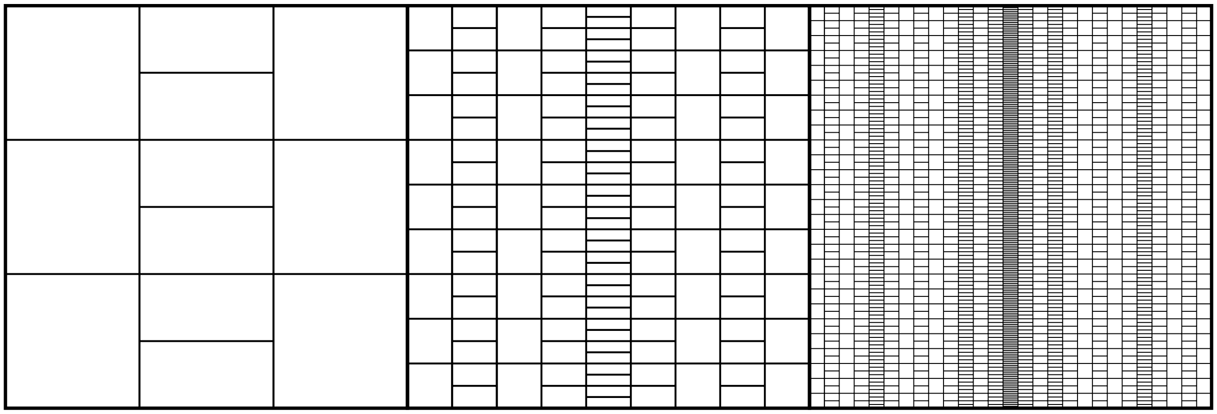


Figure 4: The tiling $\boldsymbol{H}^1 \mid \boldsymbol{H}^2 \mid \boldsymbol{H}^3$

Let $\boldsymbol{H}$ denote the tiling in Figure 7(a), and define $\mathcal{H}_n := G(\boldsymbol{H}^0 \mid \boldsymbol{H}^1 \mid \cdots \mid \boldsymbol{H}^n)$; see Figure 4, where we have omitted $\boldsymbol{H}^0$ for ease of illustration. The next theorem represents our primary goal for the remainder of this subsection. Note that $\boldsymbol{H}^0 = \{\rho\}$ consists of a single tile, and that $\{(\mathcal{H}_n, \rho)\}$ forms a Cauchy sequence in $(\mathcal{G}_\bullet, \mathbb{d}_{\mathrm{loc}})$, since $(\mathcal{H}_n, \rho)$ is naturally a rooted subgraph of $(\mathcal{H}_{n+1}, \rho)$. Letting $\mathcal{H}_\infty$ denote its limit, we will establish the following.

Theorem 2.9. *The infinite planar graph $\mathcal{H}_\infty$ transient and has uniform polynomial volume growth.*

Uniform growth is established in Lemma 2.16 and transience in Corollary 2.22.

2.2.2 Volume growth

The next lemma is straightforward.

Lemma 2.10. *Consider $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{T}$ and $G = G(\boldsymbol{S} \circ \boldsymbol{T})$. For any $X \in \boldsymbol{S} \circ \boldsymbol{T}$, it holds that $|B_G(X, \mathrm{diam}(G(\boldsymbol{T})))| \geqslant |\boldsymbol{T}|$.*

If $\boldsymbol{T}$ is a tiling, let us partition the edge set $E(G(\boldsymbol{T})) = E_1(\boldsymbol{T}) \cup E_2(\boldsymbol{T})$ into horizontal and vertical edges. For $A \in \boldsymbol{T}$ and $i \in \{1, 2\}$, let $N_{\boldsymbol{T}}(A, i)$ denote the set of tiles adjacent to A in $G(\boldsymbol{T})$ along the ith direction, and denote $N_{\boldsymbol{T}}(A) := N_{\boldsymbol{T}}(A, 1) \cup N_{\boldsymbol{T}}(A, 2)$. Moreover, we define:

$$\begin{aligned}
\alpha_{\boldsymbol{T}}(A, i) &:= \max\left\{\frac{\ell_j(A)}{\ell_j(B)} : B \in N_{\boldsymbol{T}}(A, i), j \in \{1, 2\}\right\}, \quad i \in \{1, 2\} \\
\alpha_{\boldsymbol{T}}(A) &:= \max_{i \in \{1,2\}} \alpha_{\boldsymbol{T}}(A, i) \\
\alpha_{\boldsymbol{T}} &:= \max\{\alpha_{\boldsymbol{T}}(A) : A \in \boldsymbol{T}\} \\
L_{\boldsymbol{T}} &:= \max\{\ell_i(A) : A \in \boldsymbol{T}, i \in \{1, 2\}\}.
\end{aligned}$$

We will take $\alpha_{\boldsymbol{T}} := 1$ if $\boldsymbol{T}$ contains a single tile. It is now straightforward to check that $\alpha_{\boldsymbol{T}}$ bounds the degrees in $G(\boldsymbol{T})$.

Lemma 2.11. *For a tiling $\boldsymbol{T}$ and $A \in \boldsymbol{T}$, it holds that*

$$\deg_{G(\boldsymbol{T})}(A) \leqslant 4(1 + \alpha_{\boldsymbol{T}}) \leqslant 8\alpha_{\boldsymbol{T}}.$$

Proof. After accounting for the four corners of A, every other tile $B \in N_{\boldsymbol{T}}(A,i)$ intersects A in a segment of length at least $\ell_i(B) \geqslant \ell_i(A)/\alpha_{\boldsymbol{T}}$. The second inequality follows from $\alpha_{\boldsymbol{T}} \geqslant 1$. □

Lemma 2.12. *Consider $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{T}$ and let $G = G(\boldsymbol{S} \circ \boldsymbol{T})$. Then for any $X \in \boldsymbol{S} \circ \boldsymbol{T}$, it holds that*

$$|B_G(X, 1/(\alpha_{\boldsymbol{S}}^4 L_{\boldsymbol{T}}))| \leqslant 192\alpha_{\boldsymbol{S}}^2 |\boldsymbol{T}|. \tag{2.2}$$

Proof. For a tile $Y \in \boldsymbol{S} \circ \boldsymbol{T}$, let $\hat{Y} \in \boldsymbol{S}$ denote the unique tile for which $Y \subseteq \hat{Y}$. Let us also define

$$\tilde{N}_{\boldsymbol{S}}(\hat{X}) := \{\hat{X}\} \cup N_{\boldsymbol{S}}(\hat{X}) \cup N_{\boldsymbol{S}}\left(N_{\boldsymbol{S}}(\hat{X},1),2\right) \cup N_{\boldsymbol{S}}\left(N_{\boldsymbol{S}}(\hat{X},2),1\right),$$

which is the set of vertices that can be reached from $\hat{X}$ by following at most one edge in each direction.

We will show that

$$\left[\!\left[B_G(X, 1/(\alpha_{\boldsymbol{S}}^3 L_{\boldsymbol{T}}))\right]\!\right] \subseteq \left[\!\left[\tilde{N}_{\boldsymbol{S}}(\hat{X})\right]\!\right]. \tag{2.3}$$

It follows that

$$|B_G(X, 1/(\alpha_{\boldsymbol{S}}^3 L_{\boldsymbol{T}}))| \leqslant |\boldsymbol{T}| \cdot |\tilde{N}_{\boldsymbol{S}}(\hat{X})| \leqslant |\boldsymbol{T}| \cdot 3\left(\max_{A \in \boldsymbol{S}} \deg_{G(\boldsymbol{S})}(A)\right)^2,$$

and then (3.3) follows from Lemma 3.14.

To establish (3.5), consider any path $\langle X = X_0, X_1, X_2, \ldots, X_h\rangle$ in G with $\hat{X}_h \notin \tilde{N}_{\boldsymbol{S}}(\hat{X})$. Let $k \leqslant h$ be the smallest index for which $\hat{X}_k \notin \tilde{N}_{\boldsymbol{S}}(\hat{X})$. Then:

$$X_0, X_1, \ldots, X_{k-1} \subseteq \left[\!\left[\tilde{N}_{\boldsymbol{S}}(\hat{X})\right]\!\right] \tag{2.4}$$

$$X_{k-1} \cap \left(\partial \left[\!\left[\tilde{N}_{\boldsymbol{S}}(\hat{X})\right]\!\right] \cap (0,1)^2\right) \neq \emptyset. \tag{2.5}$$

Now (3.6) implies that

$$\ell_i(X_j) \leqslant L_{\boldsymbol{T}} \ell_i(\hat{X}_j) \leqslant L_{\boldsymbol{T}} \alpha_{\boldsymbol{S}}^2 \ell_i(\hat{X}), \qquad j \leqslant k-1, i \in \{1,2\}. \tag{2.6}$$

And (3.7) shows that

$$\sum_{j=0}^{k-1} \ell_i(X_j) \geqslant \min\left\{\ell_i(Y) : Y \in \tilde{N}_{\boldsymbol{S}}(\hat{X})\right\} \geqslant \ell_i(\hat{X})/\alpha_{\boldsymbol{S}}^2, \qquad i \in \{1,2\}. \tag{2.7}$$

To clarify why this is true, note that

$$\hat{X} + \left[-\ell_1(\hat{X})/\alpha_{\boldsymbol{S}}^2, \ell_1(\hat{X})/\alpha_{\boldsymbol{S}}^2\right] \times \left[-\ell_2(\hat{X})/\alpha_{\boldsymbol{S}}^2, \ell_2(\hat{X})/\alpha_{\boldsymbol{S}}^2\right] \subseteq \left[\!\left[\tilde{N}_{\boldsymbol{S}}(\hat{X})\right]\!\right],$$

where '+' here is the Minkowski sum $R + S := \{r + s : r \in R, s \in S\}$. Indeed, this inequality motivates our definition of the "ℓ_∞ neighborhood" $\tilde{N}$ above.

Combining (3.8) and (3.9) now gives

$$h - 1 \geqslant k - 1 \geqslant \frac{1}{\alpha_{\boldsymbol{S}}^4 L_{\boldsymbol{T}}},$$

completing the proof. □

Consider a tiling $\boldsymbol{T} \in \mathcal{T}$ and for $i \in \{1,2\}$, let $\boldsymbol{T}_i \subseteq \boldsymbol{T}$ denote the set of tiles that touch a side of $[0,1]^2$ parallel to the ith axis. Define

$$\eta_{\boldsymbol{T}} := \max \left\{ \frac{\ell_j(A)}{\ell_j(B)} : A, B \in \boldsymbol{T}_i, i, j \in \{1,2\} \right\}.$$

Lemma 2.13. *For any $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{T}$, it holds that*

$$\alpha_{\boldsymbol{S} \circ \boldsymbol{T}} \leqslant \max\left(\eta_{\boldsymbol{T}} \alpha_{\boldsymbol{S}}, \alpha_{\boldsymbol{T}}\right).$$

Proof. Consider $A \in \boldsymbol{S} \circ \boldsymbol{T}$ and $B \in N_{\boldsymbol{S} \circ \boldsymbol{T}}(A)$. Let $\hat{A}, \hat{B} \in \boldsymbol{S}$ be the unique tiles for which $A \subseteq \hat{A}$ and $B \subseteq \hat{B}$. If $\hat{A} = \hat{B}$, then

$$\frac{\ell_j(B)}{\ell_j(A)} \leqslant \alpha_{\boldsymbol{T}}, \quad j \in \{1,2\}.$$

If $\hat{A} \neq \hat{B}$, then A, B must both originate from tiles in $\boldsymbol{T}_i$ for some $i \in \{1,2\}$, hence

$$\frac{\ell_j(A)}{\ell_j(B)} \leqslant \eta_{\boldsymbol{T}} \frac{\ell_j(\hat{A})}{\ell_j(\hat{B})} \leqslant \eta_{\boldsymbol{T}} \alpha_{\boldsymbol{S}}, \quad j \in \{1,2\}.$$ □

We can now analyze the graphs $\{\boldsymbol{H}^n : n \geqslant 0\}$.

Lemma 2.14. *For $n \geqslant 0$, we have $|\boldsymbol{H}^n| = 12^n$, and $3^n \leqslant \operatorname{diam}(\boldsymbol{H}^n) \leqslant 3^{n+1}$.*

Proof. The first claim is straightforward by induction. For the second claim, note that $\ell_1(A) = 3^{-n}$ for every $A \in \boldsymbol{H}^n$. Moreover, there are 3^n tiles touching the left-most boundary of $[0,1]^2$. Therefore to connect any $A, B \in \boldsymbol{H}^n$ by a path in $G(\boldsymbol{H}^n)$, we need only go from A to the left-most column in at most 3^n steps, then use at most 3^n steps of the column, and finally move at most 3^n steps to B. □

Lemma 2.15. *For any $n \geqslant 0$, it holds that*

$$\left|B_{G(\boldsymbol{H}^n)}(A, r)\right| \asymp r^{\log_3(12)}, \qquad \forall A \in \boldsymbol{H}^n, 1 \leqslant r \leqslant \operatorname{diam}(G(\boldsymbol{H}^n)).$$

Proof. Writing $\boldsymbol{H}^n = \boldsymbol{H}^{n-k} \circ \boldsymbol{H}^k$ and employing Lemma 3.13 together with Lemma 2.14 gives

$$\left|B_{G(\boldsymbol{H}^n)}(A, 3^{k+1})\right| \geqslant |\boldsymbol{H}^k| = 12^k, \qquad \forall A \in \boldsymbol{H}^n, k \in \{0,1,\ldots,n\}.$$

The desired lower bound now follows using monotonicity of $|B_{G(\boldsymbol{H}^n)}(A, r)|$ with respect to r.

To prove the upper bound, first note $\eta_{\boldsymbol{H}} = 1$. Therefore Lemma 2.13 shows that $\alpha_{\boldsymbol{H}^n} \leqslant \alpha_{\boldsymbol{H}} = 2$. Moreover, we have $L_{\boldsymbol{H}^k} = 3^{-k}$, hence Lemma 3.15 gives

$$\left|B_{G(\boldsymbol{H}^n)}(A, 3^k/8)\right| \leqslant 256 \cdot 12^k, \qquad \forall A \in \boldsymbol{H}^n, k \in \{0,1,\ldots,n\},$$

completing the proof. □

Finally, this allows us to establish a uniform polynomial growth rate for $\mathcal{H}_\infty$.

Lemma 2.16. *It holds that*

$$|B_{\mathcal{H}_\infty}(v, r)| \asymp r^{\log_3(12)} \qquad \forall v \in V(\mathcal{H}_\infty), r \geqslant 1.$$

Proof. Recall first the natural identification $\boldsymbol{H}^n \hookrightarrow V(\mathcal{H}_\infty)$ under which $V(\mathcal{H}_\infty) = \bigcup_{n\geqslant 0} \boldsymbol{H}^n$ is a partition. Consider $v \in V(\mathcal{H}_\infty)$ and let $n \geqslant 0$ be such that $v \in \boldsymbol{H}^n$. Now Lemma 2.15 in conjunction with Lemma 2.14 yields the bounds:

$$\begin{aligned}
&|B_{\mathcal{H}_\infty}(v,r)| \geqslant |B_{\mathcal{H}_\infty}(v,r) \cap \boldsymbol{H}^n| \gtrsim r^{\log_3(12)} && r \leqslant 3^{n+3} \\
&|B_{\mathcal{H}_\infty}(v,r)| \geqslant |\boldsymbol{H}^k| = 12^k \gtrsim r^{\log_3(12)} && r \in [3^{k+3}, 3^{k+4}), k \geqslant n \\
&|B_{\mathcal{H}_\infty}(v,r)| = |B_{\mathcal{H}_\infty}(v,r) \cap \boldsymbol{H}^{n-1}| + |B_{\mathcal{H}_\infty}(v,r) \cap \boldsymbol{H}^n| + |B_{\mathcal{H}_\infty}(v,r) \cap \boldsymbol{H}^{n+1}| && \\
&\qquad\qquad \lesssim r^{\log_3(12)} && r \leqslant 3^{n-1} \\
&|B_{\mathcal{H}_\infty}(v,r)| \leqslant \sum_{j \leqslant \max(k,n+1)} |\boldsymbol{H}^j| \leqslant 2 \cdot 12^{\max(k,n+1)} \lesssim r^{\log_3(12)} && r \in [3^{k-1}, 3^k), k \geqslant n.
\end{aligned}$$

These four bounds together verify the desired claim. □

2.2.3 Effective resistances

Consider a weighted, undirected graph $G = (V, E, c)$ with edge conductances $c : E \to \mathbb{R}_+$. For $p \geqslant 1$, denote $\ell_p(V) := \{f : V \to \mathbb{R} \mid \sum_{u\in V} |f(u)|^p < \infty\}$, and equip $\ell_2(V)$ with the inner product $\langle f, g\rangle = \sum_{u\in V} f(u)g(u)$.

For $s, t \in \ell_1(V)$ with $\|s\|_1 = \|t\|_1$, we define the *effective resistance*

$$R^G_{\mathrm{eff}}(s,t) := \left\langle s - t, L^\dagger_G(s-t)\right\rangle,$$

where L_G is the combinatorial Laplacian of G, and $L^\dagger_G$ is the Moore-Penrose pseudoinverse. Here, L_G is the operator on $\ell_2(V)$ defined by

$$L_G f(v) = \sum_{u:\{u,v\}\in E} c(\{u,v\})\,(f(v) - f(u))\,.$$

If G is unweighted, we assume it is equipped with unit conductances $c \equiv \mathbb{1}_{E(G)}$.

Equivalently, if we consider mappings $\theta : E \to \mathbb{R}$, and define the energy functional

$$\mathcal{E}_G(\theta) := \sum_{e\in E} c(e)^{-1}\theta(e)^2,$$

then $R^G_{\mathrm{eff}}(s,t)$ is the minimum energy of a flow with demands $s - t$. (See, for instance, [LP16, Ch. 2].) For two finite sets $A, B \subseteq V$ in a graph, we define

$$R^G_{\mathrm{eff}}(A \leftrightarrow B) := \inf\left\{R^G_{\mathrm{eff}}(s,t) : \mathrm{supp}(s) \subseteq A, \mathrm{supp}(t) \subseteq B, s,t \in \ell_1(V), \|s\|_1 = \|t\|_1 = 1\right\},$$

and we recall the following standard characterization (see, e.g., [LP16, Thm. 2.3]).

If we define additionally $c_v := \sum_{u\in V(L)} c(\{u,v\})$ for $v \in V$, then we can recall that weighted random walk $\{X_t\}$ on G with Markovian law

$$\mathbb{P}\left[X_{t+1} = v \mid X_t = u\right] = \frac{c(\{u,v\})}{c_u}, \qquad u, v \in V.$$

Theorem 2.17 (Transience criterion)**.** *A weighted graph $G = (V, E, c)$ is transient if and only if there is a vertex $v \in V$ and an increasing sequence $V_1 \subseteq \cdots \subseteq V_n \subseteq V_{n+1} \subseteq \cdots$ of finite subsets of vertices satisfying $\bigcup_{n\geqslant 1} V_n = V$ and*

$$\sup_{n\geqslant 1} R_{\text{eff}}^G(\{v\} \leftrightarrow V \setminus V_n) < \infty.$$

For a tiling $\boldsymbol{T}$ of a closed rectangle R, let $L(\boldsymbol{T})$ and $R(\boldsymbol{T})$ denote the sets of tiles that intersect the left and right edges of R, respectively. We define

$$\rho(\boldsymbol{T}) := R_{\text{eff}}^{G(\boldsymbol{T})}\left(\mathbb{1}_{L(\boldsymbol{T})}/|L(\boldsymbol{T})|, \mathbb{1}_{R(\boldsymbol{T})}/|R(\boldsymbol{T})|\right),$$

Observation 2.18. For any $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{T}$, we have $|L(\boldsymbol{S}\circ\boldsymbol{T})| = |L(\boldsymbol{S})|\cdot|L(\boldsymbol{T})|$ and $|R(\boldsymbol{S}\circ\boldsymbol{T})| = |R(\boldsymbol{S})|\cdot|R(\boldsymbol{T})|$. In particular, $|L(\boldsymbol{H}^n)| = |R(\boldsymbol{H}^n)| = 3^n$.

Lemma 2.19. *Suppose that $\boldsymbol{S}, \boldsymbol{T}$ are tilings satisfying the conditions of Definition 2.8. Suppose furthermore that all rectangles in $R(\boldsymbol{S})$ have the same height, and the same is true for $L(\boldsymbol{T})$. Then we have*

$$\rho(\boldsymbol{S} \mid \boldsymbol{T}) \leqslant \rho(\boldsymbol{S}) + \rho(\boldsymbol{T}) + \frac{1}{\max(|R(\boldsymbol{S})|, |L(\boldsymbol{T})|)}.$$

Proof. By the triangle inequality for effective resistances, it suffices to prove that

$$R_{\text{eff}}^G\left(\mathbb{1}_{R(\boldsymbol{S})}/|R(\boldsymbol{S})|, \mathbb{1}_{L(\boldsymbol{T})}/|L(\boldsymbol{T})|\right) \leqslant \frac{1}{\max(|R(\boldsymbol{S})|, |L(\boldsymbol{T})|)},$$

where $G = G(\boldsymbol{S} \mid \boldsymbol{T})$. We construct a flow from $R(\boldsymbol{S})$ to $L(\boldsymbol{T})$ as follows: If $A \in R(\boldsymbol{S})$, $B \in L(\boldsymbol{T})$ and $\{A, B\} \in E(G)$, then the flow value on $\{A, B\}$ is

$$F_{AB} := \frac{\text{len}(A \cap B)}{\ell_2(A)} \frac{1}{|R(\boldsymbol{S})|}.$$

Denoting $m := \max(|R(\boldsymbol{S})|, |L(\boldsymbol{T})|)$, we clearly have $F_{AB} \leqslant 1/m$. Moreover,

$$\sum_{A\in R(\boldsymbol{S})} \sum_{\substack{B\in L(\boldsymbol{T}):\\ \{A,B\}\in E(G)}} F_{AB} = 1,$$

hence

$$\sum_{A\in R(\boldsymbol{S})} \sum_{\substack{B\in L(\boldsymbol{T}):\\ \{A,B\}\in E(G)}} F_{AB}^2 \leqslant 1/m,$$

completing the proof. □

Using the simple inequalities $\rho(\boldsymbol{T}) \geqslant 1/|L(\boldsymbol{T})|$ and $\rho(\boldsymbol{T}) \geqslant 1/|R(\boldsymbol{T})|$, this yields the following.

Corollary 2.20. *For any tilings $\boldsymbol{S}, \boldsymbol{T}$ satisfying the assumptions of Lemma 2.19, it holds that*

$$\rho(\boldsymbol{S} \mid \boldsymbol{T}) \leqslant \rho(\boldsymbol{S}) + \rho(\boldsymbol{T}) + \min(\rho(\boldsymbol{S}), \rho(\boldsymbol{T})).$$

Lemma 2.21. *For every $n \geqslant 1$, it holds that*

$$\rho(\boldsymbol{H}^n) \lesssim (5/6)^n.$$

Proof. Fix $n \geqslant 2$. Recalling Figure 1(b), let us consider $\boldsymbol{H}^n$ as consisting of three (identical) tilings stacked vertically, and where each of these three tilings is written as $\boldsymbol{H}^{n-1} \mid \boldsymbol{S} \mid \boldsymbol{H}^{n-1}$ where $\boldsymbol{S}$ consists of two copies of $\boldsymbol{H}^{n-1}$ stacked vertically. Applying Lemma 2.19 to $\boldsymbol{H}^{n-1} \mid \boldsymbol{S} \mid \boldsymbol{H}^{n-1}$ gives

$$\begin{aligned}\rho(\boldsymbol{H}^n) &\leqslant (1/3)^2 \cdot 3\left(2\rho(\boldsymbol{H}^{n-1}) + \rho(\boldsymbol{S}) + \frac{1}{\max(|R(\boldsymbol{H}^{n-1})|, |L(\boldsymbol{S})|)} + \frac{1}{\max(|L(\boldsymbol{H}^{n-1})|, |R(\boldsymbol{S})|)}\right) \\ &\leqslant (1/3)^2 \cdot 3\left(2\rho(\boldsymbol{H}^{n-1}) + (1/2)^2 \cdot 2\rho(\boldsymbol{H}^{n-1}) + \frac{2}{2 \cdot 3^{n-1}}\right) \\ &= (5/6)\rho(\boldsymbol{H}^{n-1}) + 3^{-n},\end{aligned}$$

where in the second inequality we have employed Observation 2.18. This yields the desired result by induction on n. ◻

Corollary 2.22. *The graphs $\mathcal{H}_n = G(\boldsymbol{H}^0 \mid \boldsymbol{H}^1 \mid \cdots \mid \boldsymbol{H}^n)$ satisfy*

$$\sup_{n \geqslant 1} \rho(\mathcal{H}_n) < \infty. \tag{2.8}$$

Hence $\mathcal{H}_\infty$ is transient.

Proof. Employing Lemma 2.19, Observation 2.18, and Lemma 2.21 together yields

$$\rho(\mathcal{H}_n) \lesssim \sum_{j=1}^{n} \left[(5/6)^j + 3^{-j}\right],$$

verifying (2.8). Now Theorem 2.17 yields the transience of $\mathcal{H}_\infty$. ◻

2.3 Generalizations and unimodular constructions

Consider a sequence $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$ with $\gamma_i \in \mathcal{N}$. Define a tiling $\boldsymbol{T}_\gamma \in \mathcal{T}$ as follows: The unit square is partitioned into b columns of width $1/b$, and for $i \in \{1, 2, \ldots, b\}$, the ith column has γ_i rectangles of height $1/\gamma_i$. For instance, the tiling $\boldsymbol{H}$ from Figure 7(a) can be written $\boldsymbol{H} = \boldsymbol{T}_{\langle 3,6,3 \rangle}$.

We will assume throughout this subsection that $\min(\gamma) = b$ and $\gamma_1 = \gamma_b$. Let us use the notation $|\gamma| := \gamma_1 + \cdots + \gamma_b$. The proof of the next lemma follows just as for Lemma 2.14 using $\min(\gamma) = b$ so that there is a column in $\boldsymbol{T}_\gamma^n$ of height b^n.

Lemma 2.23. *For $n \geqslant 0$, it holds that $|\boldsymbol{T}_\gamma^n| = |\gamma|^n$, and $b^n \leqslant \operatorname{diam}(\boldsymbol{T}_\gamma^n) \leqslant 3b^n$.*

Note that $\eta_{\boldsymbol{T}_\gamma} = 1$ since $\gamma_1 = \gamma_b$. The next lemma follows from the same reasoning used in the proof of Lemma 2.15. The dependence of the implicit constant on $|\gamma|/b$ comes from $\alpha_{\boldsymbol{T}_\gamma} \leqslant |\gamma|/b$.

Lemma 2.24. *For any $n \geqslant 0$, it holds that*

$$|B_G(A, r)| \asymp_{|\gamma|/b} r^{\log_b(|\gamma|)} \qquad \forall A \in \boldsymbol{T}_\gamma^n,\ 1 \leqslant r \leqslant \operatorname{diam}(G(\boldsymbol{T}_\gamma^n)).$$

2.3.1 Degrees of growth

Consider $b, k \in \mathcal{N}$ with $k \geqslant b \geqslant 3$, and define the sequence

$$\gamma^{(b,k)} := \left\langle b, \underbrace{\left\lceil \frac{k-3}{b-3} \right\rceil b, \ldots, \left\lceil \frac{k-3}{b-3} \right\rceil b}_{(k-3) \bmod (b-3) \text{ copies}}, b, \underbrace{\left\lfloor \frac{k-3}{b-3} \right\rfloor b, \ldots, \left\lfloor \frac{k-3}{b-3} \right\rfloor b}_{[(b-3)-(k-3)] \bmod (b-3) \text{ copies}}, b, b \right\rangle.$$

Denote $\boldsymbol{T}_{(b,k)} := \boldsymbol{T}_{\gamma^{(b,k)}}$ and note that $|\gamma^{(b,k)}| = bk$. Define $d_g(b,k) := \log_b(bk)$, and $\Gamma_{b,k} := \sum_{i=1}^{b} 1/\gamma_i^{(b,k)}$.

Observation 2.25. The following facts hold for $k \geqslant b \geqslant 3$ and $n \geqslant 0$:

(a) There are b^n tiles in the left- and right-most columns of $\boldsymbol{T}_{(b,k)}^n$.

(b) If a pair of consecutive columns in $\boldsymbol{T}_{(b,k)}^n$ have heights h and h', then $\min(h,h')$ divides $\max(h,h')$.

Now observe that Lemma 2.24 yields the following.

Corollary 2.26. *The family of graphs $\mathcal{F} = \{G(\boldsymbol{T}_{(b,k)}^n) : n \geqslant 0\}$ has uniform polynomial growth of degree $d_g(b,k)$ in the sense that*

$$|B_G(x,r)| \asymp_k r^{d_g(b,k)}, \qquad \forall G \in \mathcal{F}, x \in V(G), 1 \leqslant r \leqslant \operatorname{diam}(G).$$

For any rational $p/q \geqslant 2$, one can achieve $d_g(b,k) = p/q$ by taking $b = 3^q$ and $k = 3^{p-q}$.

Next, we analyze the effective resistances across $\boldsymbol{T}_{(b,k)}^n$.

Lemma 2.27. *For every $n \geqslant 1$, it holds that*

$$\rho\left(\boldsymbol{T}_{(b,k)}^n\right) \lesssim \Gamma_{b,k}^n.$$

Proof. Fix $n \geqslant 2$ and write $\boldsymbol{T}_{(b,k)}^n = \boldsymbol{T}_{(b,k)} \circ \boldsymbol{T}_{(b,k)}^{n-1}$ as $\boldsymbol{A}_1 \mid \boldsymbol{A}_2 \mid \cdots \mid \boldsymbol{A}_b$ where, for $1 \leqslant i \leqslant b$, each $\boldsymbol{A}_i$ is a vertical stack of $\gamma_i^{(b,k)}$ copies of $\boldsymbol{T}_{(b,k)}^{n-1}$. Since $\rho(\boldsymbol{A}_i) = \rho(\boldsymbol{T}_{(b,k)}^{n-1})/\gamma_i^{(b,k)}$ by the parallel law for effective resistances, applying Lemma 2.19 to $\boldsymbol{A}_1 \mid \boldsymbol{A}_2 \mid \cdots \mid \boldsymbol{A}_b$ gives

$$\rho\left(\boldsymbol{T}_{(b,k)}^n\right) \leqslant \sum_{i-1}^{b} \rho(\boldsymbol{T}_{(b,k)}^{n-1})/\gamma_i^{(b,k)} + \sum_{i=1}^{b-1} \frac{1}{\min\left(R(\boldsymbol{A}_i), L(\boldsymbol{A}_{i+1})\right)} \leqslant \rho(\boldsymbol{T}_{(b,k)}^{n-1})\Gamma_{b,k} + b^{1-n},$$

where in the second inequality we have employed $\min\left(R(\boldsymbol{A}_i), L(\boldsymbol{A}_{i+1})\right) \geqslant b^n$ which follows from Observation 2.25(a). Finally, observe that $\Gamma_{b,k} \geqslant 1/\gamma_1^{(b,k)} + 1/\gamma_b^{(b,k)} = 2/b$, and therefore the desired claim follows by induction. □

The next result establishes Theorem 2.1.

Theorem 2.28. *For every $k > b$, the graphs $\mathcal{T}_n^{(b,k)} := G\left(\boldsymbol{T}_{(b,k)}^0 \mid \boldsymbol{T}_{(b,k)}^1 \mid \cdots \mid \boldsymbol{T}_{(b,k)}^n\right)$ satisfy*

$$\sup_{n \geqslant 1} \rho\left(\mathcal{T}_n^{(b,k)}\right) < \infty. \tag{2.9}$$

Hence the limit graph $\mathcal{T}_\infty^{(b,k)}$ is transient. Moreover, $\mathcal{T}_\infty^{(b,k)}$ has uniform polynomial growth of degree $d_g(b,k)$.

Proof. Employing Lemma 2.19, Observation 2.25(a), and Lemma 2.27 together yields

$$\rho\left(\mathcal{T}_n^{(b,k)}\right) \lesssim \sum_{j=1}^{n} \left(\Gamma_{b,k}^j + b^{-j}\right).$$

For $k > b$, we have $\max(\gamma^{(b,k)}) > b$ and $\min(\gamma^{(b,k)}) = b$, hence $\Gamma_{b,k} < 1$, verifying (2.9). Now Theorem 2.17 yields transience of $\mathcal{T}_\infty^{(b,k)}$.

Uniform polynomial growth of degree $d_g(b,k)$ follows from Corollary 2.26 as in the proof of Lemma 2.16. □

2.3.2 The distributional limit

Fix $k \geqslant b \geqslant 3$ and take $G_n := G(\boldsymbol{T}_{(b,k)}^n)$. Since the degrees in $\{G_n\}$ are uniformly bounded, the sequence has a subsequential distributional limit, and in all arguments that follow, we could consider any such limit. But let us now argue that if μ_n is the law of (G_n, ρ_n) with $\rho_n \in V(G_n)$ chosen according to the stationary measure, then the measures $\{\mu_n : n \geqslant 0\}$ have a distributional limit.

Lemma 2.29. *For any $k \geqslant b \geqslant 3$, there is a reversible random graph $(G_{b,k}, \rho)$ such that $\{(G_n, \rho_n)\} \Rightarrow (G_{b,k}, \rho)$. Moreover, almost surely $G_{b,k}$ has uniform polynomial volume growth of degree $d_g(b,k)$.*

Proof. It suffices to prove that $\{(G_n, \rho_n)\}$ has a limit $(G_{b,k}, \rho)$. Reversibility of the limit then follows automatically (as noted in Section 2.1.1), and the degree of growth is an immediate consequence of Corollary 2.26. It will be slightly easier to show that the sequence $\{(G_n, \hat{\rho}_n)\}$ has a distributional limit, with $\hat{\rho}_n \in V(G_n)$ chosen uniformly at random.

Let $\mu_{n,r}$ be the law of $B_{G_n}(\hat{\rho}_n, r)$. It suffices to show that the measures $\{\mu_{n,r} : n \geqslant 0\}$ converge for every fixed $r \geqslant 1$, and then a standard application of Kolmogorov's extension theorem proves the existence of a limit.

For a tiling $\boldsymbol{T}$ of a rectangle R, let $\partial \boldsymbol{T}$ denote the set of tiles that intersect some side of R. Define the neighborhood $N_r(\partial \boldsymbol{T}_{(b,k)}^n) := \{v \in \boldsymbol{T}_{(b,k)}^n : d_{G_n}(v, \partial \boldsymbol{T}_{(b,k)}^n) \leqslant r\}$ and abbreviate $d = d_g(b,k)$. Then $|\partial \boldsymbol{T}_{(b,k)}^n| \leqslant 4b^n$, so Corollary 2.26 gives

$$\left|N_r(\partial \boldsymbol{T}_{(b,k)}^n)\right| \lesssim_k b^n r^d.$$

Since $|\boldsymbol{T}_{(b,k)}^n| = (bk)^n$, it follows that

$$1 - \mathbb{P}\left[\mathcal{E}_{r,n}\right] \lesssim_k k^{-n} r^d,$$

where $\mathcal{E}_{r,n}$ is the event $\{B_{G_n}(\hat{\rho}_n, r) \cap \partial \boldsymbol{T}^n_{(b,k)} = \emptyset\}$.

Now write $\boldsymbol{T}^n_{(b,k)} = \boldsymbol{T}_{(b,k)} \circ \boldsymbol{T}^{n-1}_{(b,k)}$, and note that $\hat{\rho}_n$ falls into one of the $|\gamma^{(b,k)}| = bk$ copies of G_{n-1} and is, moreover, uniformly distributed in that copy. Therefore we can naturally couple $(G_n, \hat{\rho}_n)$ and $(G_{n-1}, \hat{\rho}_{n-1})$ by identifying $\hat{\rho}_n$ with $\hat{\rho}_{n-1}$. Moreover, conditioned on the event $\mathcal{E}_{r,n-1}$, we can similarly couple $B_{G_n}(\hat{\rho}_n, r)$ and $B_{G_{n-1}}(\hat{\rho}_{n-1}, r)$.

It follows that, for every $r \geqslant 1$,

$$d_{TV}(\mu_{n-1,r}, \mu_{n,r}) \leqslant 1 - \mathbb{P}[\mathcal{E}_{r,n-1}] \lesssim_k k^{-n} r^d .$$

As the latter sequence is summable, it follows that $\{\mu_{n,r}\}$ converges for every fixed $r \geqslant 1$, completing the proof. □

2.3.3 Speed of the random walk

Let $\{X_t\}$ denote the random walk on $G_{b,k}$ with $X_0 = \rho$. Our first goal will be to prove a lower bound on the speed of the walk. Define:

$$d_w(b,k) := d_g(b,k) + \log_b(\Gamma_{b,k}).$$

We will show that $d_w(b,k)$ is related to the speed exponent for the random walk.

Theorem 2.30. *Consider any $k \geqslant b \geqslant 3$. It holds that for all $T \geqslant 1$,*

$$\mathbb{E}\left[d_{G_{b,k}}(X_T, X_0) \mid X_0 = \rho\right] \gtrsim_k T^{1/d_w(b,k)}. \tag{2.10}$$

Before proving the theorem, let us observe that it yields Theorem 2.3. Fix $k \geqslant b \geqslant 3$. Observe that for any positive integer $p \geqslant 1$, we have $d_g(b^p, k^p) = d_g(b,k)$. On the other hand,

$$\begin{aligned} d_g(b^p, k^p) - d_w(b^p, k^p) &= -\log_{b^p}(\Gamma_{b^p,k^p}) \\ &\geqslant -\log_{b^p}\left(3b^{-p} + (b/k)^p\right) - o_p(1) \\ &\geqslant \min\left(1, \log_b(k) - 1\right) - o_p(1) \\ &\geqslant \min\left(1, d_g(b^p, k^p) - 2\right) - o_p(1)\,. \end{aligned} \tag{2.11}$$

So for every $\varepsilon > 0$, there is some $p = p(\varepsilon)$ such that

$$d_w(b^p, k^p) \leqslant \max\left(2, d_g(b^p, k^p) - 1\right) + \varepsilon\,,$$

and moreover G_{b^p,k^p} almost surely has uniform polynomial growth of degree $d_g(b,k)$. Combining this with the construction of Corollary 2.26 for all rational $d \geqslant 2$ yields Theorem 2.3.

Hitting times vs. rate of escape

For a (possibly weighted) graph G, a vertex $u \in V(G)$, and the random walk $\{X^G_t\}$ on G, we use the notation:

$$H_G(u,j) := \mathbb{E}\left[\min\left\{t \geqslant 0 : d_G(X^G_0, X^G_t) \geqslant j\right\} \mid X^G_0 = u\right].$$

Consider now a reversible random graph (G,ρ) and let $\{X_t\}$ denote the standard random walk on G. The next lemma is an extension of the arguments of [LP13, §4.1] from transitive graphs to reversible random graphs. It shows that in order to lower bound the speed, it suffices to give upper bounds on hitting times.

Lemma 2.31. *Suppose (G, ρ) is a reversible random graph and $\mathbb{E}[H_G(\rho, j)] \leqslant T$ for some $T \geqslant 1$. Then:*

$$\mathbb{E}\left[d_G(X_0, X_T) \mid X_0 = \rho\right] \geqslant \frac{j-1}{4}.$$

Proof. Let us assume throughout that $X_0 = \rho$. First, we claim that for every $T \geqslant 0$,

$$\mathbb{E}\left[d_G(X_0, X_T)\right] \geqslant \frac{1}{2} \max_{0 \leqslant t \leqslant T} \mathbb{E}\left[d_G(X_0, X_t) - d_G(X_0, X_1)\right]. \tag{2.12}$$

Let $s' \leqslant T$ be such that

$$\mathbb{E}[d_G(X_0, X_{s'})] = \max_{0 \leqslant t \leqslant T} \mathbb{E}[d_G(X_0, X_t)].$$

Then there exists an even time $s \in \{s', s'-1\}$ such that $\mathbb{E}[d_G(X_0, X_s)] \geqslant \mathbb{E}\left[d_G(X_0, X_{s'}) - d_G(X_0, X_1)\right]$.

Consider $\{X_t\}$ and an identically distributed walk $\{\tilde{X}_t\}$ such that $\tilde{X}_t = X_t$ for $t \leqslant s/2$ and $\tilde{X}_t$ evolves independently after time $s/2$. By the triangle inequality, we have

$$d_G(X_0, \tilde{X}_T) + d_G(\tilde{X}_T, X_s) \geqslant d_G(X_0, X_s).$$

But by reversibility of (G, ρ), both of the triples $(G, X_0, \tilde{X}_T)$ and $(G, \tilde{X}_T, X_s)$ have the same law as (G, X_0, X_T). Taking expectations yields (2.12).

Define $\alpha := \frac{1}{j} \max_{0 \leqslant t \leqslant 2T} \mathbb{E}[d_G(X_0, X_t)]$ and $H_j := \min\{t \geqslant 0 : d_G(X_0, X_t) \geqslant j\}$. First, observe that the triangle inequality implies

$$d_G(X_0, X_{2T}) \geqslant \mathbb{1}_{\{H_j \leqslant 2T\}} \left(j - d_G(X_{H_j}, X_{2T})\right). \tag{2.13}$$

By reversibility of (G, ρ), we also have

$$\mathbb{E}\left[\mathbb{1}_{\{H_j \leqslant 2T\}} \cdot d_G(X_{H_j}, X_{2T})\right] = \mathbb{E}\left[\mathbb{1}_{\{H_j \leqslant 2T\}} \cdot d_G(X_0, X_{2T-H_j})\right] \leqslant \mathbb{P}(H_j \leqslant 2T)\alpha j.$$

Thus taking expectations in (2.13) yields

$$\mathbb{E}[d_G(X_0, X_{2T})] \geqslant \mathbb{P}(H_j \leqslant 2T)(1-\alpha)j \geqslant \frac{1}{2}(1-\alpha)j,$$

recalling our assumption that $\mathbb{E}[H_j] \leqslant T$. On the other hand, (2.12) shows that

$$\mathbb{E}\ d_G(X_0, X_{2T}) \geqslant \frac{1}{2}(\alpha j - 1)$$

Averaging these two inequalities yields the desired result. □

The linearized graphs

Fix integers $k \geqslant b \geqslant 3$ and $n \geqslant 1$, and let us consider now the (weighted) graph $L = L^n_{(b,k)}$ derived from $G = G(\boldsymbol{T}^n_{(b,k)})$ by identifying every column of rectangles into a single vertex. Thus $|V(L)| = b^n$.

We connect two vertices $u, v \in V(L)$ if their corresponding columns C_u and C_v in G are adjacent, and we define the conductances $c(\{u, v\}) := |E_G(C_u, C_v)|$, where $E_G(S, T)$ denotes the number of edges between two subsets $S, T \subseteq V(G)$.

Let us order the vertices of L from left to right as $V(L) = \{\ell_1, \ldots, \ell_{b^n}\}$. The series law for effective resistances gives the following.

Observation 2.32. For $1 \leqslant i \leqslant t \leqslant j \leqslant b^n$, we have

$$R^L_{\text{eff}}(\ell_i \leftrightarrow \ell_j) = R^L_{\text{eff}}(\ell_i \leftrightarrow \ell_t) + R^L_{\text{eff}}(\ell_t \leftrightarrow \ell_j)$$

We will use this to bound the resistance between any pair of columns.

Lemma 2.33. *If* $1 \leqslant s < t \leqslant b^n$, *then*

$$R^L_{\text{eff}}(\ell_s \leftrightarrow \ell_t) \cdot (|C_s| + \cdots + |C_t|) \lesssim_k (\Gamma_{b,k} \cdot bk)^{\log_b(t-s)} . \tag{2.14}$$

Proof. Denote $h := \lceil \log_b(t-s) \rceil$, $\boldsymbol{T} := \boldsymbol{T}_{(b,k)}$, and $\Gamma := \Gamma_{b,k}$. Write $\boldsymbol{T}^n = \boldsymbol{T}^{n-h} \circ \boldsymbol{T}^h$ and along this decomposition, partition $\boldsymbol{T}^n$ into b^{n-h} sets of tiles $\mathcal{D}_1, \ldots, \mathcal{D}_{b^{n-h}}$, where each $\mathcal{D}_i$ is formed from adjacent columns

$$\mathcal{D}_i := C_{(i-1)\cdot b^h+1} \cup \cdots \cup C_{i \cdot b^h} .$$

Suppose that, for $1 \leqslant i \leqslant b^{n-h}$, the tiling $\boldsymbol{T}^{n-h}$ has β_i tiles in its ith column. Then $\mathcal{D}_i$ consists of β_i copies of $\boldsymbol{T}^h$ stacked atop each other.

Thus we have $|\mathcal{D}_i| = \beta_i |\boldsymbol{T}^h|$, and furthermore $\rho(\mathcal{D}_i) \leqslant \rho(\boldsymbol{T}^h)/\beta_i$, hence

$$\rho(\mathcal{D}_i) \cdot |\mathcal{D}_i| \leqslant \rho(\boldsymbol{T}^h) \cdot |\boldsymbol{T}^h| \lesssim \Gamma^h \cdot (bk)^h , \tag{2.15}$$

where the last inequality uses Lemma 2.27.

Let $1 \leqslant i \leqslant j \leqslant b^{n-h}$ be such that $C_s \subseteq \mathcal{D}_i$ and $C_t \subseteq \mathcal{D}_j$. Since $t \leqslant s + b^h$, and each set $\mathcal{D}_i$ consists of b^h consecutive columns, it must be that $j \leqslant i+1$. If $i = j$, then $|C_s| + \cdots + |C_t| \leqslant |\mathcal{D}_i|$, and Observation 2.32 gives

$$R^L_{\text{eff}}(\ell_s \leftrightarrow \ell_t) \leqslant \rho(\mathcal{D}_i),$$

thus (2.15) yields (2.14), as desired.

Suppose, instead, that $j = i + 1$. From Lemma 2.13, we have $\alpha_{\boldsymbol{T}^{n-h}} \leqslant \alpha_{\boldsymbol{T}} \leqslant |\gamma|/b \leqslant k$. Therefore $1/k \leqslant \beta_{i+1}/\beta_i \leqslant k$. Thus using Corollary 2.20 and (2.15) gives

$$\rho(\mathcal{D}_i \cup \mathcal{D}_{i+1}) = \rho(\mathcal{D}_i \mid \mathcal{D}_{i+1}) \leqslant 2\rho(\mathcal{D}_i) + \rho(\mathcal{D}_{i+1}) \leqslant 3k\rho(\boldsymbol{T}^h)/\beta_i .$$

Since it also holds that $|\mathcal{D}_i| + |\mathcal{D}_{i+1}| = (\beta_i + \beta_{i+1})|\boldsymbol{T}^h| \leqslant 2k\beta_i|\boldsymbol{T}^h|$, Observation 2.32 gives

$$R^L_{\text{eff}}(\ell_s \leftrightarrow \ell_t) \cdot (|C_s| + \cdots + |C_t|) \leqslant \rho(\mathcal{D}_i \cup \mathcal{D}_{i+1})|\mathcal{D}_i \cup \mathcal{D}_{i+1}| \lesssim_k \rho(\boldsymbol{T}^h)|\boldsymbol{T}^h| ,$$

and again (2.15) establishes (2.14), completing the proof. □

Lemma 2.34. *For any* $1 \leqslant j \leqslant b^n/2$ *and* $1 \leqslant i \leqslant b^n$, *we have*

$$H_L(\ell_i, j) \lesssim_k j^{d_w(b,k)}$$

Proof. Denote $s := \max(1, i-j)$ and $t := \min(b^n, i+j)$. Then the standard connection between hitting times and effective resistances [?] in conjuction with Lemma 2.33 yields

$$\begin{aligned} H_L(\ell_i, j) &\leqslant 2 \left(\sum_{i=s}^{t} \sum_{v \in C_i} c_v \right) \max \left(R^L_{\text{eff}}(\ell_s \leftrightarrow \ell_i), R^L_{\text{eff}}(\ell_i \leftrightarrow \ell_t) \right) \\ &\lesssim_k (|C_s| + \cdots + |C_t|) \max \left(R^L_{\text{eff}}(\ell_s \leftrightarrow \ell_i), R^L_{\text{eff}}(\ell_i \leftrightarrow \ell_t) \right) \\ &\lesssim_k (bk \cdot \Gamma_{b,k})^{\log_b(j)} , \end{aligned}$$

where in the second line we have used that the degrees in $G(\boldsymbol{T}^n)$ are bounded by a function of k (cf. Lemma 3.14). □

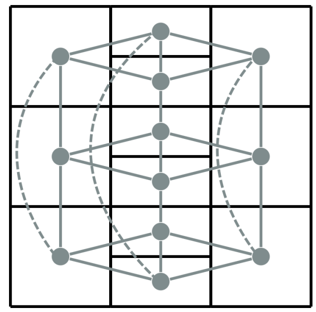


Figure 5: The cylindrical graph $\tilde{G}$ for $G = G(\boldsymbol{T}_{\langle 3,6,3\rangle})$. The new edges are dashed.

Hitting times in $G_{b,k}$

Consider now the graphs $G_n := G(\boldsymbol{T}^n_{(b,k)})$ for some $k \geqslant b \geqslant 3$ and $n \geqslant 1$. Let us define the cylindrical version $\tilde{G}_n$ of G_n with the same vertex set, but additionally and edge from the top tile to the bottom tile in every column (see Figure 5). If we choose $\tilde{\rho}_n \in V(\tilde{G}_n)$ according to the stationary measure on $\tilde{G}_n$, then clearly $\{(\tilde{G}_n, \tilde{\rho}_n)\} \Rightarrow (G_{b,k}, \rho)$ as well.

Define also $L_n := L^n_{b,k}$. Because of Observation 2.25(b), the graph $\tilde{G}_n$ has vertical symmetry: The automorphism group of $\tilde{G}_n$ acts transitively within columns. Let $\pi_n : V(G_n) \to V(L_n)$ denote the projection map and observe that

$$d_{\tilde{G}_n}(u,v) \geqslant d_{L_n}(\pi_n(u), \pi_n(v)), \qquad \forall u,v \in V(G_n),$$

therefore

$$H_{\tilde{G}_n}(u,j) \lesssim H_{L_n}(\pi_n(u), j), \qquad \forall u \in V(G_n), \tag{2.16}$$

This follows from the "vertical symmetry" mentioned above: The random walk on $\tilde{G}_n$ projects to the random walk on L_n. The loss of an implicit constant in (2.16) is due to the fact that the projected walk $\pi_n(X_t^{\tilde{G}_n})$ has holding probabilities, but these are bounded above by 2/3 (as every vertex in $\tilde{G}_n$ has two vertical neighbors, and at least one horizontal neighbor).

From (2.16) and Lemma 2.34, we conclude that

$$H_{\tilde{G}_n}(u,j) \lesssim_k j^{d_w(b,k)}, \qquad \forall u \in V(G), j \leqslant b^n/2. \tag{2.17}$$

With this in hand, we can establish speed lower bounds in the limit $(G_{b,k}, \rho)$.

Proof of Theorem 2.30. Observe that (2.17) implies that $\mathbb{E}[H_{G_{b,k}}(\rho, j)] \lesssim_k j^{d_w(b,k)}$ for every $j \geqslant 1$. Now Lemma 2.31 yields the desired bound (2.10). ◻

2.3.4 Annular resistances

We will establish Theorem 2.5 by proving the following.

Theorem 2.35. *For any $k \geqslant b \geqslant 3$, there is a constant $C = C(k)$ such that for $G = G_{b,k}$, almost surely*

$$R^G_{\text{eff}}(B_G(\rho, R) \leftrightarrow V(G) \setminus B_G(\rho, 2R)) \leqslant C R^{\log_b(\Gamma_{b,k})}, \quad \forall R \geqslant 1.$$

To see that this yields Theorem 2.5, consider some $k \geqslant b^2$, corresponding to the restriction $d_g(b,k) \geqslant 3$. Then for all positive integers $p \geqslant 1$, we have $d_g(b^p, k^p) = d_g(b,k)$ and recalling (2.11),

$$\lim_{p\to\infty} \log_b(\Gamma_{b^p,k^p}) = -1\,.$$

To prove Theorem 2.35, it suffices to show the following.

Lemma 2.36. *For every $n \geqslant 1$, $k \geqslant b \geqslant 3$, there is a constant $C = C(k)$ such that for $G = G(\boldsymbol{T}^n_{(b,k)})$, we have*

$$R^G_{\mathrm{eff}}(B_G(x,R) \leftrightarrow V(G) \setminus B_G(x,2R)) \leqslant C R^{\log_b(\Gamma_{b,k})}, \quad \forall x \in V(G), 1 \leqslant R \leqslant \mathrm{diam}(G)/C\,.$$

Proof. Denote $\boldsymbol{T} := \boldsymbol{T}_{(b,k)}$. Consider some value $1 \leqslant R \leqslant \mathrm{diam}(G)/C$, and define $h := \lfloor \log_b(R/3) \rfloor$.

Let $C_1, \ldots, C_{b^n}$ denote the columns of $\boldsymbol{T}^n$ and writing $\boldsymbol{T}^n = \boldsymbol{T}^{n-h} \circ \boldsymbol{T}^h$, let us partition the columns into consecutive sets $\mathcal{D}_1, \ldots, \mathcal{D}_{b^{n-h}}$ (as in the proof of Lemma 2.33), where $\mathcal{D}_i = C_{(i-1)b^h+1} \cup \cdots \cup C_{ib^h}$. For $1 \leqslant i \leqslant b^{n-h}$, let β_i denote the number of tiles in the ith column of $\boldsymbol{T}^{n-h}$ so that $\mathcal{D}_i$ consists of β_i copies of $\boldsymbol{T}^h$ stacked vertically.

Fix some vertex $x \in V(G)$ and suppose that $x \in \mathcal{D}_s$ for some $1 \leqslant s \leqslant b^{n-h}$. Denote $\Delta := 9b$. By choosing C sufficiently large, we can assume that $b^{n-h} > \Delta$, so that either $s \leqslant b^{n-h} - \Delta$ or $s \geqslant 1 + \Delta$. Let us assume that $s \leqslant b^{n-h} - \Delta$, as the other case is treated symmetrically. Define $t := \lceil s + 2 + 6b \rceil$ so that $t - s \leqslant \Delta$, and

$$d_G(\mathcal{D}_s, \mathcal{D}_t) \geqslant b^h(t-s-1) \geqslant (t-s-1)\frac{R}{3b} > 2R\,. \tag{2.18}$$

Denote $\xi := \gcd(\beta_s, \beta_{s+1}, \ldots, \beta_t)$. We claim that

$$\xi \gtrsim_k \max(\beta_s, \beta_{s+1}, \ldots, \beta_t)\,. \tag{2.19}$$

This follows because $\min(\beta_i, \beta_{i+1}) \mid \max(\beta_i, \beta_{i+1})$ for all $1 \leqslant i < b^n$ (cf. Observation 2.25(b)), and moreover the ratio $\max(\beta_i, \beta_{i+1})/\min(\beta_i, \beta_{i+1})$ is bounded by a function depending only on k. Since $t - s \lesssim_k 1$, this verifies (2.19).

Denote $\hat{\mathcal{D}} := \mathcal{D}_s \cup \cdots \cup \mathcal{D}_t$. One can verify that $\hat{\mathcal{D}}$ is a vertical stacking of ξ copies of $\hat{\boldsymbol{T}} := \boldsymbol{T}_{\langle \beta_s/\xi, \ldots, \beta_t/\xi \rangle} \circ \boldsymbol{T}^h$, and Corollary 2.20 implies that

$$\rho(\hat{\boldsymbol{T}}) \lesssim_k \rho(\boldsymbol{T}^h) \lesssim \Gamma^h_{b,k}\,, \tag{2.20}$$

with the final inequality being the content of Lemma 2.27.

Let $\hat{\boldsymbol{A}}$ be the copy of $\hat{\boldsymbol{T}}$ that contains x, and let $\boldsymbol{S}$ be the copy of $\boldsymbol{T}^h$ in $\boldsymbol{T}^n = \boldsymbol{T}^{n-h} \circ \boldsymbol{T}^h$ that contains x. Since ξ divides β_s, it holds that $\boldsymbol{S} \subseteq \hat{\boldsymbol{A}}$ and $L(\boldsymbol{S}) \subseteq L(\hat{\boldsymbol{A}})$. We further have

$$|L(\hat{\boldsymbol{A}})| = |L(\hat{\boldsymbol{T}})| = (\beta_s/\xi)|L(\boldsymbol{T}^h)| = (\beta_s/\xi)|L(\boldsymbol{S})| \lesssim_k |L(\boldsymbol{S})|\,.$$

This yields

$$R^G_{\mathrm{eff}}(\mathbb{1}_{L(\boldsymbol{S})}/|L(\boldsymbol{S})| \leftrightarrow R(\hat{\boldsymbol{A}})) \lesssim_k R^G_{\mathrm{eff}}(\mathbb{1}_{L(\hat{\boldsymbol{A}})}/|L(\hat{\boldsymbol{A}})| \leftrightarrow R(\hat{\boldsymbol{A}})), \tag{2.21}$$

where we have used the hybrid notation: For $s \in \ell_1(V)$,

$$R^G_{\mathrm{eff}}(s \leftrightarrow U) := \inf\left\{R^G_{\mathrm{eff}}(s,t) : \mathrm{supp}(t) \subseteq U, \|t\|_1 = \|s\|_1\right\}.$$

Therefore,

$$R_{\text{eff}}^{G}(L(\boldsymbol{S}) \leftrightarrow R(\hat{\boldsymbol{A}})) \leqslant R_{\text{eff}}^{G}(\mathbb{1}_{L(\boldsymbol{S})}/|L(\boldsymbol{S})| \leftrightarrow R(\hat{\boldsymbol{A}}))$$
$$\overset{(2.21)}{\lesssim_k} R_{\text{eff}}^{G}(\mathbb{1}_{L(\hat{\boldsymbol{A}})}/|L(\hat{\boldsymbol{A}})| \leftrightarrow R(\hat{\boldsymbol{A}})) = \rho(\hat{\boldsymbol{T}}) \overset{(2.20)}{\lesssim_k} \Gamma_{b,k}^{h}\,.$$

Since $\operatorname{diam}_G(\boldsymbol{S}) \leqslant 3b^h \leqslant R$ and $x \in \boldsymbol{S}$, it holds that $\boldsymbol{S} \subseteq B_G(x, R)$. On the other hand, since $x \in \boldsymbol{S}$, (2.18) shows that $B_G(x, 2R) \cap R(\hat{\boldsymbol{A}}) = \emptyset$. We conclude that

$$R_{\text{eff}}^{G}(B_G(x,R) \leftrightarrow V(G) \setminus B_G(x,2R)) \leqslant R_{\text{eff}}^{G}(L(\boldsymbol{S}) \leftrightarrow R(\hat{\boldsymbol{A}})) \lesssim_k \Gamma_{b,k}^{h} \lesssim_k R^{\log_b(\Gamma_{b,k})},$$

as desired. □

2.3.5 Complements of balls are connected

Let us finally prove Theorem 2.4. Recall the setup from Section 2.3.3: $G_n = G(\boldsymbol{T}_{(b,k)}^{n})$, $\tilde{G}_n$ denotes the cylindrical version, and $\{(\tilde{G}_n, \tilde{\rho}_n)\} \Rightarrow (G_{b,k}, \rho)$.

Proof of Theorem 2.4. We will show that, for every $R \geqslant 1$, almost surely the complement of a ball $B_G(\rho, R)$ in $G_{b,k}$ is connected. In conjuction with Lemma 2.29 and Corollary 2.26, this establishes Theorem 2.4.

Note that for any $x \in V(\tilde{G}_n)$ and $R \leqslant (b^n)/3$, it holds that the complement of $B_{\tilde{G}_n}(x, R)$ is connected in $\tilde{G}_n$, as $B_{\tilde{G}}(x, R)$ cannot "wrap around" the cylinder. Moreover, if we define $U := V(\tilde{G}_n) \setminus B_{\tilde{G}_n}(x, R)$, then for any $x, y \in U$, it holds that for some constant $\kappa = \kappa(k) > 1$,

$$d_{\tilde{G}_n[U]}(x,y) \leqslant \kappa \left(d_{\tilde{G}_n}(x,y) + R\right)\,.$$

This implies that for all $x \in V(G)$, $R \leqslant (b^n)/3$ and $R' > R$, it holds that $B_{\tilde{G}_n}(x, R') \setminus B_{\tilde{G}_n}(x, R)$ is connected in $\tilde{G}_n[B_{\tilde{G}_n}(\rho, 3\kappa R') \setminus B_{\tilde{G}_n}(\rho, R)]$.

Therefore almost surely, for all $R' > R \geqslant 1$, it holds that $B_{G_{b,k}}(\rho, R') \setminus B_{G_{b,k}}(\rho, R)$ is connected in $G_{b,k}[B_{G_{b,k}}(\rho, 3\kappa R') \setminus B_{G_{b,k}}(\rho, R)]$. In particular, this implies that the complement of $B_{G_{b,k}}(\rho, R)$ is connected in $G_{b,k}$, completing the proof. □

3 Non-existence of annular separators in geometric graphs

3.1 Introduction

The well-known Lipton-Tarjan separator theorem [LT79] asserts that any n-vertex planar graph has a balanced separator with $O(\sqrt{n})$ vertices. By the Koebe-Andreev-Thurston circle packing theorem, every planar graph can be realized as the tangency graph of interior-disjoint circles in the plane. One can define d-dimensional geometric graphs by analogy: Take a collection of "almost non-overlapping" bodies $\{S_v \subseteq \mathbb{R}^d : v \in V\}$, where each S_v is "almost round," and the associated geometric graph contains an edge $\{u, v\}$ if S_u and S_v "almost touch."

As a prototypical example, suppose we require that every point $x \in \mathbb{R}^d$ is contained in at most k of the bodies $\{S_v\}$, each S_v is a Euclidean ball, and two bodies are considered adjacent whenever

$S_u \cap S_v \neq \emptyset$. These are precisely the intersubsection graphs of k-ply systems of balls, studied by Miller, Teng, Thurston, and Vavasis [MTTV97]. Those authors also provide a generalization of the Lipton-Tarjan separator theorem: For $k = O(1)$, such an intersubsection graph contains a balanced separator of size $O(n^{1-1/d})$.

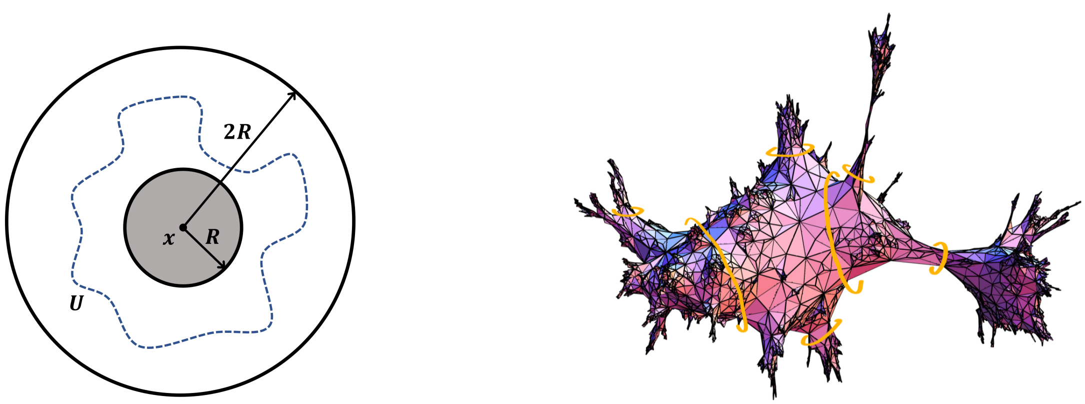
Similarly, finite-element graphs associated to simplicial complexes with bounded aspect ratio can be viewed as subgraphs of geometric overlap graphs [MTTV98], and one again obtains balanced separators of size $O(n^{1-1/d})$. This covers a number of scenarios commonly arising in applications of the finite-element method; we refer to the discussion of well-shaped meshes in [ST07, §6.2].

We will not be too concerned with the particular notion of geometric graph used since our construction satisfies all these commonly employed sets of assumptions. Indeed, it can be cast as the tangency graph of a sphere packing, where adjacent spheres have uniformly comparable radii. It can also be cast as the 1-skeleton of a d-dimensional simplicial complex whose simplices have uniformly bounded aspect ratio as studied. Such graphs were studied by, for instance, by Plotkin, Rao, and Smith [PRS94] in their work on shallow minors (see also the followup work [Ten98]).

Annular separators. Note that the preceding results deal with *global* separators that separate the entire graph into two roughly equal pieces. In many settings, especially those arising in physical simulation, it useful to consider *local* separators. Let $G = (V, E)$ be an undirected graph with path metric d_G, and define graph balls and graph spheres, respectively, by

$$B_G(x, R) := \{y \in V : d_G(x, y) \leqslant R\}$$
$$S_G(x, R) := \{y \in V : d_G(x, y) = R\}.$$

Suppose that for some $C > 1$, we we want to separate $S_G(x, R)$ from $S_G(x, CR)$ by removing a small set of nodes $U \subseteq B_G(x, CR) \setminus B_G(x, R)$. We refer to U as an *annular separator.* See Figure 6.



(a) Separating $S_G(x, R)$ from $S_G(x, 2R)$

(b) A random triangulation[2] of $\mathbb{S}^2$

Figure 6: Annular separators

[2]Depiction of a random triangulation is due to Nicolas Curien.

It is easy to see that even if G is a planar graph, small annular separators don't necessary exist. But in many cases, one can find annular separators of size $O(R)$. For instance, this is true for most vertices at most scales in a uniformly random triangulation of the 2-dimensional sphere [Kri05] (a fact which extends experimentally to a variety of other models of random planar maps, e.g., those studied in [GHS20]). These models also have the properties that the cardinality of graph balls tends to grow asymptotically like $|B_G(x,R)| \sim R^k$ (as $R \to \infty$, up to lower-order fluctuations), where the exponent k depends on the model. For random triangulations, one has $k = 4$ [**?**].

Benjamini and Papasoglou [BP11] give an explanation of this phenomenon as follows. Suppose that G is an infinite planar graph and we assume, additionally, that *G has uniform polynomial growth of degree k:* There exist numbers $C, k \geqslant 1$ such that

$$C^{-1}R^k \leqslant |B_G(x,R)| \leqslant CR^k, \quad \forall x \in V, R \geqslant 1.$$

Then for every $x \in V(G)$ and $R \geqslant 1$, there is a set $U \subseteq B_G(x,2R) \setminus B_G(x,R)$ whose removal disconnects $S_G(x,R)$ from $S_G(x,2R)$ in G, and such that $|U| \leqslant O(R)$. This applies equally well to finite graphs: Indeed, the authors actually show that if the graph metric restricted to $B_G(x,2R)$ has doubling constant λ, then one can find an annular separator of size at most $C_\lambda R$.

We remark that there are rich families of planar graphs with uniform polynomial growth arising in a variety of contexts; see [BS01, BK02, EL20]. Indeed, one can obtain planar graphs with uniform polynomial growth of degree k for all real degrees $k > 1$. Moreover, many models of random planar graphs have an almost sure asymptotic version of this property [DG20, GHS20].

The authors of [BP11] asked whether an analog of this phenomenon holds in higher dimensions. For instance, if G is a graph with uniform polynomial growth that can be geometrically represented in $\mathbb{R}^3$, does it hold that G has annular separators of size $O(R^2)$? We give examples showing that for $d \geqslant 3$, this phenomenon fails in a strong way.

Say that a graph G is *sphere-packed in $\mathbb{R}^d$* if G is the tangency graph of a collection of interior-disjoint spheres in $\mathbb{R}^d$. Say that G is *M-uniformly sphere-packed in $\mathbb{R}^d$* if the collection of spheres can be taken such that the radii of any two tangent spheres lies in the interval $[M^{-1}, M]$, and that G is *uniformly sphere-packed in $\mathbb{R}^d$* if this holds for some $M < \infty$.

Theorem 3.1 (Arbitrarily large annular separators)**.** *For every $d \geqslant 3$ and $s \geqslant 1$, there is a number $c > 0$ and a graph G satisfying:*

1. *G has uniform polynomial growth of degree $O(s)$.*
2. *G is uniformly sphere-packed in $\mathbb{R}^d$.*
3. *For every $x \in V(G)$, there are at least cR^s vertex-disjoint paths from $S_G(x,R)$ to $S_G(x,R')$ for any $R' > R \geqslant 1$.*

Clearly if G has uniform polynomial growth of degree d, then one of the R-many spheres $S(x, R+1), S(x, R+2) \ldots, S(x, 2R)$ must be an annular separator of size $O(R^{d-1})$. We show that the moment the growth degree exceeds d, there are graphs sphere-packed in $\mathbb{R}^d$ that don't have $(d-1)$-dimensional annular separators.

Theorem 3.2 (Nearly-dimensional growth rate)**.** *For every $d \geqslant 3$ and $\varepsilon > 0$, there are numbers $c, \delta > 0$ and a graph G satisfying:*

1. *G has uniform polynomial growth of degree at most $d + \varepsilon$.*

2. *G is uniformly sphere-packed in $\mathbb{R}^d$.*

3. *For every $x \in V(G)$, there are at least $cR^{(d-1)+\delta}$ vertex-disjoint paths from $S_G(x, R)$ to $S_G(x, R')$ for any $R' > R \geqslant 1$.*

Note that the two preceding theorems refer to infinite graphs. A version for families of finite graphs appears in Theorem 3.9.

Preliminaries. We will consider primarily connected, undirected graphs $G = (V, E)$, which we equip with the associated path metric d_G. We write $V(G)$ and $E(G)$, respectively, for the vertex and edge sets of G. If $U \subseteq V(G)$, we write $G[U]$ for the subgraph induced on U.

For $v \in V$, let $\deg_G(v)$ denote the degree of v in G. Let $\mathrm{diam}(G) := \sup_{x,y \in V} d_G(x, y)$ denote the diameter of G (which is only finite for G finite and connected), and for a subset $S \subseteq V$, denote $\mathrm{diam}_G(S) := \sup_{x,y \in S} d_G(x, y)$. For $v \in V$ and $r \geqslant 0$, we use $B_G(v, r) = \{u \in V : d_G(u, v) \leqslant r\}$ to denote the closed ball in G. For subsets $S, T \subseteq V$, we write $d_G(S, T) := \inf\{d_G(s, t) : s \in S, t \in T\}$.

For two expressions A and B, we use the notation $A \lesssim B$ to denote that $A \leqslant CB$ for some *universal* constant C. The notation $A \lesssim_{\alpha,\beta,\cdots} B$ denotes that $A \leqslant C(\alpha, \beta, \cdots)B$ where $C(\alpha, \beta, \cdots)$ denotes a number depending only on the parameters α, β, etc. We write $A \asymp B$ for the conjunction $A \lesssim B \wedge B \lesssim A$.

3.2 Tilings of the unit cube

Fix the dimension $d \geqslant 2$. Our constructions are based on tilings of subsets of $\mathbb{R}^d$ by axis-parallel hyperrectangles, a generalization of the planar constructions in [EL20]. A d-dimensional *tile* is an axis-parallel closed hyperrectangle $A \subseteq \mathbb{R}^d$, i.e., a set of the form $A = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$ for numbers satisfying $a_i < b_i$ for each $i = 1, 2, \ldots, d$.

We will encode such a tile as a $(d + 1)$-tuple $(p(A), \ell_1(A), \ell_2(A), \ldots, \ell_d(A))$, where $p(A) := (a_1, a_2, \ldots, a_d)$ and $\ell_i(A) := b_i - a_i$ is the length of the projection of A along the ith axis.

A *tiling* $\boldsymbol{T}$ is a finite collection of interior-disjoint tiles. Denote $[\![\boldsymbol{T}]\!] := \bigcup_{A \in \boldsymbol{T}} A$. If $R \subseteq \mathbb{R}^d$, we say that $\boldsymbol{T}$ *is a tiling of* R if $[\![\boldsymbol{T}]\!] = R$. We associate to a tiling its *tangency graph* $G(\boldsymbol{T})$ with vertex set $\boldsymbol{T}$ and with an edge between two tiles $A, B \in \boldsymbol{T}$ whenever $A \cap B$ has non-zero $(d - 1)$-dimensional volume.

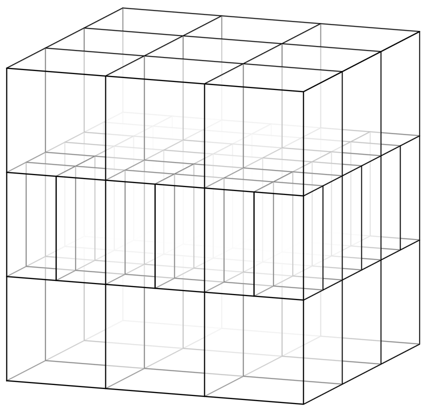
Denote by $\mathcal{T}_d$ the set of all tilings of the unit d-dimensional cube $[0, 1]^d$. See Figure 7(a) for a tiling of $[0, 1]^3$ and Figure 7(b) for a representation of its tangency graph. For the remainder of the paper, we will consider only tilings $\boldsymbol{T}$ for which $G(\boldsymbol{T})$ is connected.

Definition 3.3 (Tiling product)**.** For $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{T}_d$, define the product $\boldsymbol{S} \circ \boldsymbol{T} \in \mathcal{T}_d$ as the tiling formed by replacing every tile in $\boldsymbol{S}$ by an (appropriately scaled) copy of $\boldsymbol{T}$. More precisely: For every $A \in \boldsymbol{S}$ and $B \in \boldsymbol{T}$, there is a tile $R \in \boldsymbol{S} \circ \boldsymbol{T}$ with $\ell_i(R) := \ell_i(A)\ell_i(B)$, and

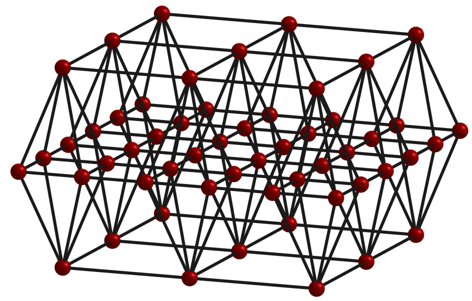
$$p_i(R) := p_i(A) + p_i(B)\ell_i(A),$$

for each $i = 1, 2, \ldots, d$.

If $\boldsymbol{T} \in \mathcal{T}_d$ and $n \geqslant 0$, we use $\boldsymbol{T}^n := \boldsymbol{T} \circ \cdots \circ \boldsymbol{T}$ to denote the n-fold tile product of $\boldsymbol{T}$ with itself, where $\boldsymbol{T}^0 := \boldsymbol{I}_d$ and $\boldsymbol{I}_d := \{[0, 1]^d\} \in \mathcal{T}_d$ is the identity tiling. The following observation shows that this is well-defined.



(a) $\boldsymbol{T}^{(3)}_{\langle 3,6,3\rangle}$



(b) The tangency graph

Figure 7: A tiling and its tangency graph

Observation 3.4. The tiling product is associative: $(\boldsymbol{S} \circ \boldsymbol{T}) \circ \boldsymbol{U} = \boldsymbol{S} \circ (\boldsymbol{T} \circ \boldsymbol{U})$ for all $\boldsymbol{S}, \boldsymbol{T}, \boldsymbol{U} \in \mathcal{T}_d$. Moreover, if $\boldsymbol{I_d} \in \mathcal{T}_d$ consists of the single tile $[0,1]^d$, then $\boldsymbol{T} \circ \boldsymbol{I_d} = \boldsymbol{I_d} \circ \boldsymbol{T}$ for all $\boldsymbol{T} \in \mathcal{T}_d$.

3.2.1 The construction

Given a sequence $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$ with $\gamma_i \in \mathcal{N}$, we define an associated tiling $\boldsymbol{T}^{(d)}_\gamma \in \mathcal{T}_d$ as follows: For $1 \leqslant i \leqslant b$, fill $[0,1]^{d-1} \times [(i-1)/b, i/b]$ with γ_i^{d-1} copies of $[0, 1/\gamma_i]^{d-1} \times [0, 1/b]$ formed into a $\underbrace{\gamma_i \times \cdots \times \gamma_i}_{d-1 \text{ times}}$ grid. For example, see $\boldsymbol{T}^{(d)}_{\langle 3,6,3\rangle}$ in Figure 7(a). The following observation will be useful.

Observation 3.5. For $\gamma = \langle \gamma_1, \ldots, \gamma_a \rangle$, $\gamma' = \left\langle \gamma'_1, \ldots, \gamma'_b \right\rangle$, we have $\boldsymbol{T}^{(d)}_\gamma \circ \boldsymbol{T}^{(d)}_{\gamma'} = \boldsymbol{T}^{(d)}_{\gamma \otimes \gamma'}$ where

$$\gamma \otimes \gamma' = \left\langle \gamma_1\gamma'_1, \ldots, \gamma_1\gamma'_b, \ldots, \gamma_a\gamma'_1, \ldots, \gamma_a\gamma'_b \right\rangle.$$

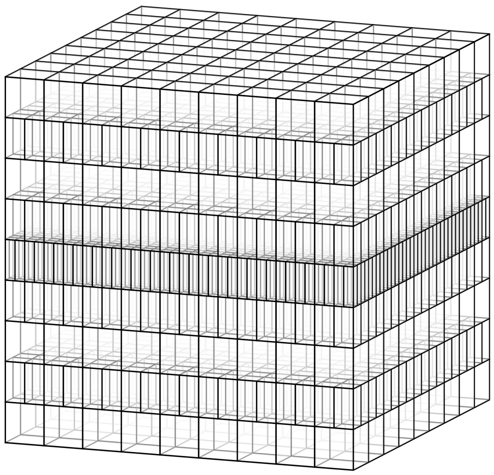
In particular, for $n \geqslant 1$ it holds that $\left(\boldsymbol{T}^{(d)}_\gamma\right)^n = \boldsymbol{T}^{(d)}_{\gamma^{\otimes n}}$, where $\gamma^{\otimes n} := \gamma \otimes \gamma \otimes \cdots \otimes \gamma$ is the n-fold tensor product of γ with itself. See Figure 8(a) for a representation of $\boldsymbol{T}^{(3)}_{\langle 3,6,3\rangle^{\otimes 2}}$. We will use these two representations interchangeably throughout the paper.

We use the notations $|\gamma| := b$ and

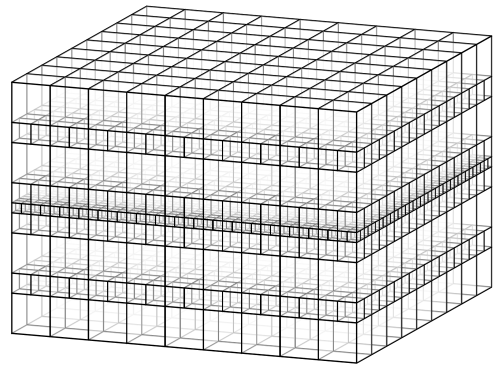
$$|\gamma|^{(d)} := |\boldsymbol{T}^{(d)}_\gamma| = \gamma_1^{d-1} + \cdots + \gamma_b^{d-1}.$$

Note that $|\gamma \otimes \gamma'|^{(d)} = |\gamma|^{(d)}|\gamma'|^{(d)}$, hence

$$\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}} = |\gamma^{(d)}|^n. \tag{3.1}$$



(a) $\boldsymbol{T}^{(3)}_{\langle 3,6,3\rangle} \circ \boldsymbol{T}^{(3)}_{\langle 3,6,3\rangle} = \boldsymbol{T}^{(3)}_{\langle 3,6,3\rangle^{\otimes 2}}$



(b) $\tilde{\boldsymbol{T}}^{(3)}_{\langle 3,6,3\rangle^{\otimes 2}}$

Figure 8: A tile product and its cube packing

Further, denote

$$k^{(d)}(\gamma) := \frac{\log\left(|\gamma|^{(d)}\right)}{\log|\gamma|}.$$

We now state three key lemmas proved that are proved in subsequent subsections, and then use them to prove our main theorems. The first establishes uniform polynomial volume growth for the graphs $\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}}$. It is proved in Section 3.3.

Lemma 3.6 (Volume growth)**.** *Consider $\gamma = \langle \gamma_1, \ldots, \gamma_b\rangle$ with $\min(\gamma) = \gamma_1 = \gamma_b = b$. Then for all $d \geqslant 2$, there is a number $C = C(d, \gamma)$ such that the family of graphs $\mathcal{F} = \left\{G\left(\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}}\right) : n \geqslant 0\right\}$ has uniform polynomial growth of degree $k = k^{(d)}(\gamma)$ in the sense that*

$$C^{-1}R^k \leqslant |B_G(x,R)| \leqslant CR^k, \qquad \forall G \in \mathcal{F}, x \in V(G), 1 \leqslant R \leqslant \mathrm{diam}(G).$$

The second lemma, proved in Section 3.4 establishes a lower bound on the size of annular separators.

Lemma 3.7 (Separator size)**.** *For every $d \geqslant 2$ and $b \geqslant 1$, there is a number $c = c(d,b)$ such that for any $\gamma = \langle \gamma_1, \ldots, \gamma_b\rangle$ with $\min(\gamma) = b$, the following holds. Denote $\tilde{k} := k^{(d-1)}(\gamma)$. For any $n \geqslant 1$, if $G = G(\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}})$, and $v \in V(G)$, then there are at least $cR^{\tilde{k}}$ disjoint paths from $S_G(v,R)$ to $S_G(v,R')$, for any $1 \leqslant R < R' \leqslant \mathrm{diam}(G)/(3(d+1))$.*

If G is an undirected graph, let us write $[G]_m$ for the *m-subdivision of G*, where each edge of G is replaced by a path of length m. We will sometimes consider $V(G) \subseteq V([G]_m)$ via the obvious identification. The next lemma is proved in Section 3.5.

Lemma 3.8. *Consider a sequence $\gamma = \langle \gamma_1, \dots, \gamma_b \rangle$ with $\gamma_1 = \gamma_b$, and such that*

$$\max\left\{\frac{\gamma_{i+1}}{\gamma_i}, \frac{\gamma_i}{\gamma_{i+1}}\right\} \in \mathbb{N}, \quad \forall 1 \leqslant i < b. \tag{3.2}$$

Then for every $d \geqslant 2$, there are numbers $m = m(d, \gamma)$ and $M = M(d, \gamma)$ such that for every $n \geqslant 0$: If $G = G(\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}})$, then $[G]_m$ is M-uniformly sphere-packed in $\mathbb{R}^d$.

With these results in our hand, let us first prove a finitary version of our main theorems. The corresponding infinite version appears in Section 3.2.2. Define

$$s(d,k) := d - 1 + (k - d)\left(1 - \frac{1}{d-1}\right).$$

This represents our basic tradeoff: One can construct d-dimensional geometric graphs with uniform polynomial growth of degree $k + \varepsilon$ and such that every annular separator has size $\Omega(R^{s(d,k)})$.

Theorem 3.9 (Finite graph families)**.** *For every $d \geqslant 3$, $k \geqslant d$, and $\varepsilon > 0$, there is a family $\mathcal{F}$ of finite graphs satisfying:*

1. *For some $\tilde{k} \leqslant k + \varepsilon$ and every $G \in \mathcal{F}$,*

$$|B_G(x, R)| \asymp_{d,\varepsilon} R^{\tilde{k}}, \quad \forall x \in V(G), 1 \leqslant R \leqslant \mathrm{diam}(G).$$

2. *Each G is M-uniformly sphere-packed in $\mathbb{R}^d$ for some $M \lesssim_{d,\varepsilon} 1$.*
3. *There is a number $c \gtrsim_{d,\varepsilon} 1$ such that for every $G \in \mathcal{F}$ and $x \in V(G)$, there are at least $cR^{s(d,k)}$ vertex-disjoint paths from $S_G(x, R)$ to $S_G(x, R')$ for every $1 \leqslant R < R' \leqslant c\,\mathrm{diam}(G)$.*

In light of Lemma 3.6–Lemma 3.8, we can take $\mathcal{F} = \left\{[G(\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}})]_m : n \geqslant 0\right\}$ as long as we can find a sequence γ suited to the parameters. To this end, consider $d \geqslant 2$ and parameters $h, p, q \in \mathbb{N}$ such that $p \geqslant qd$. Define the sequence

$$\gamma^{(p,q,h)} := \langle b, \underbrace{tb, \dots, tb}_{b-2 \text{ copies}}, b \rangle,$$

where $b := h^{q(d-1)}$ and $t := h^{p-dq}$. Note that $t \geqslant 1$ since $p \geqslant qd$. Moreover, $\gamma^{(p,q,h)}$ satisfies (3.2) by construction.

Lemma 3.10. *Let $d \geqslant 2$ and $k := p/q$ be as above. Then for all $\varepsilon > 0$, there is some $h_0 \in \mathbb{N}$ so that for all $h \geqslant h_0$ the following statements hold:*

1. $\left|k^{(d)}(\gamma^{(p,q,h)}) - k\right| < \varepsilon.$
2. $\left|k^{(d-1)}(\gamma^{(p,q,h)}) - s(d,k)\right| < \varepsilon.$

Proof. It holds that

$$\begin{aligned} k^{(d)}(\gamma^{(p,q,h)}) = \log_b(|\gamma^{(p,q,h)}|^{(d)}) &= \log_b\left(b^{d-1}(2+(b-2)t^{d-1})\right) \\ &= \log_b\left(b^{d-1}(2+(b-2)h^{(d-1)(p-dq)})\right) \\ &= \log_b\left(b^{d-1}(2+(b-2)b^{k-d})\right). \end{aligned}$$

Furthermore, we have

$$\lim_{h\to\infty} \log_b(|\gamma^{(p,q,h)}|^{(d)}) = \lim_{b\to\infty} \log_b(b^{d-1}(2+(b-2)b^{(k-d)})) = k.$$

Therefore for all $\varepsilon > 0$, for all sufficiently large values of h, it holds that $|k^{(d)}(\gamma^{(p,q,h)}) - k| \leqslant \varepsilon$. Similarly, we have

$$\lim_{b\to\infty} \log_b(|\gamma^{(p,q,h)}|^{(d-1)}) = s(d,k),$$

hence we can choose h sufficiently large to as to satisfy the second condition as well. □

3.2.2 Construction of the infinite graphs

For this subsection alone, we consider tilings of the nonnegative orthant $[0,\infty)^d$.

Note that for any sequence $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$, one can view $G(\boldsymbol{T}_\gamma^{(d)})$ as the tangency graph of a packing of cubes by changing the height of cubes in layer i from $1/b$ to b/γ_i. See Figure 8(b) for an example. Let $\tilde{\boldsymbol{T}}_\gamma^{(d)}$ denote this rescaled tiling. By convention, we insist that one corner of the tiling still lies at the origin, implying that

$$[\![\tilde{\boldsymbol{T}}_\gamma^{(d)}]\!] = [0,1]^{d-1} \times [0, H(\gamma)],$$

where

$$H(\gamma) := b\left(\frac{1}{\gamma_1} + \cdots + \frac{1}{\gamma_b}\right).$$

Since for $n \geqslant 0$ we have $\tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes(n+1)}} = \tilde{\boldsymbol{T}}^{(d)}_{\gamma} \circ \tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes n}}$, it follows that $\tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes n}} \subseteq \tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes(n+1)}}$ and hence we have the chain of inclusions:

$$\tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes 0}} \subseteq \cdots \subseteq \tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes(n-1)}} \subseteq \tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes n}} \subseteq \tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes(n+1)}} \subseteq \cdots$$

This gives rise, in a straightforward way, to the infinite tiling $\tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes\mathbb{N}}}$, with $[\![\tilde{\boldsymbol{T}}^{(d)}_{\gamma^{\otimes\mathbb{N}}}]\!] = [0,\infty)^d$. We define the infinite tangency graph $\hat{G}_\gamma^{(d)} := G(\boldsymbol{T}^{(d)}_{\gamma^{\otimes\mathbb{N}}})$.

Theorem 3.11. *If $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$ has $\gamma_1 = \gamma_b = b = \min(\gamma)$, then $\hat{G}_\gamma^{(d)}$ has uniform polynomial growth of degree $k^{(d)}(\gamma)$, and for every $x \in V(G)$ and $R' > R \geqslant 1$, there are at least $cR^{k^{(d-1)}(\gamma)}$ vertex-disjoint paths from $S_G(x,R)$ to $S_G(x,R')$, where $c \gtrsim_{\gamma,d} 1$.*

Proof. For $n \geqslant 1$, denote $G_n := G(\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}})$ and $\hat{G} := \hat{G}_\gamma^{(d)}$. We can think of G_n as an induced subgraph of $\hat{G}$ in the obvious way. Consider a vertex $v \in V(\hat{G})$ and radii $R' > R \geqslant 1$. Then there is some $n \geqslant 1$ such that $B_{\hat{G}}(v, 2R') \subseteq V(G_n)$.

In this case, the it holds that $d_{\hat{G}}(x, y) = d_{G_n}(x, y)$ for all $x, y \in B_{\hat{G}}(v, R')$, since any path originating in $B_{\hat{G}}(v, R')$ and leaving $V(G_n)$ must have length at least $2R'$. Therefore $B_{\hat{G}}(v, R') = B_{G_n}(v, R')$, and the uniform polynomial growth and seperator size assertions then follow immediately from Lemma 3.6 and Lemma 3.7. □

Recall the definition of $[G]_m$ from Lemma 3.8. The following theorem is proved in Section 3.5.

Theorem 3.12. *If $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$ satisfies $\gamma_1 = \gamma_b$ and (3.2), then there is some number $m = m(d, \gamma)$ such that $[\hat{G}_\gamma^{(d)}]_m$ is uniformly sphere-packed in $\mathbb{R}^d$.*

3.3 Volume growth analysis

Our goal is now to prove Lemma 3.6. The next subsection provides a few key lemmas about the size of balls in products of tilings, which are mostly straightforward generalizations of the bounds in [EL20] (for the case $d = 2$). With these in hand, we prove Lemma 3.6 in Section 3.3.2.

3.3.1 Volume growth in tile products

The next lemma is straightforward.

Lemma 3.13. *Consider $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{T}_d$ and $G = G(\boldsymbol{S} \circ \boldsymbol{T})$. If $X \in \boldsymbol{S} \circ \boldsymbol{T}$, then $|B_G(X, \mathrm{diam}(G(\boldsymbol{T})))| \geqslant |\boldsymbol{T}|$.*

Let $\{e_1, \ldots, e_d\}$ denote the standard basis of $\mathbb{R}^d$. If $\boldsymbol{T} \in \mathcal{T}_d$, we write $E_i(\boldsymbol{T}) \subseteq E(G(\boldsymbol{T}))$ for the set of edges in the ith direction, i.e., those edges $\{A, B\} \in E(G(\boldsymbol{T}))$ where $A \cap B$ is orthogonal to e_i. Thus we have a partition $E(G(\boldsymbol{T})) = E_1(\boldsymbol{T}) \cup \cdots \cup E_d(\boldsymbol{T})$.

For $A \in \boldsymbol{T}$ and $1 \leqslant i \leqslant d$, denote

$$N_{\boldsymbol{T}}(A, i) = \{A' \in \boldsymbol{T} : \{A, A'\} \in E_i(\boldsymbol{T})\},$$

and $N_{\boldsymbol{T}}(A) := N_{\boldsymbol{T}}(A, 1) \cup \cdots \cup N_{\boldsymbol{T}}(A, d)$. Moreover, we define

$$\begin{aligned}
\alpha_{\boldsymbol{T}}(A, i) &:= \max\left\{\frac{\ell_j(A)}{\ell_j(B)} : B \in N_{\boldsymbol{T}}(A, i), 1 \leqslant j \leqslant d\right\} \\
\alpha_{\boldsymbol{T}}(i) &:= \max\{\alpha_{\boldsymbol{T}}(A, i) : A \in \boldsymbol{T}\} \\
\alpha_{\boldsymbol{T}} &:= \max\{\alpha_{\boldsymbol{T}}(i) : 1 \leqslant i \leqslant d\} \\
L_{\boldsymbol{T}} &:= \max\{\ell_i(A) : A \in \boldsymbol{T}, 1 \leqslant i \leqslant d\}.
\end{aligned}$$

We will take $\alpha_{\boldsymbol{T}} := 1$ if $\boldsymbol{T}$ contains a single tile. It is now straightforward to check that $\alpha_{\boldsymbol{T}}$ bounds the degrees in $G(\boldsymbol{T})$.

Lemma 3.14. *For a tiling $\boldsymbol{T} \in \mathcal{T}_d$ and $A \in \boldsymbol{T}$, it holds that*

$$\deg_{G(\boldsymbol{T})}(A) \leqslant 3^d \cdot \alpha_{\boldsymbol{T}}^{2d}.$$

Proof. Denote

$$\tilde{A} := \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d : p_i(A) - \alpha_{\boldsymbol{T}} \ell_i(A) \leqslant x_i \leqslant p_i(A) + (1 + \alpha_{\boldsymbol{T}})\ell_i(A)\right\},$$

where $p_i(A)$ denotes the ith coordinate of $p(A)$. Clearly

$$\mathrm{vol}_d(\tilde{A}) = \prod_{i=1}^{d} ((1 + 2\alpha_{\boldsymbol{T}})\ell_i(A)) = (1 + 2\alpha_{\boldsymbol{T}})^d \,\mathrm{vol}_d(A).$$

Furthermore, by the definition of $\alpha_{\boldsymbol{T}}$, it holds that $A' \subseteq \tilde{A}$ for $A' \in N_{\boldsymbol{T}}(A)$. And for any such A', it holds that $\mathrm{vol}_d(A') \geqslant \alpha_{\boldsymbol{T}}^{-d} \,\mathrm{vol}_d(A)$. Hence,

$$|N_{\boldsymbol{T}}(A)|\alpha_{\boldsymbol{T}}^{-d} \,\mathrm{vol}_d(A) \leqslant (1 + 2\alpha_{\boldsymbol{T}})^d \,\mathrm{vol}_d(A).$$

Using $\alpha_{\boldsymbol{T}} \geqslant 1$, this yields $\deg_{G(\boldsymbol{T})}(A) \leqslant 3^d \alpha_{\boldsymbol{T}}^{2d}$. ◻

Lemma 3.15. *Consider $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{T}_d$ and let $G = G(\boldsymbol{S} \circ \boldsymbol{T})$. Then for any $X \in \boldsymbol{S} \circ \boldsymbol{T}$, it holds that*

$$|B_G(X, 1/(\alpha_{\boldsymbol{S}}^{2d} L_{\boldsymbol{T}}))| \leqslant (3\alpha_{\boldsymbol{S}}^2)^{d^2} (d + 1)|\boldsymbol{T}|. \tag{3.3}$$

Proof. For $Y \in \boldsymbol{S} \circ \boldsymbol{T}$, let $\hat{Y} \in \boldsymbol{S}$ denote the unique tile for which $Y \subseteq \hat{Y}$. Let $\Gamma_{\boldsymbol{S}}(\hat{X})$ denote the set of paths γ in $G(\boldsymbol{S})$ that originate from $\hat{X}$ and such that $|\gamma \cap E_i(\boldsymbol{S})| \leqslant 1$ for each $i = 1, \ldots, d$. In other words, the set of paths in $G(\boldsymbol{S})$ starting at $\hat{X}$ and containing at most one edge in every direction. Denote by $\tilde{N}_{\boldsymbol{S}}(\hat{X}) := \bigcup_{\gamma \in \Gamma_{\boldsymbol{S}}(\hat{X})} V(\gamma)$ the set of vertices reachable via such paths.

Note that because we allow one edge in every direction,

$$\hat{X} + [-\ell_1(\hat{X})/\alpha_{\boldsymbol{S}}^d, \ell_1(\hat{X})/\alpha_{\boldsymbol{S}}^d] \times \cdots \times [-\ell_d(\hat{X})/\alpha_{\boldsymbol{S}}^d, \ell_d(\hat{X})/\alpha_{\boldsymbol{S}}^d] \subseteq \left[\!\left[\tilde{N}_{\boldsymbol{S}}(\hat{X})\right]\!\right], \tag{3.4}$$

where '+' here is the Minkowski sum $R + S := \{r + s : r \in R, s \in S\}$.

We will now show that

$$\left[\!\left[B_G(X, 1/(\alpha_{\boldsymbol{S}}^{2d} L_{\boldsymbol{T}}))\right]\!\right] \subseteq \left[\!\left[\tilde{N}_{\boldsymbol{S}}(\hat{X})\right]\!\right]. \tag{3.5}$$

It will follow that

$$|B_G(X, 1/(\alpha_{\boldsymbol{S}}^{2d} L_{\boldsymbol{T}}))| \leqslant |\boldsymbol{T}| \cdot |\tilde{N}_{\boldsymbol{S}}(\hat{X})| \leqslant |\boldsymbol{T}| \cdot (d + 1) \left(\max_{A \in \boldsymbol{S}} \deg_{G(\boldsymbol{S})}(A)\right)^d,$$

and then (3.3) follows from Lemma 3.14.

To establish (3.5), consider any path $\langle X = X_0, X_1, X_2, \ldots, X_h \rangle$ in G with $\hat{X}_h \notin \tilde{N}_{\boldsymbol{S}}(\hat{X})$. Let $k \leqslant h$ be the smallest index for which $\hat{X}_k \notin \tilde{N}_{\boldsymbol{S}}(\hat{X})$. Then:

$$X_0, X_1, \ldots, X_{k-1} \subseteq \left[\!\left[\tilde{N}_{\boldsymbol{S}}(\hat{X})\right]\!\right] \tag{3.6}$$

$$X_{k-1} \cap \left(\partial \left[\!\left[\tilde{N}_{\boldsymbol{S}}(\hat{X})\right]\!\right] \cap (0,1)^d\right) \neq \emptyset. \tag{3.7}$$

Now (3.6) implies that for $j \leqslant k - 1$, we have $\hat{X}_j \in \tilde{N}_{\boldsymbol{S}}(\hat{X})$, which implies $\ell_i(\hat{X}_j) \leqslant \alpha_{\boldsymbol{S}}^d \ell_i(\hat{X})$ since $\hat{X}_j$ can be reached from $\hat{X}$ by a path of length at most d in $G(\boldsymbol{S})$. It follows that

$$\ell_i(X_j) \leqslant L_{\boldsymbol{T}} \ell_i(\hat{X}_j) \leqslant L_{\boldsymbol{T}} \alpha_{\boldsymbol{S}}^d \ell_i(\hat{X}), \qquad j \leqslant k - 1, 1 \leqslant i \leqslant d. \tag{3.8}$$

And (3.7) together with (3.4) shows that

$$\sum_{j=0}^{k-1} \ell_i(X_j) \geqslant \ell_i(\hat{X})/\alpha_{\boldsymbol{S}}^d, \qquad 1 \leqslant i \leqslant d. \tag{3.9}$$

Combining (3.8) and (3.9) now gives

$$h-1 \geqslant k-1 \geqslant \frac{1}{\alpha_{\boldsymbol{S}}^{2d} L_{\boldsymbol{T}}},$$

verifying (3.5) and completing the proof. □

3.3.2 Volume growth in iterated products

Our goal is now to prove Lemma 3.6. To this end, fix $d \geqslant 2$.

Observation 3.16. For $1 \leqslant i \leqslant d-1$, we have $\alpha_{\boldsymbol{T}_\gamma^{(d)}}(i) = 1$. It further holds that

$$\alpha_{\boldsymbol{T}_\gamma^{(d)}} = \alpha_{\boldsymbol{T}_\gamma^{(d)}}(d) = \max\left\{\frac{\gamma_i}{\gamma_j} : 1 \leqslant i, j \leqslant b, |i-j| = 1\right\}.$$

Given Observation 3.5 and Observation 3.16, the next lemma is straightforward.

Lemma 3.17. *Consider $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$ and $\gamma' = \langle \gamma'_1, \ldots, \gamma_{b'} \rangle$. If $\gamma'_1 = \gamma'_{b'}$, then*

$$\alpha_{\boldsymbol{T}_\gamma^{(d)} \circ \boldsymbol{T}_{\gamma'}^{(d)}} = \alpha_{\boldsymbol{T}_{\gamma\otimes\gamma'}^{(d)}} \leqslant \max(\alpha_{\boldsymbol{T}_\gamma^{(d)}}, \alpha_{\boldsymbol{T}_{\gamma'}^{(d)}}).$$

Proof. Note that, by definition for $1 \leqslant j \leqslant b'$ and $0 \leqslant i \leqslant b-1$, we have $(\gamma \otimes \gamma')_{ib'+j} = \gamma_i \gamma'_j$. Hence,

$$\frac{(\gamma \otimes \gamma')_{ib'+j+1}}{(\gamma \otimes \gamma')_{ib'+j}} = \begin{cases} \frac{\gamma'_{j+1}}{\gamma'_j} & j < b' \\ \frac{\gamma_{i+1}}{\gamma_i}\frac{\gamma'_1}{\gamma'_{b'}} = \frac{\gamma_{i+1}}{\gamma_i} & j = b', \end{cases}$$

where we used $\gamma'_1 = \gamma'_{b'}$ in the second case. □

Corollary 3.18. *Let $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$. If $\gamma_1 = \gamma_b$, then for $n \geqslant 1$ we have*

$$\alpha_{\boldsymbol{T}_{\gamma^{\otimes n}}^{(d)}} \leqslant \alpha_{\boldsymbol{T}_\gamma^{(d)}}.$$

Lemma 3.19. *Consider $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$ with $\gamma_1 = \gamma_b = b$. For every $n \geqslant 0$, it holds that*

$$b^n - 1 \leqslant \operatorname{diam}\left(G(\boldsymbol{T}_{\gamma^{\otimes n}}^{(d)})\right) \leqslant (d+1)b^n. \tag{3.10}$$

Proof. Let C denote the "bottom" layer of tiles in $\boldsymbol{T}_{\gamma^{\otimes n}}^{(d)}$, i.e., those contained in $[0,1]^{d-1} \times [0, 1/b^n]$. Take $G := G(\boldsymbol{T}_{\gamma^{\otimes n}}^{(d)})$. For every $A \in \boldsymbol{T}_{\gamma^{\otimes n}}^{(d)}$, it holds that $\ell_d(A) = b^{-n}$. In particular, this yields the lower bound in (3.10) since any path from the bottom to the top (in dimension d) requires b^n tiles.

This also implies that for $A \in \boldsymbol{T}_{\gamma^{\otimes n}}^{(d)}$ we have $d_G(A, C) \leqslant b^n$. Furthermore, as $\gamma_1 = b$ by assumption, we have $(\gamma^{\otimes n})_1 = b^n$, and therefore by construction $G[C]$ is a $(d-1)$-dimensional $b^n \times b^n \times \cdots \times b^n$ grid. Hence we have $\operatorname{diam}_G(C) \leqslant (d-1)b^n$, and now the upper bound in (3.10) follows by the triangle inequality. □

Proof of Lemma 3.6. Denote $G := G(\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}})$ and fix $A \in \boldsymbol{T}^{(d)}_{\gamma^{\otimes n}}$. For any $0 \leqslant j \leqslant n$, write $\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}} = \boldsymbol{T}^{(d)}_{\gamma^{\otimes(n-j)}} \circ \boldsymbol{T}^{(d)}_{\gamma^{\otimes j}}$ and let $\hat{A}$ denote the copy of $\boldsymbol{T}^{(d)}_{\gamma^{\otimes j}}$ containing A. By Lemma 3.19, we have $\operatorname{diam}(\hat{A}) \leqslant (d+1)b^j$, and thus $B_G(A, (d+1)b^j) \supseteq \hat{A}$.

Employing Lemma 3.13 therefore gives

$$\left|B_G(A, (d+1)b^j)\right| \geqslant |\hat{A}| = \left|\boldsymbol{T}^{(d)}_{\gamma^{\otimes j}}\right| = \left(|\gamma|^{(d)}\right)^j, \qquad \forall A \in \boldsymbol{T}^{(d)}_{\gamma^{\otimes n}},\ j \in \{0, 1, \ldots, n\}.$$

The desired lower bound now follows using monotonicity of $|B_G(A, r)|$ with respect to r.

To prove the upper bound, first note that by Corollary 3.18 we have

$$\alpha_{\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}}} \leqslant \alpha_{\boldsymbol{T}^{(d)}_{\gamma}} \lesssim_\gamma 1.$$

Moreover, as $\min(\gamma) = b$, we have $L_{\boldsymbol{T}^{(d)}_{\gamma^{\otimes j}}} = b^{-j}$, hence applying Lemma 3.15 with $S = \boldsymbol{T}^{(d)}_{\gamma^{\otimes(n-j)}}$ and $T = \boldsymbol{T}^{(d)}_{\gamma^{\otimes j}}$ gives

$$\left|B_G(A, b^j/\xi)\right| \lesssim_{d,\gamma} |\boldsymbol{T}^{(d)}_{\gamma^{\otimes j}}| = \left(|\gamma|^{(d)}\right)^j, \qquad \forall A \in \boldsymbol{T}^{(d)}_{\gamma^{\otimes n}},\ j \in \{0, 1, \ldots, n\},$$

for some $\xi \lesssim_{\gamma,d} 1$, completing the proof. □

3.4 The size of annular separators

We now prove that the graphs $G(\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}})$ do not have small annular separators.

Definition 3.20 (Projection of tiles)**.** For $d \geqslant 2$ and a tile $A \subseteq \mathbb{R}^d$, we write $\Pi_{d-1}(A) \subseteq \mathbb{R}^{d-1}$ for the projection of A onto the last $d-1$ coordinates. Furthermore, for a tiling $\boldsymbol{T} \in \mathcal{T}_d$, we define

$$\Pi_{d-1}(\boldsymbol{T}) := \{\Pi_{d-1}(A) \mid A \in \boldsymbol{T}\},$$

and for a tile $B \in \Pi_{d-1}(\boldsymbol{T})$,

$$\Pi^{-1}_{d-1}(B; \boldsymbol{T}) := \{A \in \boldsymbol{T} : \Pi_{d-1}(A) = B\}.$$

Observation 3.21. For any sequence $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$, the following hold:

(a) $\Pi_{d-1}(\boldsymbol{T}^{(d)}_\gamma) = \boldsymbol{T}^{(d-1)}_\gamma$.

(b) For all $B \in \Pi_{d-1}(\boldsymbol{T}^{(d)}_\gamma)$, the tangency graph $G(\Pi^{-1}_{d-1}(B; \boldsymbol{T}^{(d)}_\gamma))$ is a path.

(c) For all $B \neq B' \in \Pi_{d-1}(\boldsymbol{T}^{(d)}_\gamma)$, the sets $\Pi^{-1}_{d-1}(B; \boldsymbol{T}^{(d)}_\gamma)$ and $\Pi^{-1}_{d-1}(B'; \boldsymbol{T}^{(d)}_\gamma)$ are disjoint.

See Figure 9.

We will now use the family $\{\Pi^{-1}_{d-1}(B; \boldsymbol{T}^{(d)}_\gamma) : B \in \Pi_{d-1}(\boldsymbol{T}^{(d)}_\gamma)\}$ of pairwise disjoint paths to locate paths across annuli in $G(\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}})$. We first remark that diameter of each such path is long in G.

Lemma 3.22. *For every $B \in \Pi_{d-1}(\boldsymbol{T}_\gamma^{(d)})$,*

$$\mathrm{diam}_{G(\boldsymbol{T}_\gamma^{(d)})}(\Pi_{d-1}^{-1}(B; \boldsymbol{T}_\gamma^{(d)})) \geqslant \frac{1}{L_{\boldsymbol{T}_\gamma^{(d)}}} - 1,$$

Proof. Let $\Pi_1 : \mathbb{R}^d \to \mathbb{R}$ denote the projection onto the first coordinate. Then the Euclidean diameter of $\Pi_1(\Pi_{d-1}^{-1}(B; \boldsymbol{T}_\gamma^{(d)}))$ is precisely 1. On the other hand, for any path P in $G(\boldsymbol{T}_\gamma^{(d)})$, the Euclidean diameter of $\Pi_1(P)$ is at most $(\mathrm{len}(P) + 1)L_{\boldsymbol{T}_\gamma^{(d)}}$, and the result follows. □

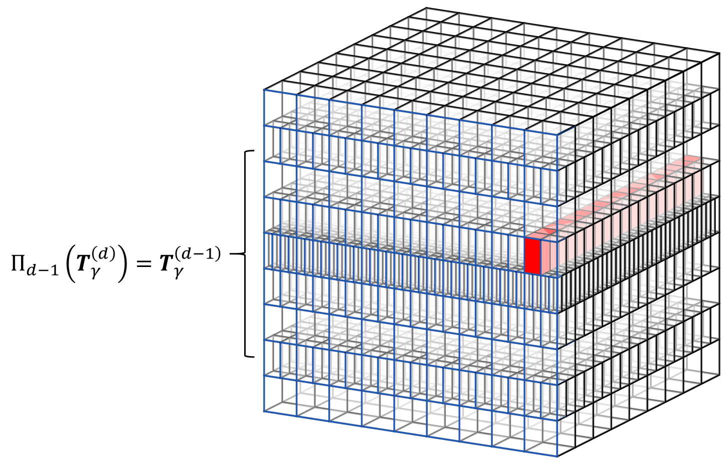


Figure 9: Projection to the last $d-1$ coordinates and $\Pi_{d-1}^{-1}(B; \boldsymbol{T}_\gamma^{(d)})$ for one $B \in \Pi_{d-1}(\boldsymbol{T}_\gamma^{(d)})$.

With this in hand, we can now exhibit many disjoint paths across annuli in G.

Proof of Lemma 3.7. Define $G := G(\boldsymbol{T}_{\gamma^{\otimes n}}^{(d)})$ and $h := \lfloor \log_b(R/(d+1)) \rfloor$. Let $A \in \boldsymbol{T}_{\gamma^{\otimes n}}^{(d)}$ be arbitrary. Write $\boldsymbol{T}_{\gamma^{\otimes n}}^{(d)} = \boldsymbol{T}_{\gamma^{\otimes (n-h)}}^{(d)} \circ \boldsymbol{T}_{\gamma^{\otimes h}}^{(d)}$, and let $\hat{A} \subseteq \boldsymbol{T}_{\gamma^{\otimes n}}^{(d)}$ be the copy of $\boldsymbol{T}_{\gamma^{\otimes h}}^{(d)}$ that contains A. By Lemma 3.19, we have $\mathrm{diam}_G(\hat{A}) \leqslant (d+1)b^h \leqslant R$, where the latter inequality follows from our definition of h. Thus $\hat{A} \subseteq B_G(A, R)$.

But as $\hat{A}$ is a translation of $\boldsymbol{T}_{\gamma^{\otimes h}}^{(d)}$, by Observation 3.21(a), it holds that $\Pi_{d-1}(\boldsymbol{T}_{\gamma^{\otimes h}}^{(d)})$ is a translation of $\boldsymbol{T}_{\gamma^{\otimes h}}^{(d-1)}$. This yields

$$|\Pi_{d-1}\left(B_G(A, R)\right)| \geqslant |\Pi_{d-1}(\hat{A})| = |\boldsymbol{T}_{\gamma^{\otimes h}}^{(d-1)}| = \left(|\gamma|^{(d-1)}\right)^h.$$

Using Observation 3.21(b)–(c), the sets $\{\Pi_{d-1}^{-1}(B; \boldsymbol{T}_{\gamma^{\otimes n}}^{(d)}) : B \in \Pi_{d-1}(B_G(A, R))\}$ form a collection of vertex-disjoint paths in G, and Lemma 3.22 gives a lower bound on the diameter of every such path in G. It follows that for

$$R' < \frac{1}{2}\left(\frac{1}{L_{\boldsymbol{T}_{\gamma^{\otimes n}}^{(d)}}} - 1\right), \tag{3.11}$$

there are at least $(|\gamma|^{(d-1)})^h$ vertex-disjoint paths originating in $B_G(A,R)$ and leaving $B_G(A,R')$.

Finally, note that, by assumption, we have $\min(\gamma) \geqslant b$, and therefore $\min(\gamma^{\otimes n}) \geqslant b^n$, implying that

$$L_{\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}}} = b^{-n}.$$

Since $\operatorname{diam}(G) \leqslant (d+1)b^n$ by Lemma 3.19, the constraint (3.11) is implied by

$$R' < \frac{1}{2}\left(\frac{\operatorname{diam}(G)}{d+1} - 1\right).$$

We may assume that $\operatorname{diam}(G) > 6(d+1)$, otherwise the statement of the lemma is vaccuous (since we may choose the constant c sufficiently small depending on d), in which case this is implied by

$$R' \leqslant \frac{\operatorname{diam}(G)}{3(d+1)},$$

completing the proof. □

3.5 Sphere-packing representations

We will now prove Lemma 3.8 and Theorem 3.12. To this end, we require some regularity from our cube packings. Say that two closed, axis-parallel cubes $A, B \subseteq \mathbb{R}^d$ are *neatly tangent* if:

(i) A and B are interior-disjoint.

(ii) If A and B intersect along $(d-1)$-dimensional faces $F_A \subseteq A$ and $F_B \subseteq B$, then either $F_A \subseteq A \cap B$ or $F_B \subseteq A \cap B$.

A collection $\mathcal{C}$ of closed, axis-parallel cubes is *a neat cube packing* if every pair $A \neq B \in \mathcal{C}$ is either neatly tangent or else disjoint. The *aspect ratio of the packing* $\mathcal{C}$ is defined by

$$\alpha(\mathcal{C}) := \max\left\{\frac{\ell(A)}{\ell(B)} : A, B \in \mathcal{C}, A \cap B \neq \emptyset\right\},$$

where $\ell(A)$ denotes the sidelength of a cube A. Say that a graph G is *admits an α-uniform neat cube packing in* $\mathbb{R}^d$ if G is the tangency graph of a neat cube packing $\mathcal{C}$ with $\alpha(\mathcal{C}) \leqslant \alpha$.

Lemma 3.23. *If $\gamma = \langle \gamma_1, \ldots, \gamma_b \rangle$ satisfies $\gamma_1 = \gamma_b = b = \min(\gamma)$ and (3.2) holds, then the graphs $G(\boldsymbol{T}^{(d)}_\gamma)$ and $\hat{G}^{(d)}_\gamma$ admit an α-uniform neat cube packing in $\mathbb{R}^d$ with $\alpha = \alpha_{\boldsymbol{T}^{(d)}_\gamma}$.*

Proof. First take $\mathcal{C} := \tilde{\boldsymbol{T}}^{(d)}_\gamma$, as defined in Section 3.2.2. Under the integrality assumptions on γ, it holds that $\mathcal{C}$ is a neat packing, since one of the ratios γ_{i+1}/γ_i or γ_i/γ_{i+1} is an integer for every $1 \leqslant i < b$ (see Figure 8(b) for an illustration). Moreover, we have

$$\alpha(\mathcal{C}) = \alpha_{\boldsymbol{T}^{(d)}_\gamma} = \alpha_{\tilde{\boldsymbol{T}}^{(d)}_\gamma}.$$

For the second assertion, we take $C := \boldsymbol{T}^{(d)}_{\gamma^{\otimes\infty}}$ (as defined in Section 3.2.2). Corollary 3.18 asserts that

$$\alpha_{\boldsymbol{T}^{(d)}_{\gamma^{\otimes n}}} \leqslant \alpha_{\boldsymbol{T}^{(d)}_{\gamma}}, \quad \forall n \geqslant 0,$$

hence $\alpha_{\boldsymbol{T}^{(d)}_{\gamma^{\otimes\infty}}} < \infty$, as desired. □

Given the preceding lemma, the next result suffices to prove Lemma 3.8 and Theorem 3.12.

Lemma 3.24. *If G admits an α-uniform neat cube packing in $\mathbb{R}^d$, then there are numbers $m \leqslant O(\alpha d)$ and $M \leqslant O(\alpha^2 d)$ such that the subdivision $[G]_m$ is M-uniformly sphere-packed in $\mathbb{R}^d$.*

To prove this, we need a simple result on sphere packings that satisfy prescribed tangencies. For a general closed axis-parallel cube $C \subseteq \mathbb{R}^d$ and $\varepsilon > 0$, we define ∂C to denote the boundary of C in $\mathbb{R}^d$, and we use $\|\cdot\|$ for the standard Euclidean distance. For a point $x \in \mathbb{R}^d$ and a set $S \subseteq \mathbb{R}^d$, define $\mathrm{dist}_2(x, S) := \inf\{\|x - y\| : y \in S\}$. Let us also define

$$\partial_\varepsilon C := \{x \in \partial C : \#\{F \in \mathcal{F}_C : \mathrm{dist}_2(x, F) > \varepsilon\ell(C)\} = 2d - 1\},$$

where $\ell(C)$ is the sidelength of C, and $\mathcal{F}_C$ is the collection of the $2d$ facets of C (i.e., the $(d-1)$-dimensional faces of C). These are the boundary points that lie on exactly once face of C and are $\varepsilon\ell(C)$-far from every other.

Lemma 3.25. *For every $d \geqslant 2$ and $\varepsilon \in (0, 1/2)$, the following holds. Consider a closed axis-parallel cube $C \subseteq \mathbb{R}^d$ of sidelength ℓ, and a set of points $P \subseteq \partial_\varepsilon C$ such that $\|x - y\| \geqslant \varepsilon\ell$ for every $x \neq y \in P$. Then there is a finite collection of interior-disjoint spheres $\mathcal{S}$ contained in C such that:*

1. *The radius of every sphere in $\mathcal{S}$ is at least $\varepsilon\ell/(60d)$.*
2. *For every $p \in P$, there is a sphere $S(p) \in \mathcal{S}$ tangent to p.*
3. *The tangency graph of $\mathcal{S}$ is the m-subdivision of a star graph whose leaves are the spheres $\{S(p) : p \in P\}$, with $m = 2 + \lceil \frac{6d}{\varepsilon} \rceil$.*

Proof. By scaling and translation, it suffices to prove the lemma for $C = [0,1]^d$. Denote $c_0 := (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$, and define the sphere

$$S_0 := \{x \in \mathbb{R}^d : \|x - c_0\| = 1/4\}.$$

For each $p \in P$, define $S(p)$ to be the unique sphere of radius $\varepsilon/8$ that is contained in $[0,1]^d$ and such that $S(p) \cap \partial[0,1]^d = \{p\}$. Such a sphere exists because $p \in \partial_\varepsilon[0,1]^d$.

Let c_p denote the center of $S(p)$, and let $[c_0, c_p]$ denote the line segment from c_0 to c_p. Define z_p to be the point where $[c_0, c_p]$ intersects S_0, and define z'_p to be the point where $[c_0, c_p]$ intersects $S(p)$. Let $[z_p, z'_p] \subseteq [c_0, c_p]$ denote the line segment connecting z_p to z'_p.

As $\|x - y\| \geqslant \varepsilon$ for all $x \neq y \in P$, it holds that $\|z'_x - z'_y\| \geqslant \varepsilon - 4(\varepsilon/8) \geqslant \varepsilon/2$. Note that $[0,1]^d \subseteq B_2(c_0, \sqrt{d}/2)$, where the latter object is the Euclidean ball of radius $\sqrt{d}/2$ about c_0. Let $\tilde{z}_x$ denote the point where the line through $[c_0, c_p]$ intersects $\partial B_2(c_0, \sqrt{d}/2)$. Then we have

$\|\tilde{z}_x - \tilde{z}_y\| \geqslant \|z'_x - z'_y\|$, and by similarity of the triangles defined by $\{c_0, \tilde{z}_x, \tilde{z}_y\}$ and $\{c_0, z'_x, z'_y\}$, it holds that

$$\|z'_x - z'_y\| \geqslant \frac{\|\tilde{z}_x - \tilde{z}_y\|}{\sqrt{d}} \geqslant \frac{\varepsilon}{2\sqrt{d}}.$$

It follows that

$$\min\{\|a-b\| : a \in [z_x, z'_x], b \in [z_y, z'_y]\} \geqslant \|z_x - z_y\| \geqslant \frac{\varepsilon}{2\sqrt{d}}. \tag{3.12}$$

Note also that for every $p \in P$,

$$\frac{1}{8} \leqslant \frac{1}{4} - \frac{\varepsilon}{4} \leqslant \|c_0 - c_p\| - \left(\frac{1}{4} + \frac{\varepsilon}{8}\right) \leqslant \left\|z_p - z'_p\right\| \leqslant \|c_0 - c_p\| \leqslant \mathrm{diam}_2([0,1]^d) \leqslant \sqrt{d}, \tag{3.13}$$

where we have used $\varepsilon < 1/2$.

Let $\gamma_p : [0, \|z_p - z'_p\|] \to [z_p, z'_p]$ be a parameterization of $[z_p, z'_p]$ by arclength. Define $k := \lceil \frac{6d}{\varepsilon} \rceil$ and $r_p := \frac{\|z_p - z'_p\|}{2k}$, and let $\tilde{S}_p$ be the collecion of interior-disjoint spheres of radius r_p centered at the points

$$\gamma_p(r_p), \gamma_p(2r_p), \gamma_p(4r_p), \ldots, \gamma_p(2(k-1)r_p).$$

Note that the tangency graph of $\tilde{S}_p$ is a path, and that the first sphere is tangent to S_0, while the last is tangent to $S(p)$.

By (3.13), for each $p \in P$, we have $r_P \in \left[\frac{\varepsilon}{60d}, \frac{\varepsilon}{6\sqrt{d}}\right]$. In particular, (3.12) implies that if $S \in \tilde{S}_p$ and $S' \in \tilde{S}_{p'}$ for $p \neq p' \in P$, then S and S' are disjoint.

Thus the collection of spheres

$$\mathcal{S} := \{S_0\} \cup \{S(p) : p \in P\} \cup \bigcup_{p \in P} \tilde{S}_p$$

satisfies the conditions of the lemma. □

Proof of Lemma 3.24. Suppose that C is a neat cube packing whose tangency graph is G and such that $\alpha(C) \leqslant \alpha$. For every pair $A, B \in C$ with $A \cap B \neq \emptyset$, let $c(A, B)$ be the center of mass of $A \cap B$.

Define the set of points $P_A := \{c(A, B) : B \in C, A \cap B \neq \emptyset\}$. Since C is a neat cube packing and we have $\ell(B) \geqslant \ell(A)/\alpha$, it follows that with $\varepsilon := 1/(2\alpha)$, we have $P_A \subseteq \partial_\varepsilon A$, and

$$\|x - y\| \geqslant \varepsilon \ell(A), \quad \forall x \neq y \in P_A.$$

Therefore we can use Lemma 3.25 to replace each cube $A \in C$ by a corresponding collection $\mathcal{S}_A$ of spheres whose tangency graph is the m-subdivision of a star, the leaves of which are tangent to the points in P_A. Note that Lemma 3.25 gives $m = 2 + \lceil 12\alpha d \rceil$.

Moreover, any two adjacent spheres have their ratio of radii contained in the interval $[M^{-1}, M]$ for

$$M := \alpha \frac{60d}{\varepsilon} = 120\alpha^2 d.$$

Thus if we define $\mathcal{S} := \bigcup_{A \in C} \mathcal{S}_A$, then $[G]_{2m}$ is the tangency graph of $\mathcal{S}$, and the corresponding sphere-packing is M-uniform. □

4 Multiscale entropic regularization for MTS on general metric spaces

4.1 Introduction

Let (X, d) be a finite metric space with $|X| = n > 1$. The Metrical Task Systems (MTS) problem, introduced in [BLS92] is defined as follows. The input is a sequence $\langle c_t : X \to \mathbb{R}_+ \mid t = 1, 2, \dots \rangle$ of nonnegative cost functions on the state space X. At every time t, an online algorithm maintains a state $\rho_t \in X$.

The corresponding cost is the sum of a *service cost* $c_t(\rho_t)$ and a *movement cost* $d(\rho_{t-1}, \rho_t)$. Formally, an *online algorithm* is a sequence of mappings $\boldsymbol{\rho} = \langle \rho_1, \rho_2, \dots, \rangle$ where, for every $t \geqslant 1$, $\rho_t : (\mathbb{R}_+^X)^t \to X$ maps a sequence of cost functions $\langle c_1, \dots, c_t \rangle$ to a state. The initial state $\rho_0 \in X$ is fixed. The *total cost of the algorithm* $\boldsymbol{\rho}$ *in servicing* $\boldsymbol{c} = \langle c_t : t \geqslant 1 \rangle$ is defined as the sum of the service and movement costs:

$$\begin{aligned}
\mathrm{serv}_{\boldsymbol{\rho}}(\boldsymbol{c}) &:= \sum_{t \geqslant 1} c_t(\rho_t(c_1, \dots, c_t)) \\
\mathrm{move}_{\boldsymbol{\rho}}(\boldsymbol{c}) &:= \sum_{t \geqslant 1} d(\rho_{t-1}(c_1, \dots, c_{t-1}), \rho_t(c_1, \dots, c_t)) \\
\mathrm{cost}_{\boldsymbol{\rho}}(\boldsymbol{c}) &:= \mathrm{serv}_{\boldsymbol{\rho}}(\boldsymbol{c}) + \mathrm{move}_{\boldsymbol{\rho}}(\boldsymbol{c}).
\end{aligned}$$

The cost of the *offline optimum*, denoted $\mathrm{cost}^*(\boldsymbol{c})$, is the infimum of $\sum_{t \geqslant 1}[c_t(\rho_t) + d(\rho_{t-1}, \rho_t)]$ over *any* sequence $\langle \rho_t : t \geqslant 1 \rangle$ of states.

A *randomized online algorithm* $\boldsymbol{\rho}$ is said to be α*-competitive* if for every $\rho_0 \in X$, there is a constant $\beta > 0$ such that for all cost sequences $\boldsymbol{c}$:

$$\mathbb{E}\left[\mathrm{cost}_{\boldsymbol{\rho}}(\boldsymbol{c})\right] \leqslant \alpha \cdot \mathrm{cost}^*(\boldsymbol{c}) + \beta\,.$$

Such an algorithm is said to be α*-competitive for service costs and* α'*-competitive for movement costs* if there is a constant $\beta > 0$ such that for all cost sequences $\boldsymbol{c}$:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{serv}_{\boldsymbol{\rho}}(\boldsymbol{c})\right] &\leqslant \alpha \cdot \mathrm{cost}^*(\boldsymbol{c}) + \beta \\
\mathbb{E}\left[\mathrm{move}_{\boldsymbol{\rho}}(\boldsymbol{c})\right] &\leqslant \alpha' \cdot \mathrm{cost}^*(\boldsymbol{c}) + \beta.
\end{aligned}$$

For the n-point uniform metric, a simple coupon-collector argument shows that the competitive ratio is $\Omega(\log n)$, and this is tight [BLS92]. A long-standing conjecture is that this $\Theta(\log n)$ competitive ratio holds for an arbitrary n-point metric space. The lower bound has almost been established [BBM06, BLMN05]; for any n-point metric space, the competitive ratio is $\Omega(\log n / \log\log n)$. Following a long sequence of works (see, e.g., [Sei99, BKRS00, BBBT97, Bar96, FM03, FRT04]), an upper bound of $O((\log n)^2)$ was shown in [BCLL21].

Competitive analysis via gradient descent. Let us consider an equivalent fractional perspective on MTS where the online algorithm maintains, at every point in time, a probability distribution $\mu_t \in \mathbb{R}_+^X$, and we interpret the costs similarly as a vector $c_t \in \mathbb{R}_+^X$. The cost of the algorithm is then given by

$$\sum_{t \geqslant 1} \left(\langle \mu_t, c_t \rangle + \mathbb{W}_X^1(\mu_{t-1}, \mu_t) \right),$$

where $\mathbb{W}^1_X$ is the L^1 transportation cost between two probability distributions on (X, d). This perspective is convenient, as now the state of the algorithm is given by a point in the probability simplex $\Delta_X \subseteq \mathbb{R}^X_+$.

This yields a natural first algorithm for solving MTS:

$$\mu_{t+1} := \operatorname{proj}_{\Delta_X}(\mu_t - \eta c_t), \tag{4.1}$$

where $\eta > 0$ is some parameter we can choose and $\operatorname{proj}_{\Delta_X}$ denotes the Euclidean projection onto the convex body Δ_X. Moreover, it gives a natural way of relating the cost incurred by the algorithm to the cost incurred by *any other* state $\nu \in \Delta_X$: It is a basic exercise in convex geometry to show that

$$\|\mu_{t+1} - \nu\|^2 - \|\mu_t - \nu\|^2 \leqslant \eta\langle c_t, \nu - \mu_t\rangle. \tag{4.2}$$

In other words, if $\langle c_t, \mu_t\rangle > \langle c_t, \nu\rangle$, then μ_t approaches ν proportionally in the squared Euclidean distance.

Thus we cannot consistently incur more service cost than any fixed state. This does not provide a competitive algorithm because there is, in general, no convenient relationship between the Euclidean distance $\|\mu_t - \mu_{t+1}\|$ and the transportation distance $\mathbb{W}^1_X(\mu_t, \mu_{t+1})$.

But one can replace the Euclidean distance by any Bregman divergence $\mathbb{D}_\Phi$ associated to a strictly convex function Φ. Equivalently, we perform the projection (4.1) in the local inner product

$$\langle u, v\rangle_{\mu_t} := \langle \nabla^2\Phi(\mu_t)u, v\rangle.$$

Thus by choosing an appropriate geometry on Δ_X, one can hope to obtain a competitive algorithm. Such algorithms often go by the name *mirror descent* and the regularizer Φ is called the *mirror map* (we will often use the term *regularizer* interchangeably).

This framework is proposed in [ABBS10, BCN14] and applied to the k-server problem in [BCL+18], and to MTS in [BCLL21] and [CL19]. In all these papers, the algorithms apply only to ultrametrics (equivalently, to hierarchically separated tree metrics (HSTs)). In [BCLL21], mirror descent is used to analyze the algorithm on weighted stars, and these algorithms are glued together in an ad-hoc way to handle HSTs. In [CL19], stronger bounds (known as "refined guarantees") are obtained by finding an appropriate regularizer on arbitrary HSTs. In both cases, general finite metric spaces are then handled via random embeddings into HSTs.

In the present work, we apply this method directly to MTS on general metric spaces and match the best-known competitive ratio. Previously, it was unknown how to achieve any poly$(\log n)$ competitive ratio for general metric spaces using mirror descent and achieving this was posed as an open problem by Bubeck[3].

We consider this an important step in advancing the underlying philosophy. Note that past approaches to MTS have involved a series of ad-hoc, complicated algorithms, along with clever potential function analyses. In contrast, in the mirror descent approach, once one specifies a convex body and a regularizer, both the algorithm and the method of analysis fall out naturally. Indeed, the most subtle part of competitive analysis lies in connecting the cost an online algorithm incurs to the cost of some offline optimum, and this is done entirely through the general Bregman divergence analog of (4.2), which becomes

$$\mathbb{D}_\Phi(\nu \,\|\, \mu_{t+1}) - \mathbb{D}_\Phi(\nu \,\|\, \mu_t) \leqslant \langle c_t, \nu - \mu_t\rangle.$$

[3]Posed in his talk at HALG 2019.

4.2 The multiscale noisy metric entropy

To obtain poly($\log n$)-competitive algorithms for MTS, previous approaches [BCLL21, CL19] employ a regularizer that can be cast as a multiscale entropy for probability distributions on an underlying tree metric. To handle general metric spaces, we will consider probability distributions on a lifted convex body that is specified by a directed aylic graph whose sinks are the points of (X, d). See Figure 10 for a pictoral representation when the metric space is a path.

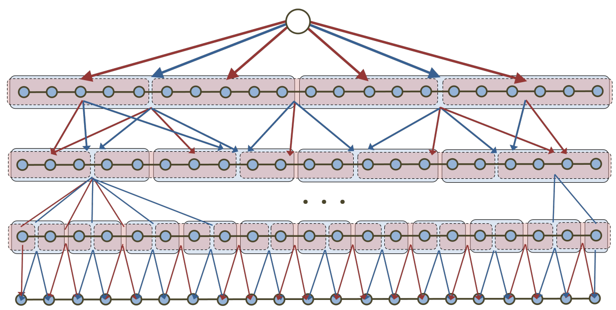


Figure 10: A hierarchical flow DAG over the path

The hierarchical flow DAG. Consider a finite set X and a directed aylic weighted graph $\mathcal{D} = (V, A)$ with $X \subseteq V$ and such that

(i) $\mathcal{D}$ has a single source $\mathbb{r} \in V$, and

(ii) The set of sinks in $\mathcal{D}$ is X.

We say that $\mathcal{D}$ *is a DAG over* X. In what follows, we use the notation $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geqslant 0\}$ and $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$. For an arc $(u, v) \in A$, we will often use the shorthand uv.

A vector $F \in \mathbb{R}_+^A$ is called a *flow in* $\mathcal{D}$ if holds that

$$\sum_{v : uv \in A} F_{uv} = \sum_{v : vu \in A} F_{vu}, \quad \forall u \in V \setminus (X \cup \{\mathbb{r}\}). \tag{4.3}$$

For a flow F and $u \in V \setminus X$, define $F_u := \sum_{v : uv \in A} F_{uv}$. For a sink $x \in X$, we define $F_x := \sum_{u : ux \in A} F_{ux}$ as the flow into x. Say that F is a *unit flow in* $\mathcal{D}$ if $F_{\mathbb{r}} = 1$, and let $\mathcal{F}_{\mathcal{D}} \subseteq \mathbb{R}_+^A$ denote the convex set of all unit flows in $\mathcal{D}$.

A *(directed) path* γ *in* $\mathcal{D}$ is a sequence $\gamma = \langle u_1 u_2, u_2 u_3, \ldots, u_{m-1} u_m \rangle$ with $u_i u_{i+1} \in A$ for each $i \in \{1, \ldots, m-1\}$. We will occasionally also specify a path as a sequence of vertices. We use $\bar{\gamma}$ to denote the final vertex u_m of γ. Let $\mathcal{P}_{\mathcal{D}}$ denote the set of all paths in $\mathcal{D}$ from $\mathbb{r}$ to some sink.

The multiscale entropy. Let $\omega \in \mathbb{R}_{++}^A$ denote a vector of nonnegative arc lengths that are decreasing along paths, i.e., such that $\omega_{uv} > \omega_{vw}$ whenever $uv, vw \in A$. Let $\theta \in \mathbb{R}_{++}^A$ specify a probability distribution on the edges leaving every vertex, i.e.,

$$\sum_{v : uv \in A} \theta_{uv} = 1, \qquad \forall u \in V \setminus X. \tag{4.4}$$

Define the associated values

$$\eta_{uv} := 1 + \log(1/\theta_{uv}) \tag{4.5}$$
$$\delta_{uv} := \theta_{uv}/\eta_{uv}. \tag{4.6}$$

We refer to the triple $\hat{\mathcal{D}} := (\mathcal{D}, \omega, \theta)$ as a *marked DAG*. For a given normalization parameter $\kappa > 0$, such a marked DAG yields a multiscale entropy functional $\Phi_{\hat{\mathcal{D}}} : \mathcal{F}_{\mathcal{D}} \to \mathbb{R}_+$ defined by

$$\Phi_{\hat{\mathcal{D}}}(F) := \frac{1}{\kappa} \sum_{uv \in A} \frac{\omega_{uv}}{\eta_{uv}} (F_{uv} + \delta_{uv} F_u) \log \left(\frac{F_{uv}}{F_u} + \delta_{uv} \right).$$

One can consult [CL19] for a detailed discussion of multiscale entropies of this form on HSTs.

Two notions of depth. We define two notions of depth associated to $\hat{\mathcal{D}}$. The first is the *combinatorial depth* $\Delta_0(\mathcal{D})$ which is the maximum number of arcs in any path from $\mathbb{r}$ to some sink X. For $\gamma \in \mathcal{P}_{\mathcal{D}}$, let us define

$$\theta(\gamma) := \prod_{uv \in \gamma} \theta_{uv}, \tag{4.7}$$

and let the *information depth* be defined as

$$\Delta_I(\hat{\mathcal{D}}) := \max_{\gamma \in \mathcal{P}_{\mathcal{D}}} \log(1/\theta(\gamma)).$$

Note that $\theta(\cdot)$ induces a probability distribution on $\mathcal{P}_{\mathcal{D}}$, and as clearly for $\gamma \in \mathcal{P}_{\mathcal{D}}$ it holds that $\theta(\gamma) \geqslant e^{-\Delta_I(\hat{\mathcal{D}})}$ we have

$$\log |\mathcal{P}_{\mathcal{D}}| \leqslant \Delta_I(\hat{\mathcal{D}}). \tag{4.8}$$

4.2.1 Mirror descent dynamics

Let us now fix a marked DAG $\hat{\mathcal{D}}$ and take $\Phi := \Phi_{\hat{\mathcal{D}}}$. We seek to define a continuous path $F : [0, \infty] \to \mathcal{F}_{\mathcal{D}}$ that represents the dynamics of projected vector flow in response to a continuous path $c(t) \in \mathbb{R}_+^X$ of costs arriving at the points of X.

A natural Euclidean flow would be specified heuristically by

$$F(t + dt) = \mathrm{proj}_{\mathcal{F}_{\mathcal{D}}} (F(t) - c(t)\, dt),$$

where for $v \in \mathbb{R}^A$, we define $\mathrm{proj}_{\mathcal{F}_{\mathcal{D}}}(v)$ as the unique point of $\mathcal{F}_{\mathcal{D}}$ with minimal Euclidean distance to v. In other words, we move a little in the direction $-c(t)$ and then project back to the feasible region $\mathcal{F}_{\mathcal{D}}$.

Instead, we will define our dynamics using the Bregman projection $\mathrm{proj}^{\Phi}_{\mathcal{F}_{\mathcal{D}}}$ associated to our multiscale entropic regularizer, where

$$\mathrm{proj}^{\Phi}_{\mathcal{F}_{\mathcal{D}}}(v) := \operatorname{argmin} \{ \mathbb{D}_{\Phi}(v' \,\|\, v) : v' \in \mathcal{F}_{\mathcal{D}} \},$$

and

$$\mathbb{D}_{\Phi}(v' \,\|\, v) := \Phi(v') - \Phi(v) - \langle \nabla\Phi(v), v' - v \rangle$$

is the Bregman divergence associated to Φ.

One can show that if $c(t)$ is continuous, then there is a path $F : [0, \infty) \to \mathcal{F}_{\mathcal{D}}$ for which the following dynamics are well-defined (for almost every $t \in [0, \infty)$):

$$F(t + dt) = \mathrm{proj}^{\Phi}_{\mathcal{F}_{\mathcal{D}}} (F(t) - c(t)\, dt)$$

This path further satisfies (for almost all $t \in [0, \infty)$) the system of partial differential equations given by

$$\partial_t \left(\frac{F_{uv}(t)}{F_u(t)}\right) = \kappa \frac{\eta_{uv}}{\omega_{uv}} \left(\frac{F_{uv}(t)}{F_u(t)} + \delta_{uv}\right) (\beta_u(t) - \hat{c}_{uv}(t)), \quad uv \in A, \tag{4.9}$$

where $\hat{c}_{uv}(t) = \mathbb{1}_{\{F_{uv}(t)>0\}} c_v(t)$ if $v \in X$, and otherwise

$$\hat{c}_{uv}(t) = \mathbb{1}_{\{F_{uv}(t)>0\}} \sum_{w : vw \in A} \frac{F_{vw}(t)}{F_v(t)} \hat{c}_{vw}(t), \tag{4.10}$$

and $\beta_u(t)$ is the unique value that guarantees

$$\partial_t \sum_{v : uv \in A} \frac{F_{uv}(t)}{F_u(t)} = 0,$$

i.e.,

$$\beta_u(t) = \frac{\sum_{v : uv \in A} \frac{\eta_{uv}}{\omega_{uv}} \left(\frac{F_{uv}(t)}{F_u(t)} + \delta_{uv}\right) \hat{c}_{uv}(t)}{\sum_{v : uv \in A} \frac{\eta_{uv}}{\omega_{uv}} \left(\frac{F_{uv}(t)}{F_u(t)} + \delta_{uv}\right)}.$$

Here we express the algorithm in continuous time for conceptual simplicity; its evolution is completely specified by the regularizer $\Phi_{\hat{\mathcal{D}}}$ and the costs $c(t)$. But the existence of a solution to (4.9) is derived from the limit of discrete-time algorithms in Section 4.4.

4.2.2 Metric compatibility

To analyze the algorithm specified by (4.9) on a metric space (X, d), we need additionally that $\hat{\mathcal{D}} = (\mathcal{D}, \omega, \theta)$ is compatible with the geometry of (X, d). Suppose that $\hat{\mathcal{D}}$ is a marked DAG over X. Say that *$\hat{\mathcal{D}}$ is τ-geometric* if it holds that for every pair of consecutive arcs $uv, vw \in A$, we have $\omega_{uv} \geqslant \tau \omega_{vw}$.

Let us define a metric on $\mathcal{P}_{\mathcal{D}}$ as follows: Suppose $\gamma_1, \gamma_2 \in \mathcal{P}_{\mathcal{D}}$ and let $u \in V$ be the first vertex at which they diverge, i.e., at which $uv_1 \in \gamma_1, uv_2 \in \gamma_2$ and $v_1 \neq v_2$. Define the distance

$$\mathrm{dist}_{\hat{\mathcal{D}}}(\gamma_1, \gamma_2) := \max(\omega_{uv_1}, \omega_{uv_2}).$$

One can check that this gives a metric on $\mathcal{P}_{\mathcal{D}}$ since the arc lengths are decreasing along source-sink paths. In fact, this defines an ultrametric on $\mathcal{P}_{\mathcal{D}}$.

Say that $\hat{\mathcal{D}}$ is *ε-expanding (with respect to (X, d))* if for every pair $\gamma_1, \gamma_2 \in \mathcal{P}_{\mathcal{D}}$, it holds that

$$\mathrm{dist}_{\hat{\mathcal{D}}}(\gamma_1, \gamma_2) \geqslant \varepsilon d(\bar{\gamma}_1, \bar{\gamma}_2),$$

where we recall that $\bar{\gamma}_1, \bar{\gamma}_2 \in X$ are the endpoints of γ_1 and γ_2, respectively.

We may extend $\mathrm{dist}_{\hat{\mathcal{D}}}$ to a distance on $\mathcal{F}_{\mathcal{D}}$ by defining $\mathbb{W}^1_{\hat{\mathcal{D}}}(F, F')$ as the L^1-transportation cost between $F, F' \in \mathcal{F}_{\mathcal{D}}$ with the underlying metric $\mathrm{dist}_{\hat{\mathcal{D}}}$, noting that F and F' can be viewed as probability distributions on $\mathcal{P}_{\mathcal{D}}$.

Say that $\hat{\mathcal{D}}$ *is L-Lipschitz (with respect to* (X, d)*)* if for every path $x_1, x_2, \ldots, x_m \in X$, there is a sequence of flows $F^{(1)}, F^{(2)}, \ldots, F^{(m)} \in \mathcal{F}_{\mathcal{D}}$ such that:

1. $F^{(i)}$ is a unit flow to x_i for every $i = 1, 2, \ldots, m$.
2. It holds that
$$\sum_{i=1}^{m-1} \mathbb{W}^1_{\hat{\mathcal{D}}}(F^{(i)}, F^{(i+1)}) \leqslant L \sum_{i=1}^{m-1} d(x_i, x_{i+1}).$$

Our main result follows from the next two theorems, which are proved in Section 4.4 and Section 4.3, respectively.

Theorem 4.1. *Suppose* (X, d) *is a metric space and* $\hat{\mathcal{D}}$ *is a* τ*-geometric marked DAG over* X*, for some* $\tau \geqslant 4$*. If* $\hat{\mathcal{D}}$ *is* ε*-expanding and L-Lipschitz with respect to* (X, d)*, then for* $\kappa = 6L$*, the MTS algorithm specified by* (4.9) *is* 1*-competitive for service costs, and* $O\left(\frac{L}{\varepsilon}\left(\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}})\right)\right)$*-competitive for movement costs.*

Theorem 4.2. *For every* n*-point metric space* (X, d)*, there is a* 12*-geometric marked DAG* $\hat{\mathcal{D}}$ *over* X *that is* 1*-expanding and* $O(\log n)$*-Lipschitz, and moreover satisfies*

$$\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}}) \leqslant O(\log n).$$

4.3 Construction of a compatible DAG over (X, d)

In Section 4.3.1, we present the main construction of a marked DAG $\hat{\mathcal{D}}$ whose vertices are net points at every scale. Achieving the crucial property $\Delta_I(\hat{\mathcal{D}}) \leqslant O(\log n)$ requires choosing the net points and the arcs of $\mathcal{D}$ carefully. In Section 4.3.2, we argue that $\hat{\mathcal{D}}$ is ε-expanding and L-Lipschitz for $\varepsilon = 1$ and $L \leqslant O(\log n)$. It may not be that $\Delta_0(\mathcal{D}) \leqslant O(\log n)$, but in Section 4.3.3 we give a generic way of obtaining this property while leaving the other essential properties intact.

4.3.1 Hierarchical nets

Fix an n-point metric space (X, d) and assume, without loss of generality, that $\mathrm{diam}(X) = 1$. Define $\varepsilon := \min\{d(x, y) : x, y \in X\}$ and $K := 1 + \lceil \log_\tau(1/\varepsilon) \rceil$.

Construction of nets. Consider a parameter $\eta > 0$. We construct an η-net $N \subseteq X$ inductively as follows. Define $N_0 := \emptyset$ and for $j \geqslant 1$, inductively define the set

$$S_j := X \setminus B_X(N_{j-1}, \eta).$$

If $S_j = \emptyset$, then we take $N := N_{j-1}$. Otherwise, let $x_j \in S_j$ be a point that maximizes $|B_X(x, \eta/3)|$ among $x \in S_j$ and define $N_j := N_{j-1} \cup \{x_j\}$.

Lemma 4.3. *The set $N \subseteq X$ is an η-net with the property that for any set $W \subseteq X$, if*

$$x^* \in \operatorname{argmax}\{|B_X(y, \eta/3)| : y \in N \cap B_X(W, 1.5\eta)\},$$

then

$$|B_X(x^*, \eta/3)| \geqslant \max\{|B_X(w, \eta/3)| : w \in W\}$$

Proof. Suppose $x_j \in N$ is the element with j minimal such that $B_X(x_j, 1.5\eta) \cap W \neq \emptyset$. Then

$$\big(B_X(x_1, \eta) \cup \cdots \cup B_X(Q_{j-1}, \eta)\big) \cap B_X(W, \eta/3) = \emptyset,$$

and hence by the greedy selection procedure,

$$|B_X(x_j, \eta/3)| = |B_X(x^*, \eta/3)|$$
$$|B_X(x_j, \eta/3)| \geqslant \max\{|B_X(w, \eta/3)| : w \in W\},$$

completing the proof. □

Denote $\tau := 12$. For each $k \in \{0, 1, \ldots, K\}$, let U_k denote a τ^{-k}-net that satisfies Lemma 4.3 with $\eta = \tau^{-k}$. We now construct a DAG $\mathcal{D} = (V, A)$ with $V := \{(u, k) : u \in U_k, k \in \{0, 1, \ldots, K\}\}$. For $k \in \{0, 1, \ldots, K-1\}$, let A_k denote the collection of pairs (u, u') for every $u \in U_k$ and $u' \in U_{k+1}$ satisfying:

$$d(u, u') \leqslant 4\tau^{-k} \tag{4.11}$$

$$|B_X(u, \tau^{-k}/3)| \geqslant \max\left\{|B_X(w, \tau^{-k}/3)| : w \in B_X(u', 6\tau^{-(k+1)})\right\}. \tag{4.12}$$

We define $A := \bigcup_{k=0}^{K-1} \{((u, k), (u', k+1)) : (u, u') \in A_k\}$, and

$$\omega_{(u,k)(u',k+1)} := 10\tau^{-k}.$$

Since $U_K = X$, we can identify the sinks in $\mathcal{D}$ with the points of X. We take $\mathbb{r} := (u, 0)$, where $U_0 = \{u\}$.

Observation 4.4. Suppose that $(u, k) \in U_k$ and $(x, K) \in V$ is reachable in $\mathcal{D}$ from (u, k). Then

$$d(u, x) \leqslant 4\tau^{-k} + 4\tau^{-(k+1)} + \cdots + 4\tau^{-(K-1)} < 5\tau^{-k}.$$

For a set $S \subseteq X$ and $k \in \{0, 1, \ldots, K\}$, define

$$\varphi_k(S) := \operatorname{argmax}\left\{|B_X(y, \tau^{-k}/3)| : y \in B_X(S, 2\tau^{-k}) \cap U_k\right\}. \tag{4.13}$$

We will require the following fact later.

Lemma 4.5. *Consider a set $S \subseteq X$ with $\operatorname{diam}_X(S) \leqslant 2\tau^{-k}$. If $u' \in S \cap U_{k+1}$, then $(\varphi_k(S), u') \in A_k$.*

Proof. Denote $u := \varphi_k(S)$. Since $u' \in S$ and $u \in B_X(S, 2\tau^{-k})$, it holds that $d(u, u') \leqslant 4\tau^{-k}$, and therefore (4.11) is satisfied. Now denote $W := B_X(u', 6\tau^{-(k+1)})$. Then $B_X(W, 1.5\tau^{-k}) \subseteq B_X(S, 2\tau^{-k})$, hence Lemma 4.3 implies that

$$|B_X(u, \tau^{-k}/3)| \geqslant \max\left\{|B_X(w, \tau^{-k}/3)| : w \in W\right\},$$

which shows that (4.12) is satisfied as well. □

For $k \in \{0, 1, \ldots, K-1\}$ and $(u, u') \in A_k$, we define

$$\theta_{(u,k),(u',k+1)} := \frac{|B_X(u', \tau^{-(k+1)}/3)|}{\sum_{w:(u,w)\in A_k} |B_X(w, \tau^{-(k+1)}/3)|}. \tag{4.14}$$

Claim 4.6. It holds that

$$\sum_{w:(u,w)\in A_k} |B_X(w, \tau^{-(k+1)}/3)| \leqslant |B_X(u, 6\tau^{-k})|.$$

Proof. Since the elements of U_{k+1} form a $\tau^{-(k+1)}$-net, the balls $\left\{B_X(w, \tau^{-(k+1)}/3) : (u, w) \in A_k\right\}$ are pairwise disjoint. Furthermore, by Observation 4.4, every such ball is contained in

$$B_X(u, 5\tau^{-k} + \tau^{-(k+1)}/3) \subseteq B_X(u, 6\tau^{-k}).$$

□

Lemma 4.7. *It holds that $\Delta_I(\mathcal{D}) \leqslant 3 \log n$, i.e., for every path $\gamma \in \mathcal{P}_{\mathcal{D}}$,*

$$\sum_{uv\in\gamma} \log(1/\theta_{u,v}) \leqslant 3 \log n.$$

Proof. Consider a path $\gamma = \langle (u_0, 0), (u_1, 1), \ldots, (u_K, K) \rangle$. From the definition (4.14) and Claim 4.6, it holds that

$$\sum_{k=0}^{K-1} \log\left(1/\theta_{(u_k,k)(u_{k+1},k+1)}\right) \leqslant \sum_{k=0}^{K-1} \log \frac{|B_X(u_k, 6\tau^{-k})|}{|B_X(u_{k+1}, \tau^{-(k+1)}/3)|}. \tag{4.15}$$

Let us denote $\ell := u_K$. By Observation 4.4, it holds that $d(\ell, u_k) \leqslant 6\tau^{-k}$ for $0 \leqslant k \leqslant K$. Therefore,

$$B_X(u_k, 6\tau^{-k}) \subseteq B_X(\ell, 12\tau^{-k}). \tag{4.16}$$

Furthermore since $(u_k, u_{k+1}) \in A_k$, by (4.12), we have

$$|B_X(u_k, \tau^{-k}/3)| \geqslant \max\left\{|B_X(w, \tau^{-k}/3)| : w \in B_X(u_{k+1}, 6\tau^{-(k+1)})\right\} \geqslant |B_X(\ell, \tau^{-k}/3)|, \tag{4.17}$$

since $d(\ell, u_{k+1}) \leqslant 6\tau^{-(k+1)}$.

By combining (4.15)–(4.17), we obtain

$$\sum_{k=0}^{K-1} \log\left(1/\theta_{(u_k,k)(u_{k+1},k+1)}\right) \leqslant \sum_{k=0}^{K-1} \log \frac{|B_X(\ell, 12\tau^{-k})|}{|B_X(\ell, \tau^{-(k+1)}/3)|} \leqslant \sum_{k=0}^{K-1} \log \frac{|B_X(\ell, \tau^{-(k-1)})|}{|B_X(\ell, \tau^{-(k+2)})|} \leqslant 3 \log n,$$

where we used $\tau = 12$ in the penultimate inequality. □

The above result together with (4.8) yield the following.

Corollary 4.8. *It holds that $|\mathcal{P}_{\mathcal{D}}| \leqslant n^3$.*

4.3.2 Distortion analysis

Lemma 4.9. *It holds that $\hat{\mathcal{D}}$ is 1-expanding with respect to (X, d).*

Proof. Suppose that $\gamma_1, \gamma_2 \in \mathcal{P}_{\mathcal{D}}$ and let $u \in V$ be the first vertex for which $uv_1 \in \gamma_1$ and $uv_2 \in \gamma_2$ with $v_1 \neq v_2$. If $u = (x, k)$, then $\omega_{uv_1} = \omega_{uv_2} = 10\tau^{-k}$ and so $\mathrm{dist}_{\hat{\mathcal{D}}}(\gamma_1, \gamma_2) = 10\tau^{-k}$. Moreover, by Observation 4.4 we have

$$d(\bar{\gamma}_1, \bar{\gamma}_2) \leqslant d(\bar{\gamma}_1, x) + d(\bar{\gamma}_2, x) \leqslant 10\tau^{-k},$$

completing the proof. ◻

For a partition P of X and $x \in X$, we let $P(x)$ denote the unique set in P containing X. We will require the following well-known random partitioning lemma.

Theorem 4.10 ([CKR01]). *For any finite metric space (X, d) and value $\Delta > 0$, there is a random partition P of X such that:*

1. $\mathrm{diam}_X(S) \leqslant \Delta$ *for every $S \in P$.*
2. *For all $x, y \in X$, it holds that*

$$\mathbb{P}\left[P(x) \neq P(y)\right] \leqslant 8\frac{d(x,y)}{\Delta} \log \frac{|B(x,\Delta)|}{|B(x,\Delta/8)|}.$$

For each $k \in \{0, 1, \ldots, K\}$, let P_k be a random partition of X satisfying the conclusion of Theorem 4.10 with $\Delta = \tau^{-k}$. Define a random map $\psi_k : X \to U_k$ as follows:

$$\psi_k(x) := \varphi_k(B_X(P_k(x), \tau^{-k}/2)),$$

where φ_k is the map defined in (4.13).

Lemma 4.11. *For every $x \in X$, it holds that $\langle(\psi_0(x), 0), (\psi_1(x), 1), \ldots, (\psi_K(x), K)\rangle$ is a path in $\mathcal{D}$.*

Proof. It suffices to show that for any $k \in \{0, 1, \ldots, K-1\}$, we have $(\psi_k(x), \psi_{k+1}(x)) \in A_k$. Define $u' = \psi_{k+1}(x)$ and $S := B_X(P_k(x), \tau^{-k}/2)$. Then $\mathrm{diam}_X(S) \leqslant 2\tau^{-k}$ and

$$d(x, u') = d(x, \psi_{k+1}(x)) \leqslant 2\tau^{-(k+1)} + \mathrm{diam}_X(B_X(P_{k+1}(x), \tau^{-(k+1)}/2)) \leqslant 4\tau^{-(k+1)} < \tau^{-k}/2,$$

where the last inequality follows from $\tau = 12$. Hence $u' \in S \cap U_{k+1}$. We can therefore apply Lemma 4.5 to conclude that $(\psi_k(x), u') = (\varphi_k(S), u') \in A_k$, completing the proof. ◻

For $x \in X$, define $\Psi(x) := \langle(\psi_0(x), 0), (\psi_1(x), 1), \ldots, (\psi_K(x), K)\rangle$. From the preceding lemma, we know that $\Psi : X \to \mathcal{P}_{\mathcal{D}}$.

Lemma 4.12. *For any $x, y \in X$, it holds that*

$$\mathbb{E}\left[\mathrm{dist}_{\hat{\mathcal{D}}}(\Psi(x), \Psi(y))\right] \leqslant O(\log n)\, d(x, y).$$

Proof. From Theorem 4.10, we have

$$\begin{aligned}\mathbb{E}\left[\mathrm{dist}_{\hat{\mathcal{D}}}(\Psi(x),\Psi(y))\right] &\leqslant \sum_{k=0}^{K} \mathbb{P}[P_k(x)\neq P_k(y)]\cdot 10\tau^{-k}\\ &\leqslant 80\, d(x,y)\sum_{k=0}^{K}\log\frac{|B(x,\tau^{-k})|}{|B(x,\tau^{-k}/8)|}\\ &\leqslant 80\log(n)\, d(x,y),\end{aligned}$$

where in the last line we used $\tau = 12 \geqslant 8$. □

Corollary 4.13. *It holds that $\hat{\mathcal{D}}$ is $O(\log n)$-Lipschitz with respect to (X,d).*

Proof. Consider any sequence $x_1,\dots,x_m$, and let us map it to the random sequence $\Psi(x_1),\dots,\Psi(x_m)$. Then from Lemma 4.12, we conclude

$$\sum_{j=1}^{m-1}\mathbb{E}\left[\mathrm{dist}_{\hat{\mathcal{D}}}(\Psi(x_j),\Psi(x_{j+1}))\right] \leqslant O(\log n)\sum_{j=1}^{m-1} d(x_j,x_{j+1}).$$

Hence there is a mapping $f : X \to \mathcal{P}_{\mathcal{D}}$ (that depends on the sequence $x_1,\dots,x_m$) such that $\sum_{j=1}^{m-1} d(f(x_j),f(x_{j+1})) \leqslant O(\log n)\sum_{j=1}^{m-1} d(x_j,x_{j+1})$, completing the proof. □

4.3.3 Compression

Let $\hat{\mathcal{D}} = (\mathcal{D},\omega,\theta)$ be the τ-geometric marked DAG constructed in Section 4.3.1. For a point $u \in V$, we let $\sigma(u)$ denote the number of paths in $\mathcal{D}$ that start at u and end in a point of X.

Observation 4.14. For $u \in V \setminus X$, it holds that

$$\sigma(u) = \sum_{v:uv\in A}\sigma(v). \tag{4.18}$$

Say an edge $uv \in A$ is *heavy* if $v \notin X$ and $\sigma(v) > \sigma(u)/2$; otherwise we say that uv *is light*. Moreover, we say a path $\gamma = \langle u_1u_2, u_2u_3,\dots,u_{m-1}u_m\rangle$ in $\mathcal{D}$ is *heavy-light* if all the edges $u_1u_2, u_2u_3,\dots,u_{m-2}u_{m-1}$ are heavy and $u_{m-1}u_m$ is light. The next lemma is straightforward and follows from (4.18).

Lemma 4.15. *For every $u \in V$, there is at most one heavy edge in $\mathcal{D}$ leaving u.*

Now we construct the marked DAG $\tilde{\mathcal{D}} = (\mathcal{D}',\omega',\theta')$ with $\mathcal{D}' = (V,A')$ as follows. We connect $u_i = (x_i,i) \in V$ to $u_j = (x_j,j) \in V$ for $1 \leqslant i < j \leqslant K$ in $\mathcal{D}'$ if there is a heavy-light path $\gamma = \langle u_iu_{i+1}, u_{i+1}u_{i+2},\dots,u_{j-1}u_j\rangle$ from u_i to u_j in $\mathcal{D}$. Note that by Lemma 4.15, at most one such path can exist. We further set

$$\begin{aligned}\omega'_{u_iu_j} &:= 10\tau^{-j+1},\\ \theta'_{u_iu_j} &:= \prod_{k=i}^{j-1}\theta_{u_ku_{k+1}}.\end{aligned}$$

Lemma 4.16. *For $uv \in A'$ with $v \notin X$ it holds that*

$$\sigma(v) \leqslant \sigma(u)/2.$$

Proof. Since $uv \in A'$, there must be a heavy-light path $\gamma = \langle w_1 w_2, \ldots, w_{m-1} w_m \rangle$ in $\mathcal{D}$ with $w_1 = u$ and $w_m = v$. Clearly the values of $\sigma(\cdot)$ are non-increasing along the (directed) paths in $\mathcal{D}$, hence we have

$$\sigma(u) \geqslant \sigma(w_2) \geqslant \cdots \geqslant \sigma(w_{m-1}).$$

Furthermore, as $w_{m-1}v$ is a light edge and $v \notin X$, it follows that

$$\sigma(v) \leqslant \sigma(w_{m-1})/2 \leqslant \sigma(u)/2,$$

as desired. □

Lemma 4.17. *It holds that $\Delta_0(\mathcal{D}') \leqslant O(\log |\mathcal{P}_{\mathcal{D}}|)$.*

Proof. We will argue that for every $\gamma \in \mathcal{P}_{\mathcal{D}'}$, one has $|\gamma| = O(\log |\mathcal{P}_{\mathcal{D}}|)$. Let $\gamma = \langle u_1 u_2, \ldots, u_{m-1} u_m \rangle$. Lemma 4.16 implies that for $1 \leqslant i \leqslant m-2$ we have $\sigma(u_{i+1}) \leqslant \sigma(u_i)/2$. Further note that we have $\sigma(u_1) = |\mathcal{P}_{\mathcal{D}}|$, and also clearly $\sigma(u_{m-1}) \geqslant 1$. Therefore,

$$m - 2 \leqslant \log_2(|\mathcal{P}_{\mathcal{D}}|),$$

completing the proof. □

We now define the map $f : \mathcal{P}_{\mathcal{D}} \to \mathcal{P}_{\mathcal{D}'}$ as follows. For a path $\gamma \in \mathcal{P}_{\mathcal{D}}$, let $f(\gamma)$ denote the path obtained by contracting all the heavy edges in γ. More precisely, for $\gamma = \langle u_1 u_2, u_2 u_3, \ldots, u_{m-1} u_m \rangle$, we define $f(\gamma)$ as follows. Denote $i_0 := 1$, and for $j = 1, 2, \ldots, m'$, let i_j denote the jth index for which $u_{i_j - 1} u_{i_j}$ is a light edge. We then denote

$$f(\gamma) := \langle u_{i_0} u_{i_1}, u_{i_1} u_{i_2}, \ldots, u_{i_{m'-1}} u_{i_{m'}} \rangle.$$

Lemma 4.18. *It holds that $\Delta_I(\tilde{D}) \leqslant \Delta_I(\hat{D})$.*

Proof. As all the edges in $\mathcal{D}'$ correspond to a path in $\mathcal{D}$, f is a surjective map. Furthermore, for $\gamma \in \mathcal{P}_{\mathcal{D}}$, one has $\theta(\gamma) = \theta'(f(\gamma))$, for $\theta(\cdot)$ defined as in (4.7) and $\theta'(\cdot)$ defined analogously, and thus we have

$$\Delta_I(\tilde{\mathcal{D}}) = \max_{\gamma' \in \mathcal{P}_{\mathcal{D}'}} \log(1/\theta'(\gamma')) = \max_{\gamma \in \mathcal{P}_{\mathcal{D}}} \log(1/\theta'(f(\gamma))) \leqslant \max_{\gamma \in \mathcal{P}_{\mathcal{D}}} \log(1/\theta(\gamma)) = \Delta_I(\hat{D}),$$

completing the proof. □

Lemma 4.19. *For all $\gamma, \gamma' \in \mathcal{P}_{\mathcal{D}}$ it holds that*

$$\mathrm{dist}_{\hat{\mathcal{D}}}(\gamma, \gamma') = \mathrm{dist}_{\tilde{\mathcal{D}}}(f(\gamma), f(\gamma')).$$

Proof. Denote $\gamma = \langle u_1u_2, \dots, u_{m-1}u_m\rangle$ and $\gamma' = \langle u'_1u'_2, \dots, u'_{m-1}u'_m\rangle$, and let $u_i = u'_i$ be the first vertex at which γ and γ' diverge so that we have

$$\mathrm{dist}_{\hat{\mathcal{D}}}(P, P') = \max(\omega_{u_iu_{i+1}}, \omega_{u_iu'_{i+1}}) = 10\tau^{-i}.$$

By Lemma 4.15, at most one of u_iu_{i+1} and $u_iu'_{i+1}$ can be heavy. Suppose that u_iu_{i+1} is light. Take $j := 1$ when $i = 1$, and otherwise let $j \leqslant i$ be the maximum index for which $u_{j-1}u_j$ is light. Further let $k \geqslant i$ be the minimum index for which $u'_ku'_{k+1}$ is a light edge. Note that k is well-defined because $u'_{m-1}u'_m$ is light. Now we have

$$\begin{aligned}\mathrm{dist}_{\tilde{\mathcal{D}}}(f(\gamma), f(\gamma')) &= \max(\omega'_{u_ju_{i+1}}, \omega'_{u_ju'_{k+1}})\\ &= \max(\omega_{u_iu_{i+1}}, \omega_{u'_ku'_{k+1}}) = \max(10\tau^{-i}, 10\tau^{-k}) = 10\tau^{-i},\end{aligned}$$

as desired. □

Lemma 4.20. *The $\tilde{D}$ is a marked DAG that is also τ-geometric.*

Proof. We first establish the τ-geometric property. Consider $u, v, w \in V$ with $uv, vw \in A'$. Denote $v = (x, i)$ for some $1 \leqslant i \leqslant K$. Then by construction, we have $\omega'_{uv} \geqslant 10\tau^{-i+1}$ and $\omega'_{vw} \leqslant 10\tau^{-i}$, completing the proof.

Next, we show that $\tilde{D}$ is a properly-constructed marked DAG. We need to establish that for $u \in V \setminus X$ it holds that

$$\sum_{v:uv\in A'} \theta'_{uv} = 1. \tag{4.19}$$

Let $v_0 := u$ and let $\gamma = \langle uv_1, v_1v_2, \dots, v_{k-1}v_k\rangle$ be the maximal heavy path going out of u for some $k \geqslant 0$, meaning that all the edges v_iv_{i+1} are heavy for $0 \leqslant i \leqslant k-1$. Lemma 4.15 implies that the choice of γ is unique.

Now using (4.19), write

$$\begin{aligned}\sum_{v\,:\,uv\in A'} \theta'_{uv} &= \sum_{j=0}^{k-1} \sum_{\substack{y\neq v_{j+1}:\\ v_jy\in A}} \theta_{v_jy} \cdot \prod_{\ell=0}^{j-1} \theta_{v_\ell v_{\ell+1}} + \prod_{\ell=0}^{k-1} \theta_{v_\ell v_{\ell+1}} \cdot \left(\sum_{y\,:\,v_ky\in A} \theta_{v_ky}\right)\\ &= \sum_{j=0}^{k-1} \sum_{\substack{y\neq v_{j+1}:\\ v_jy\in A}} \theta_{v_jy} \cdot \prod_{\ell=0}^{j-1} \theta_{v_\ell v_{\ell+1}} + \prod_{\ell=0}^{k-1} \theta_{v_\ell v_{\ell+1}}\\ &= \sum_{j=0}^{k-2} \sum_{\substack{y\neq v_{j+1}:\\ v_jy\in A}} \theta_{v_jy} \cdot \prod_{\ell=0}^{j-1} \theta_{v_\ell v_{\ell+1}} + \prod_{\ell=0}^{k-2} \theta_{v_\ell v_{\ell+1}} \cdot \left(\sum_{y\,:\,v_{k-1}y\in A} \theta_{v_{k-1}y}\right)\\ &\;\;\vdots\\ &= \sum_{\substack{y\neq v_1:\\ v_0y\in A}} \theta_{v_0y} + \theta_{v_0v_1} \cdot \left(\sum_{y\,:\,v_1y\in A} \theta_{v_1y}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{y\,:\,v_0 y \in A} \theta_{v_0 y} \\
&= 1,
\end{aligned}
$$

as desired. ◻

We are now ready to prove Theorem 4.2.

Proof of Theorem 4.2. Let us show that $\tilde{\mathcal{D}}$ satisfies the requirements of the theorem. Lemma 4.20 shows that $\tilde{\mathcal{D}}$ is a 12-geometric marked DAG.

Now note that for $\gamma \in \mathcal{P}_{\mathcal{D}}$ we have $\overline{f(\gamma)} = \bar{\gamma}$, and thus Lemma 4.19 in conjunction with Lemma 4.9 and Corollary 4.13 implies that $\tilde{D}$ is 1-expanding and $O(\log n)$-Lipschitz. Moreover, Lemma 4.18 together with Lemma 4.7 bounds the information depth of $\tilde{D}$. Finally, a bound on the combinatorial depth follows from Lemma 4.17 and Corollary 4.8. ◻

4.4 Algorithm and competitive analysis

4.4.1 Discrete-time algorithm

Let $\hat{\mathcal{D}} = (\mathcal{D}, w, \theta)$ be a marked DAG on X with $\mathcal{D} = (V, A)$. We now describe a generalization of the discrete-time dynamics of [CL19] on $\hat{\mathcal{D}}$ in response to a sequence of costs $\langle c_t : t \geqslant 1\rangle$, where $c_t \in \mathbb{R}_+^X$. Define

$$
Q_{\mathcal{D}} := \left\{ p \in \mathbb{R}_+^A \,\middle|\, \forall u \in V \setminus X : \sum_{v\,:\,uv \in A} p_{uv} = 1 \right\}.
$$

For $q \in Q_{\mathcal{D}}$ and $u \in V \setminus X$, we use $q^{(u)}$ to denote the restriction of q to the subspace spanned by subset of standard basis vectors $\{e_{uv} : uv \in A\}$, and we define the corresponding probability simplex $Q_{\mathcal{D}}^{(u)} := \{q^{(u)} : q \in Q_{\mathcal{D}}\}$. For convenience, we use $q_v^{(u)}$ for $q_{uv}^{(u)}$.

Let $\kappa > 0$ be a normalization parameter, and let the values η_{uv} and δ_{uv} be defined as in (4.5) and (4.6). For $u \in V \setminus X$ and $p \in Q_{\mathcal{D}}^{(u)}$, define

$$
\Phi^{(u)}(p) := \frac{1}{\kappa} \sum_{v\,:\,uv \in A} \frac{\omega_{uv}}{\eta_{uv}} (p_{uv} + \delta_{uv}) \log(p_{uv} + \delta_{uv}),
$$

and for $p' \in Q_{\mathcal{D}}^{(u)}$, denote

$$
\mathbb{D}^{(u)}(p \,\|\, p') := \mathbb{D}_{\Phi^{(u)}}(p \,\|\, p') = \frac{1}{\kappa} \sum_{v\,:\,uv \in A} \frac{\omega_{uv}}{\eta_{uv}} \left[(p_{uv} + \delta_{uv}) \log \frac{p_{uv} + \delta_{uv}}{p'_{uv} + \delta_{uv}} + p'_{uv} - p_{uv} \right].
$$

We now define an algorithm that takes a point $q \in Q_{\mathcal{D}}$ and a cost vector $c \in \mathbb{R}_+^X$ and outputs a point $p = \mathcal{A}(q, c) \in Q_{\mathcal{D}}$. Fix a topological ordering $\langle u_1, u_2, \ldots, u_N \rangle$ of $V \setminus X$ in $\mathcal{D}$. We define p inductively as follows. Denote $\hat{c}_x := c_x$ for $x \in X$, and for each $j = 1, 2, \ldots, N$:

$$
\hat{c}_v^{(u_j)} := \hat{c}_v \qquad \forall v : (u_j, v) \in A \tag{4.20}
$$

$$p^{(u_j)} := \operatorname{argmin}\left\{\mathbb{D}^{(u_j)}\left(p \,\|\, q^{(u_j)}\right) + \left\langle p, \hat{c}^{(u_j)}\right\rangle \;\middle|\; p \in Q_{\mathcal{D}}^{(u_j)}\right\} \tag{4.21}$$

$$\hat{c}_{u_j} := \sum_{v \,:\, u_j v \in A} p_v^{(u_j)}\, \hat{c}_v \tag{4.22}$$

We will use $\Lambda^{\mathcal{D}} : Q_{\mathcal{D}} \to \mathcal{F}_{\mathcal{D}}$ for the map which sends $q \in Q_{\mathcal{D}}$ to the (unique) $F = \Lambda^{\mathcal{D}}(q) \in \mathcal{F}_{\mathcal{D}}$ such that

$$F_{uv} = F_u q_{uv} \qquad \forall uv \in A.$$

Note that q contains more information than F; the map $\Lambda^{\mathcal{D}}$ fails to be invertible at $F \in \mathcal{F}_{\mathcal{D}}$ whenever there is some $u \in V \setminus X$ with $F_u = 0$. We will drop the superscript $\mathcal{D}$ from $\Lambda^{\mathcal{D}}$ whenever it is clear from context.

Now let p_0 be an arbitrary point in $Q_{\mathcal{D}}$. Given the cost sequence $\langle c_t : t \geqslant 1\rangle$, for $t = 1, 2, \ldots$ we define

$$p_t := \mathcal{A}(p_{t-1}, c_t), \tag{4.23}$$

and the associated MTS algorithm plays the distribution $\Lambda^{\mathcal{D}}(p_t)|_X$, i.e., at every $x \in X$ (recall that these are precisely the sinks in $\mathcal{D}$), the algorithm places probability mass equal to the flow in $\Lambda^{\mathcal{D}}(p_t)$ entering x.

For $c \in \mathbb{R}_+^X$ and $F \in \mathcal{F}_{\mathcal{D}}$ we define

$$\langle c, F\rangle_X := \sum_{v \in X} c_v F_v = \sum_{uv \in A \,:\, v \in X} c_v F_{uv}.$$

So the service cost of the algorithm until time $t \geqslant 1$ is given by

$$\sum_{s=1}^{t} \left\langle c, \Lambda^{\mathcal{D}}(p_s)\right\rangle_X,$$

and the movement cost is given by

$$\sum_{s=1}^{t} \mathbb{W}^1_{\hat{D}}\left(\Lambda^{\mathcal{D}}(p_{s-1}), \Lambda^{\mathcal{D}}(p_s)\right),$$

where we recall the L^1 transportation distance defined in Section 4.2.2.

4.4.2 Analysis via unfolding to an ultrametric

Let $\hat{\mathcal{D}} = (\mathcal{D}, \omega, \theta)$ be a τ-geometric marked DAG. As in Section 4.3.3, for a point $u \in V$, we define $\sigma(u)$ to denote the number of paths in $\mathcal{D}$ that start at u and end at X. Then if $\mathcal{D}$ is a tree and furthermore, for $uv \in A$, one defines

$$\theta_{uv} := \frac{\sigma(v)}{\sigma(u)}, \tag{4.24}$$

then the algorithm of the preceding section is exactly the same as the one for HSTs introduced in [CL19], as $\sigma(u)$ is precisely the number of leaves in the subtree rooted at u. The next result is a restatement of [CL19, Thm. 2.7].

Theorem 4.21 ([CL19]). *Let $\hat{\mathcal{D}} = (\mathcal{D}, \omega, \theta)$ be a τ-geometric marked DAG over X, for some $\tau \geqslant 4$, and such that $\mathcal{D}$ is a tree. If θ is defined as in (4.24), and $\hat{\mathcal{D}}$ is 1-expanding and L-Lipschitz, then there is some value $\kappa \asymp L$ and a number $\varepsilon = \varepsilon(\hat{\mathcal{D}}) > 0$ so that for any sequence of cost vectors $\langle c_t : t \geqslant 1 \rangle$ satisfying $\|c_t\|_\infty \leqslant \varepsilon$, the MTS algorithm specified in Section 4.4.1 is 1-competitive for service costs and $O\big(L\left(\Delta_0(\mathcal{D}) + \log|X|\right)\big)$-competitive for movement costs.*

Note that the condition on the ℓ_∞ norm of the cost vectors in the above theorem is not restrictive, since as noted in [CL19], we can always split arbitrary cost vectors into smaller pieces with each satisfying the desired ℓ_∞ bound.

Our goal now is to show that if θ is defined as in (4.24), then similar guarantees as in Theorem 4.21 hold for the algorithm on $\hat{\mathcal{D}}$, even when $\mathcal{D}$ is not a tree.

Theorem 4.22. *Let $\hat{\mathcal{D}}$ be a τ-geometric marked DAG over X, for some $\tau \geqslant 4$, and such that θ is given by (4.24). If $\hat{\mathcal{D}}$ is 1-expanding and L-Lipschitz, then there is some value $\kappa \asymp L$ and a number $\varepsilon = \varepsilon(\hat{\mathcal{D}}) > 0$ so that for any sequence of cost vectors $\langle c_t : t \geqslant 1 \rangle$ satisfying $\|c_t\|_\infty \leqslant \varepsilon$, the MTS algorithm specified in Section 4.4.1 is 1-competitive for service costs and $O\big(L\left(\Delta_0(\mathcal{D}) + \log|X|\right)\big)$-competitive for movement costs.*

Note that from (4.8) it follows that

$$\Delta_0(\mathcal{D}) + \log|\mathcal{P}_{\mathcal{D}}| \leqslant \Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}}),$$

and hence the above theorem together with Theorem 4.2 already gives a competitive algorithm with our desired bounds, though only for the specific choice of θ given by (4.24). In Section **??**, we address the case of general θ.

We prove Theorem 4.22 via a simple reduction to Theorem 4.21. Consider a τ-geometric marked DAG $\hat{\mathcal{D}} = (\mathcal{D}, w, \theta)$ on X with $\mathcal{D} = (V, A)$. Note that $d_{\hat{\mathcal{D}}}$ defines an ultrametric on $\mathcal{P}_{\mathcal{D}}$. We show that the dynamics on $\hat{\mathcal{D}}$ are "equivalent" to the dynamics on the HST corresponding to the ultrametric $(\mathcal{P}_{\mathcal{D}}, d_{\hat{\mathcal{D}}})$. More precisely, let us construct the τ-geometric marked tree $\tilde{\mathcal{D}} = (\mathcal{D}', w', \theta')$ with $\mathcal{D}' = (V', A')$ as follows. We define V' as the set of (directed) paths is $\mathcal{D}$ originating from the root. Furthermore, we connect $\gamma \in V'$ to $\gamma' \in V'$ whenever γ' is formed by adding the edge $\bar{\gamma}'$ to γ, and set

$$\omega_{\gamma\gamma'} := \omega_{\bar{\gamma}'}, \qquad \theta_{\gamma\gamma'} := \theta_{\bar{\gamma}'}.$$

One can verify that $\tilde{\mathcal{D}}$ is a τ-geometric marked tree over $\mathcal{P}_{\mathcal{D}}$. Moreover, since $\mathcal{D}'$ is a tree, there is a natural identification between the elements of $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{D}'}$ so that for $\gamma, \gamma' \in \mathcal{P}_{\mathcal{D}}$ it holds that

$$d_{\hat{\mathcal{D}}}(\gamma, \gamma') = d_{\tilde{\mathcal{D}}}(\gamma, \gamma'). \tag{4.25}$$

Now for $p \in Q_{\mathcal{D}}$, define $\tilde{p} \in Q_{\mathcal{D}'}$ to be the natural extension of p in $\mathcal{D}'$ so that for $\gamma\gamma' \in A'$ one has $\tilde{p}_{\gamma\gamma'} = p_{\bar{\gamma}'}$. Furthermore, for a cost sequence $c \in \mathbb{R}_+^X$ define its extension $\tilde{c} \in \mathbb{R}^{\mathcal{P}_{\mathcal{D}}}$ as the vector with $\tilde{c}_\gamma = c_{\bar{\gamma}}$ for $\gamma \in \mathcal{P}_{\mathcal{D}}$. Finally, let $\mathcal{A}$ denote the single-step discrete dynamics on $\hat{\mathcal{D}}$ as defined in Section 4.4.1, and similarly let $\mathcal{A}'$ denote the discrete dynamics on $\tilde{\mathcal{D}}$. Then the following lemma is straightforward.

Lemma 4.23. *Let $p \in Q_{\mathcal{D}}$, $c \in \mathbb{R}_+^X$. Then it holds that*

$$\langle \Lambda^{(\mathcal{D})}(p), c \rangle_X = \langle \Lambda^{(\mathcal{D}')}(\tilde{p}), \tilde{c} \rangle_{\mathcal{P}_{\mathcal{D}}}. \tag{4.26}$$

Furthermore, for $q = \mathcal{A}(p, c)$ we have

$$\mathcal{A}'(\tilde{p}, \tilde{c}) = \tilde{q}. \tag{4.27}$$

We are now ready to prove the main result of this section.

Proof of Theorem 4.22. Let $p_0 \in Q_{\mathcal{D}}$ and $q_0 \in Q_{\mathcal{D}'}$ with $q_0 = \tilde{p_0}$. Given the cost sequence $\langle c_t : t \geqslant 1 \rangle$, for $t \geqslant 1$ let

$$p_t = \mathcal{A}(p_{t-1}, c_t)$$

and

$$q_t = \mathcal{A}'(q_{t-1}, \tilde{c}_t).$$

Then by repeatedly applying (4.27) we get that for $t \geqslant 1$ we have $q_t = \tilde{p_t}$. Therefore from (4.26) and (4.25) it follows that the service and movement costs of the dynamics on $\hat{\mathcal{D}}$ and $\tilde{\mathcal{D}}$ are equal. Hence the competitiveness guarantees for the dynamics on $\hat{\mathcal{D}}$ follow from an application of Theorem 4.21 to the dynamics on $\tilde{\mathcal{D}}$, completing the proof. □

5 Counting and sampling perfect matchings in regular expanding non-bipartite graphs

5.1 Introduction

Given a (general) graph $G = (V, E)$ with $2n = |V|$ vertices, the problem of counting the number of perfect matchings in G is one of the most fundamental open problems in the field of counting. Jerrum and Sinclair in their landmark result [JS89] designed a Monte Carlo Markov Chain (MCMC) algorithm for this task and proved that such an algorithm runs in polynomial time if the ratio of the number of near perfect matchings to the number of perfect matchings is bound by a polynomial (in n). As a consequence one would be able to count perfect matchings if G is *very* dense, i.e., it has min-degree at least n. Not much is known beyond this case, despite several exciting results when the given graph G is bipartite [JV96, LSW00, Bar99, JSV04, BSVV06].

This problem is also extensively studied in combinatorics. Around 40 years ago, Falikman and Egorychev [Ego81, Fal81] proved the van-der-Waerden conjecture, thus showing that if G is a d-regular bipartite graph, then it has at least $(d/e)^n$ perfect matchings. This bound was further improved by Schrijver [Sch98] and simpler and more general proofs were found [Gur06, AOV21]. But it remains a mystery whether van-der-Waerden conjecture extends to non-bipartite graphs. Lovasz, Plummer most famously made the following conjecture:

Conjecture 5.1 ([LP86, Conjecture 8.1.8])**.** *For $d \geqslant 3$, there exist constants $c_1(d), c_2(d) > 1$ such that any d-regular $k-1$-edge connected graph G with $2n$ vertices contains at least $c_1(d)c_2(d)^n$ perfect matchings and $c_2(d) \to \infty$ as $d \to \infty$.*

To this date the above conjecture is only proved for $d = 3$ [EKK+11], although the same proof shows that the conjecture holds for all $d \geqslant 3$ as long as $c_2(d)$ is allowed to be a fixed constant.

At a high-level, the study of perfect matchings in general graphs faces the following barriers:

- Unlike bipartite graphs, the perfect matching polytope of a general graph has exponentially many constraints, and it is believed that there does not exist any poly-size convex program to test whether a given graph has perfect matchings. This fact significantly limits exploiting Gurvits' like techniques [Gur06] in lower-bounding the number of perfect matchings.

- In a bipartite graph, any odd alternating *walk* (that starts and ends at un-saturated vertices) can be used to extend a near perfect matching to a perfect matching. However, in a general graph, an odd alternating walk may contain odd cycles. Therefore, typical augmenting path arguments which bound the ratio of near perfect to perfect matchings fail in a non-bipartite graph (see e.g., [JV96]).

In this paper we study perfect matchings in regular *strong* expander graphs: We show that for these graphs the classical algorithm of [JS89] runs in polynomial time and can generate an approximately uniform random perfect matching. On the combinatorial side, we prove a significantly stronger version of Conjecture 5.1 for this family of graphs.

5.1.1 Main Contributions

Given a graph $G = (V, E)$, let $A_G \in \mathbb{R}^{2n\times 2n}$ be its adjacency matrix, and let $D \in \mathbb{R}^{2n\times 2n}$ be the diagonal matrix of vertex degree. The normalized adjacency matrix of G is defined as $\tilde{A}_G = D^{-1/2}AD^{-1/2}$; when G is clear in the context we may drop the subscript. Let $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_{2n}$ be the eigenvalues of $\tilde{A}$. We write

$$\sigma_2(\tilde{A}) = \max\{\lambda_2, |\lambda_{2n}|\},$$

to denote the largest eigenvalue of $\tilde{A}$ in absolute value (excluding λ_1).

Definition 5.2. For $0 < \varepsilon < 1$, we write G is an ε-spectral expander if $\sigma_2(\tilde{A}) \leqslant \varepsilon$.

For two probability distributions μ, ν defined in $\{1, \ldots, n\}$, the total variation distance of μ, ν is $\frac{1}{2}\sum_{i=1}^{n} |\mu_i - \nu_i|$.

Theorem 5.3 (Algorithm)**.** *There is a randomized algorithm that for $\varepsilon \leqslant 1/11$, $\delta > 0$, given a d-regular ε-spectral expander G on $2n$ vertices outputs a perfect matching of G from a distribution μ of total variation distance δ of the uniform distribution (of perfect matchings) in time $\text{poly}(n^{\log_{1/\varepsilon} d}, \log(1/\delta))$. Furthermore, there is a randomized algorithm that for any $\delta > 0$ approximates the number of perfect matchings of G up to $1 \pm \delta$ multiplicative factor in time $\text{poly}(n^{\log_{1/\varepsilon} d}, 1/\delta)$.*

In particular, observe that the running time of the above algorithms is polynomial in n if d is a constant or $1/\varepsilon$ is a polynomial in d and it is quasi-polynomial in n otherwise.

Theorem 5.4 (Lower Bound)**.** *For any $\varepsilon \leqslant 1/11$, every d-regular ε-spectral expander on $2n$ vertices has at least $(d/e)^n \left(\frac{\varepsilon}{2e^3d^6}\right)^{\varepsilon n}$ many perfect matchings.*

Putting the above theorem together with [EKK+11] proves Conjecture 5.1 for (strong) spectral graphs.

Recall that by a work of Friedman, a random d-regular graph is a $\varepsilon = \frac{2\sqrt{d-1}+o(1)}{d}$-spectral expander with probability $1 - 1/\text{poly}(n)$[Fri08, Bor19]. So, for a sufficiently large value of d, we can count the number perfect matchings in random d-regular graphs up to $1 \pm \delta$-multiplicatively in time polynomial in $n, 1/\delta$. Furthermore, the above theorem implies that the Lovasz-Plummer Conjecture 5.1 holds for almost all graphs.

We remark our proof technique can naturally be extended to non-regular expanders where the ratio of maximum to minimum degree is bounded. However, in the following statement we show that if this ratio is unbounded the graph may not even have a single perfect matching.

Theorem 5.5. *For $d \geqslant 3$, there exists $n_0 > 0$ such that for any $n \geqslant n_0$, there is a $O(1/\sqrt{d})$-spectral expander G on $2n$ vertices that does not have any perfect matchings.*

5.1.2 Related Works

Bollabás and McKay [BM86] showed that when $d = \Omega(\log^{1/3} n)$, as $d \to \infty$, a random d-regular graph contains $\Omega(d)^n$ many perfect matchings with probability at least $1 - 1/d^2$. Similar bounds are also implicitly given in the work of Robinson and Wormald [RW94], but to the best of our knowledge no explicit bound on the number of perfect matchings was known which holds almost surely as $n \to \infty$. Note that Theorem 5.4 implies that this statement is true with probability $1 - 1/\text{poly}(n)$ even if d is as small as a constant.

Chudnovsky and Seymour [CS12] proved that any planar cubic graph with no cut edge has at least $2^{n/655978752}$ many perfect matchings. Building on [CS12], Esperet, Kardos, King, Král, and Norine [EKK+11] showed that any d-regular $d-1$ edge connected graph has at least $2^{(1-3/d)\frac{n}{3656}}$ perfect matchings. Barvinok [Bar13] showed that any 3-regular graph in which any set S with $2 \leqslant |S| \leqslant |V| - 2$ satisfies $|E(S, \overline{S})| \geqslant 4$ has at least c^n many perfect matchings for some universal constant $c > 1$.

Jerrum and Sinclair [JS89] showed the ratio of perfect to near perfect matchings in *bipartite* Erdös-Réyni graphs is polynomial in n. Thus, one can efficiently sample a perfect matching in such graphs. However, to the best of our knowledge, no such result is known for (non-bipartite) random graphs.

Barvinok [Bar99] designed a randomized c^n approximation algorithm to the number of perfect matchings of any (general) graph, for some universal constant $c > 1$. Rudelson, Samarodnitsky, Zeitouni [RSZ16] showed that for a family of strong expander graphs Barvinok's estimator [Bar99] has a sub-exponential variance, thus obtaining a randomized polynomial time sub-exponential approximation algorithm for the number of perfect matchings of any such graphs.

Gamarnik and Katz [GK10] designed a *deterministic* $(1+\varepsilon)^n$ approximation algorithm to the number of perfect matching in expanding *bipartite* graphs.

5.1.3 Overview of Approach

At high-level our proof builds on works of [JV96, GK10]. We show that given a non-perfect matching M in a (strong) expander graph G, one can find many augmenting paths of length $O(\log \frac{n}{n-|M|})$. lemmanearpertoper Let G be a d-regular ε-spectral expander graph on $2n$ vertices for $\varepsilon \leqslant 1/11$, and let M be any (not perfect) matching in G. Then there exist at least $\lceil (n - |M|)/2 \rceil$ augmenting paths in G of length at most $\rho = O\left(\max\left(\log_{1/\varepsilon}(\frac{2\varepsilon n}{n-|M|}), 1\right)\right)$ for ρ defined in Lemma 5.14. As alluded to in the introduction, the main difficulty in proving the above theorem is that since G is not necessarily bipartite, an augmenting walk cannot necessarily be turned into an augmenting path since it may have odd cycles. To avoid this issue, first we construct a random bi-partitioning of the vertices of G by placing the endpoints of each edge of M on opposite sides. We exploit the expansion property of G to argue that, under this random bi-partition, every set expands with high probability. So, one can start from two unsaturated vertices and follow "alternating BFS trees" from each until getting to a common middle point. The expansion property allows us to show that, with high probability, after $\log_{1/\varepsilon} n$ steps we can construct an augmenting path. This method essentially tries to mimic the approach of [JV96] while exploiting the random partitioning. As an immediate corollary of the above lemma, we can upper bound the ratio of k to $k+1$ matchings in expanders. lemmaratio Let G be a d-regular ε-spectral expander graph on $2n$ vertices, and let $k \in [n]$. Let $m(j)$ denote the number of matchings of size j in G. Then we have

$$\frac{m(k)}{m(k+1)} \leqslant \frac{2(k+1)}{n-k} d^{(\rho-1)/2}$$

for ρ defined in Lemma 5.14. Building on [JS89], this lemma is already enough to prove Theorem 5.3.

To prove Theorem 5.4, we first show that for some constant $\varepsilon > 0$, G has at least $\Omega(d)^n$ many $n(1-\varepsilon)$-matchings. This part uses a greedy algorithm to find so many distinct matchings in an expander graph. Then, we exploit the above lemma to argue that the ratio of the number of $n(1-\varepsilon)$ matchings of G to the number of its perfect matchings is at most $d^{O(\varepsilon)n}$.

5.2 Preliminaries

Given a graph $G = (V, E)$ with $|V| = 2n$ and $k \in [n]$, a k-matching $M \subseteq E$ is any subset with $|M| = k$ and $e \cap e' = \emptyset$ for all $e \neq e' \in M$. For a set $S \subseteq V$, we write $G[S]$ to denote the *induced* subgraph on the set S. For a vertex $v \in V$, we write $\deg_G(v)$ to denote the degree of v in G.

Given a set of vertices $S \subseteq V$, define

$$M(S) := \{v : \exists u \in S, (u, v) \in M\}.$$

We also define $m_G(k)$ to denote the number of k-matchings in G.

Given a matching M, a walk $v_0, v_1, \ldots, v_k$ is an *alternating walk* for M if for any $1 \leqslant i \leqslant k-1$ exactly one of (v_{i-1}, v_i) and (v_i, v_{i+1}) is in M. An *augmenting path* for M is any alternating path that starts and ends with an unmatched vertex.

For a graph $G = (V, E)$ and $S, T \subseteq V$,

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

For a set $S \subseteq V$, we write

$$N_G(S) := \{u \notin S : \exists u \in S, (u, v) \in E\}$$

to denote the set of all vertices outside S that has an edge to S. When the graph G is unambiguous from the context, we may drop the subscripts.

5.2.1 Spectral Graph Theory

The following facts are the main properties of spectral expanders that we will need.

Fact 5.6 (Expander Mixing Lemma)**.** *Let G be a d-regular graph on $2n$ vertices. Then for any two sets $S, T \subseteq V$, we have*

$$\left| |E(S,T)| - \frac{d|S| \cdot |T|}{2n} \right| \leqslant d\sigma_2(\tilde{A})\sqrt{|S| \cdot |T|}$$

Lemma 5.7. *Let $G = (V, E)$ be a $2n$-vertex d-regular ε-expander, and let $S \subseteq V$. The following holds: Then, there exists $v \in S$ such that $\deg_{G[S]}(v) \geqslant \lceil d(|S|/2n - \varepsilon) \rceil$.*

Proof. By the Expander Mixing Lemma (Fact 5.6), we have $|E(S,S)| \geqslant \frac{d|S|^2}{2n} - d\varepsilon|S|$. Hence the average degree of the vertices in $G[S]$ is at least $d(|S|/2n - \varepsilon)$, and in particular there exists $v \in S$ whose degree in $G[S]$ is at least that much. ◻

Lemma 5.8 ([Tan84]). *Let G be a d-regular ε-expander on $2n$ vertices. Then for any $S \subseteq V$ we have*

$$|N(S)| \geqslant \frac{|S|}{\varepsilon^2 + (1-\varepsilon^2)|S|/2n}$$

When $|S| \leqslant 2\varepsilon n$, we immediately get the following corollary.

Corollary 5.9. *Let G be a d-regular ε-expander on $2n$ vertices. Then for any $S \subseteq V$ with $|S| \leqslant 2\varepsilon n$ we have*

$$|N(S)| \geqslant \frac{|S|}{\varepsilon^2 + \varepsilon - \varepsilon^3} \geqslant \frac{|S|}{\varepsilon^2 + \varepsilon}.$$

5.2.2 Inequalities

Theorem 5.10 (Hoeffding's Inequality). *Let $X_1, \ldots, X_k$ be independent random variables in the range $[0,1]$. Then,*

$$\mathbb{P}\sum X_i < \mathbb{E}\sum X_i - \varepsilon \leqslant \exp(-2\varepsilon^2/k).$$

Theorem 5.11 (Stirling's Formula). *For $n \geqslant 1$ we have*

$$n! \geqslant \left(\frac{n}{e}\right)^n.$$

Theorem 5.12 (Weierstrass's Inequality). *Let $0 < x_i < 1$ for $1 \leqslant i \leqslant n$. Then,*

$$\prod_{i=1}^{n}(1-x_i) \geqslant 1 - \sum_{i=1}^{n} x_i.$$

Theorem 5.13 (Hoffman-Wielandt's Inequality). *Let $A, B \in \mathbb{R}^{n\times n}$ be symmetric matrices with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and $\lambda'_1 \geqslant \cdots \geqslant \lambda'_n$, respectively. We have*

$$\sum_{i=1}^{n}(\lambda_i - \lambda'_i)^2 \leqslant \|A-B\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

5.3 Proof of the Main Lemma

The following lemma is the main result of this subsection.

Lemma 5.14. *Let $G = (V,E)$ be a d-regular ε-spectral expander graph on $2n$ vertices with $\varepsilon \leqslant 1/11$, M be any (not perfect) matching in G, and U the set of unsaturated vertices (in M). For any partitioning of $U = U_L \cup U_R$ with $|U_L| = |U_R|$ there is an augmenting path from U_L to U_R of length at most $\rho = 4\max\left(\lceil\log_{C_1(\varepsilon)}(\frac{2\varepsilon n+1}{n-|M|})\rceil, 0\right) + 1$, where $C_1(\varepsilon) = \frac{1}{\varepsilon+\varepsilon^2}$.*

Before proving this lemma we use it to prove Section 5.1.3.

Proof of Section 5.1.3. Let U be the set of unmatched vertices of M and let U' be vertices of U that are not an endpoint to any augmenting path of length at most ρ. Observe that if $|U'| <= n - |M|$, then there are at least $\lceil (n - |M|)/2 \rceil$ augmenting paths for M and we are done.

For the sake of contradiction suppose $|U'| > n - |M|$. Now arbitrarily partition U into two equal-sized sets $U_L \cup U_R$ (each of size exactly $n - |M|$) with the constraint that $U_L \subseteq U'$. So, by construction, no vertex in U_L is an endpoint of augmenting path of length at most ρ. But, by Lemma 5.14 there is an augmenting path from U_L to U_R of length at most ρ which is a contradiction. $\square$

Proof of Section 5.1.3. Given a k-matching M, by Section 5.1.3 there are at least $(n-k)/2$ augmenting paths for M (in G) of length at most ρ for ρ defined in Section 5.1.3. Note that for any vertex v of G the number of paths of length at most ρ starting at v is at most d^{ρ}. Therefore, for any $k+1$-matching M', there are at most $2(k+1)d^{(\rho-1)/2}$ k-matchings that can be mapped to M'. This is because any such matching can be obtained by "undoing" an alternating path that starts and ends at the saturated vertices of M'. Together, these imply $\frac{m(k)}{m(k+1)} \leqslant \frac{2(k+1)d^{(\rho-1)/2}}{n-k}$. $\square$

Definition 5.15 (Bipartition of G)**.** Given a matching M and $\omega : M \to \{0,1\}$, we define the bipartite graph $G_M(\omega) = (L_M(\omega), R_M(\omega), E_M(\omega))$ as follows. We drop the subscript M and ω if they are clear in the context.

All vertices of U_L are in L, all vertices of U_R are in R. For any edge $e = (u,v) \in M$, we add u to L and v to R if $\omega(e) = 0$ and we add u to R and v to L otherwise. We simply let $E_M(\omega)$ be all edges of E connecting L to R. We use μ_M to denote the uniform distribution over functions $M \to \{0,1\}$.

Lemma 5.16. *Let $G = (V,E)$ be a graph with $2n$ vertices such that for every set $S \subseteq V$ with $|S| \leqslant 2\varepsilon n$, $|N(S)| \geqslant \alpha|S|$ for $\alpha \geqslant 10$ and $0 < \varepsilon < 1$. Given a non-perfect matching M and a partition of non-saturated vertices into equal sized sets U_L, U_R, if for $t = \max(\lceil \log_{\alpha/4} \frac{2\varepsilon n+1}{|U_L|} \rceil, 0)$ there is no augmenting path of length at most $4t+1$ from U_L to U_R, then with probability $> 1/2$ (for $\omega \sim \mu_M$) there exists a set $S \subseteq L$ such that $|S| > 2\varepsilon n$, and for every $v \in S$ there is an alternating path of length at most $2t$ from U_L to v in $G_M(\omega)$.*

Proof of Lemma 5.14. First, by Corollary 5.9, since G is an ε-spectral expander and $\varepsilon < 1/11$, we can let $\alpha = 1/(\varepsilon + \varepsilon^2) \geqslant 10$. We prove the claim by contradiction. Suppose G has no augmenting path of length $\rho := 4t+1$ from U_L to U_R, for t defined in Lemma 5.16. By Lemma 5.16, for $\omega \sim \mu$, with probability $> 1/2$ there is a set $S \subseteq L$ with $|S| > 2\varepsilon n$, such that for any $v \in S$ there is an alternating path in $G_M(\omega)$ of length (at most) $2t$ from U_L to v. By renaming U_L, U_R, with probability $> 1/2$ there also exists another set $S' \subseteq R$ such that $|S'| > 2\varepsilon n$ such that for every $v \in S'$, there is an alternating path of length at most $2t$ from U_R to v in $G_M(\omega)$. By union bound, with positive probability both of these sets exist. Now, by Fact 5.6 we have

$$|E(S,S')| \geqslant \frac{d|S|\cdot|S'|}{2n} - \varepsilon d\sqrt{|S|\cdot|S'|} > d\sqrt{|S|\cdot|S'|}(\varepsilon - \varepsilon) = 0.$$

So there is an edge $(v,v') \in E(S,S')$. Now, the path formed by concatenating an alternating path from U_L to v of length $2t$, the edge (v,v'), and an alternating path from v' to U_R of length $2t$ we find alternating walk of length (at most) $\rho = 2t + 2t + 1$ from U_L to U_R in $G_M(\omega)$. But since $G_M(\omega)$ is a bipartite graph this walk can only have even length cycles; by removing these cycles we obtain an alternating path of length at most ρ from U_L to U_R (in $G_M(\omega)$). $\square$

In the rest of this subsection, we prove Lemma 5.16. First note that as $\alpha/4 > 1$, we have $\log_{\alpha/4} \frac{2\varepsilon n+1}{|U_L|} \leqslant 0$ if and only if $|U_L| > 2\varepsilon n$, and the claim is trivial in this case as we can set $S = U_L$. Now suppose $|U_L| \leqslant 2\varepsilon n$. Let us fix an arbitrary ordering on the vertices of G. Given a bipartition $G_M(\omega)$, we define a sequence of sets $U_L = L_0 \subseteq L_1 \subseteq \ldots L_T \subseteq L$, and $\emptyset = X_0 \subseteq X_1 \subseteq \ldots X_T \subseteq V$, where T is a stopping time which is the minimum of t and the first time that $|L_T| > \varepsilon n$. Given L_{i-1}, X_{i-1} for $i \geqslant 1$, we construct L_i, X_i as follows: If $|L_{i-1}| > \varepsilon n$ then we stop and we let $T = i - 1$. Otherwise, $|L_{i-1}| \leqslant 2\varepsilon n$ so by assumption of the lemma, $N(L_{i-1}) \geqslant \alpha|L_{i-1}|$. Let A_i be the lexicographically first $\alpha|L_{i-1}| - |X_{i-1}|$ neighbors of L_{i-1} which are not in X_{i-1}. In other words, we sort all neighbors of L_{i-1} which are not in X_{i-1} lexicographically and we let the first $\alpha|L_{i-1}| - |X_{i-1}|$ of them to be A_i. Note that as L_{i-1} has at least $\alpha|L_{i-1}|$ neighbors, there are at least $\alpha|L_{i-1}| - |X_{i-1}|$ "new" neighbors and so the set A_i is well-defined. We let $X_i = X_{i-1} \cup A_i$. Observe that by definition, we always have

$$|X_i| = \alpha|L_{i-1}|. \tag{5.1}$$

Finally, we let

$$L_i = L_{i-1} \cup M(A_i \cap R) = L_{i-1} \cup M(X_i \cap R).$$

The following fact follows inductively from the above construction

Fact 5.17. *For every $1 \leqslant i \leqslant T$ and every $v \in L_i$, there is an alternating path of length at most $2i$ from $U_L = L_0$ to v in $G_M(\omega)$.*

Fact 5.18. *For any $1 \leqslant i \leqslant T$, $L_{i-1} \cap M(A_i \cap R) = \emptyset$. Therefore,*

$$|L_i| = |L_{i-1}| + |A_i \cap R|.$$

Proof. For the sake of contradiction let $v \in M(A_i \cap R)$ such that $v \in L_{i-1}$ as well. Then, since v has a match, $v \notin U_L$; so we must have $i \geqslant 2$. Let $1 \leqslant j \leqslant i - 1$ be the smallest index such that $v \in L_j$. That means that, by construction, $M(v) \in A_j \cap R$. Therefore, $M(v) \in X_j \subseteq X_{i-1}$. So, $v \notin M(A_i)$. □

Since in the above construction we only "look at" the first $\alpha|L_{i-1}| - |X_{i-1}|$ new neighbors of L_{i-1} to construct L_i, it follows that all edges which have no endpoints in these sets are conditionally independent. More precisely, we obtain the following Fact.

Fact 5.19. *Let ω be chosen uniformly at random. For any $1 \leqslant i < t$, conditioned on $L_0, \ldots, L_{i-1}$, the law of ω on all edges that have no endpoints in L_{i-1}, X_{i-1} remain invariant, i.e., it is i.i.d., with expectation $1/2$ on each edge.*

Claim 5.20. For $1 \leqslant i \leqslant T$,

$$\mathfrak{P}\omega \sim \mu |A_i \cap R| \leqslant |A_i|/4 \mid L_0, \ldots, L_{i-1} < \exp(-|A_i|/8).$$

Proof. Note that given $L_0, \ldots, L_{i-1}$, $X_1, \ldots, X_i$ and $A_1, \ldots, A_i$ are uniquely determined. Let $v \in A_i$. Consider the following cases:

- $v \in U_R$. This case cannot happen because we get an augmenting path of length $2i + 1$ to U_L which is a contradiction.
- $v \in L_{i-1}$. This cannot happen because $v \in N(L_{i-1})$. This in particular shows $v \notin U_L$. So, v has a match in M.

- $M(v) \in A_i$. Then, by Definition 5.15 exactly one of $v, M(v)$ is in R.
- $M(v) \in X_{i-1}$. If $M(v) \in R$ then we must have $v \in L_{i-1}$ which cannot happen as we said in case (2). Otherwise, $M(v) \in L$, so $v \in R$.
- $M(v) \in L_{i-1}$. Then, $v \in R$.
- $v \notin L_{i-1}, M(v) \notin X_i, L_{i-1}$. In this case since $v \in A_i$, by Fact 5.19, $v \in R$ with probability $1/2$ independent of all other vertices of A_i.

Let A'_i be the set of vertices v that fall into the last case. Say we have a Bernoulli B_v with success probability $1/2$ for every $v \in A'_i$. Then, by above discussion, conditioned on $L_0, \dots, L_{i-1}$, with probability 1,

$$|A_i \cap R| \geqslant |A_i \setminus A'_i|/2 + \sum_{v \in A'_i} B_v.$$

Therefore, by the Hoeffding bound (Theorem 5.10)

$$\begin{aligned}\mathbb{P}|A_i \cap R| \leqslant |A_i/4| \mid L_0, \dots, L_{i-1} &\leqslant \mathbb{P} \sum_{v \in A'_i} B_v \leqslant |A'_i|/2 - |A_i|/4 \,\Big|\, L_0, \dots, L_{i-1} \\ &\leqslant \exp(-|A_i|^2/8|A'_i|) \leqslant \exp(-|A_i|/8)\end{aligned}$$

as desired. □

Since $|L_1| \geqslant |A_1 \cap R|$ and $|A_1| = \alpha|L_0|$,

$$\begin{aligned}\mathfrak{P}\mu|L_1| \geqslant (\alpha/4)|L_0| &\geqslant \mathfrak{P}\mu|A_1 \cap R| \geqslant (\alpha/4)|L_0| \\ &= \mathfrak{P}\mu|A_1 \cap R| \geqslant |A_1|/4 \geqslant 1 - \exp(-|A_1|/8) = 1 - \exp(-\alpha|L_0|/8) \qquad (5.2)\end{aligned}$$

where the last inequality follows form Claim 5.20.

Claim 5.21. Let ω be chosen uniformly at random. For every $2 \leqslant i \leqslant T$, we have

$$\mathfrak{P}\mu|L_i| \geqslant (\alpha/4)|L_{i-1}| \,\big|\, |L_{i-1}| \geqslant (\alpha/4)|L_{i-2}|, L_0, \dots, L_{i-1} \geqslant 1 - \exp(-(\alpha-4)|L_{i-1}|/8).$$

Proof. Suppose $|L_{i-1}| \geqslant (\alpha/4)|L_{i-1}|$. Recall that by Eq. (5.1) we have $|X_i| = \alpha|L_{i-1}|$. So we can write

$$\begin{aligned}|A_i| = |X_i \setminus X_{i-1}| &= \alpha(|L_{i-1}| - |L_{i-2}|) \\ &\geqslant \alpha(1 - 4/\alpha)|L_{i-1}| \qquad (5.3) \\ &= (\alpha - 4)|L_{i-1}|.\end{aligned}$$

Let μ' be μ conditioned on $|L_{i-1}| \geqslant (\alpha/4)|L_{i-2}|$ and $L_0, \dots, L_{i-1}$. Then,

$$\begin{aligned}\mathfrak{P}\mu'|L_i| \leqslant (\alpha/4)|L_{i-1}| &= \mathfrak{P}\mu'|L_i| - |L_{i-1}| \leqslant (\alpha/4 - 1)|L_{i-1}| \\ &= \mathfrak{P}\mu'|A_i \cap R| \leqslant (\alpha/4 - 1)|L_{i-1}| && \text{(Fact 5.18)} \\ &\leqslant \mathfrak{P}\mu'|A_i \cap R| \leqslant |A_i|/4 && \text{(Eq. (5.3))} \\ &\leqslant \exp(-|A_i|/8) && \text{(Claim 5.20)} \\ &\leqslant \exp(-(\alpha-4)|L_{i-1}|/8), && \text{(Eq. (5.3))}\end{aligned}$$

completing the proof. □

Claim 5.22. If $\alpha \geqslant 10$, for any $i \geqslant 1$ we have

$$\mathfrak{P}\mu T < i \vee (T \geqslant i \wedge |L_i| \geqslant (\alpha/4)^i |L_0|) > 1/2.$$

Proof. For $1 \leqslant j \leqslant i$, let E_j denote the event that $T < j \vee (T \geqslant j \wedge |L_j| \geqslant (\alpha/4)|L_{j-1}|)$. Then,

$$\begin{aligned}
\mathbb{P}E_i &\geqslant \mathbb{P}E_1 \wedge \cdots \wedge E_i \\
&= \mathbb{P}E_1 \prod_{j=2}^{i} \mathbb{P}E_j | E_1, \ldots, E_{j-1} \\
&\geqslant (1 - \exp(-\alpha|L_0|/8)) \prod_{j=2}^{i} \Big(\mathbb{P}T < j | E_1, \ldots, E_{j-1} \\
&\quad + \mathbb{P}T \geqslant j | E_1, \ldots, E_{j-1} \, \mathbb{E}\, 1 - \exp\left(\frac{-(\alpha-4)|L_{j-1}|}{8} \right) \Big| \, E_1, \ldots, E_{j-1}, T \geqslant j \Big) && \text{(Claim 5.21)} \\
&\geqslant (1 - \exp(-\alpha|L_0|/8)) \prod_{j=2}^{i} \left(1 - \exp\left(\frac{-(\alpha-4)}{8} (\alpha/4)^{j-1} |L_0| \right) \right) \\
&\geqslant 1 - \exp(-\alpha/8) - \sum_{j=2}^{i} \exp(-\frac{\alpha-4}{8}(\alpha/4)^{j-1}) && \text{(Theorem 5.12)} \\
&\geqslant 1 - e^{-\alpha/8} - \sum_{j=0}^{\infty} e^{-\beta(\alpha/4)^j} && (\text{for } \beta = \tfrac{\alpha(\alpha-4)}{32}) \\
&\geqslant 1 - e^{-\alpha/8} - \frac{e^{-\beta}}{1 - e^{-\beta(\alpha/4-1)}} > 1/2. && (\alpha \geqslant 10)
\end{aligned}$$

Note that in the third inequality we crucially use that if $E_1, \ldots, E_{j-1}$ occur then either $T < j$, or $T \geqslant j$ and $|L_j| \geqslant (\alpha/4)^{j-1}$. □

Setting $t = \lceil \log_{\alpha/4} \frac{2\varepsilon n+1}{|U_L|} \rceil$ by the above statement we get $\mathbb{P}T < t \vee (T = t \wedge |L_t| > 2\varepsilon n) > 1/2$. This completes the proof of Lemma 5.16.

5.4 Completing the Proofs of Theorems 5.3 and 5.4

5.4.1 The Lower-Bound

Lemma 5.23. *Let $G = (V, E)$ be an $2n$-vertex d-regular ε-expander. If $\varepsilon < 1/2$, then we have,*

$$m((1-\varepsilon)n) \geqslant \left(\frac{d}{e}\right)^{n(1-\varepsilon)} \cdot e^{-2\varepsilon n}.$$

Proof. Let $k = n(1-\varepsilon)$. We call a sequence of integers $\langle a_1, \ldots, a_k \rangle$ *valid* if $1 \leqslant a_i \leqslant \lceil d((n-i+1)/n - \varepsilon) \rceil$ for all $1 \leqslant i \leqslant k$.

Now, for valid sequence $a = \langle a_1, \ldots, a_k \rangle$, we construct a k-matching $\mathcal{M}(a)$ as follows: We are going to construct a sequence of matchings $M_0 \subseteq M_1 \subseteq \cdots \subseteq M_k$, with the property that for

$1 \leqslant i \leqslant k$, M_i is going to be a matching of size i in G. We then set $\mathcal{M}(a) := M_k$. We start with $M_0 = \emptyset$. For $i \geqslant 1$, given M_{i-1}, let S_i be the set of unmatched vertices of G with respect to M_{i-1}. Note that by construction M_{i-1} is a matching of size $i-1$, so we have $|S_i| = 2n - 2(i-1)$. Further let $\Delta_i = \max_{u \in S} \deg_{G[S_i]}(u)$, and and let u_i denote the lexicographically first vertex with degree Δ_i in S_i. Note that by Lemma 5.7, it should be that $\Delta_i \geqslant \lceil d((n-i+1)/n - \varepsilon) \rceil$, and furthermore, by validity of a we obtain $a_i \leqslant \Delta_i$. Now let let v_i be the a_i-th neighbor of u_i in $G[S_i]$ with respect to the lexicographical order. We set $M_i = M_{i-1} \cup \{(u_i, v_i)\}$. In the next claim we show that any distinct pair of valid sequences give distinct k-matchings. Therefore, the number of k-matchings of G is at least,

$$\prod_{i=1}^{k} (d\frac{n-i+1}{n} - d\varepsilon) \geqslant d^k \prod_{i=1}^{k} \frac{n-i+1-\varepsilon n}{n} \geqslant d^k \frac{k!}{n^k} \geqslant (d/e)^k \cdot (k/n)^k,$$

where the last inequality uses Theorem 5.11. By plugging in $k = (1-\varepsilon)n$ we obtain

$$m((1-\varepsilon)n) \geqslant (d/e)^{(1-\varepsilon)n} \cdot (1-\varepsilon)^{(1-\varepsilon)n} \geqslant (d/e)^{(1-\varepsilon)n} \cdot e^{-\varepsilon n},$$

where in the last inequality we used that $(1-\varepsilon)^{1-\varepsilon} \geqslant e^{-\varepsilon}$ for $\varepsilon \leqslant 1/2$. □

Claim 5.24. For any distinct valid sequences $a = \langle a_1, \ldots, a_k \rangle$ and $b = \langle b_1, \ldots, b_k \rangle$ we have $\mathcal{M}(a) \neq \mathcal{M}(b)$.

Proof. Since $a \neq b$ there is an index $1 \leqslant i \leqslant k$ such that $a_i \neq b_i$; let $1 \leqslant i \leqslant k$ be the first such index. Since $a_j = b_j$ for $1 \leqslant j \leqslant i-1$, by the above construction we have $S_i(a) = S_i(b)$. So, we would choose a unique vertex u_i in both constructions but we match it to different vertices, since $a_i \neq b_i$. Therefore $\mathcal{M}(a) \neq \mathcal{M}(b)$. □

Lemma 5.25. *Let G be a $2n$ vertex d-regular, ε-spectral expander for $\varepsilon \leqslant 1/11$. We have,*

$$\frac{m((1-\varepsilon)n)}{m(n)} \leqslant (2e/\varepsilon)^{\varepsilon n} d^{2\varepsilon n + (4\varepsilon n + 2)/\ln C_1(\varepsilon)}.$$

where $C_1(\varepsilon)$ is defined in Lemma 5.14.

Proof. For $k = \varepsilon n$, we can write

$$\begin{aligned} \frac{m(n(1-\varepsilon))}{m(n)} = \prod_{i=n-k}^{n-1} \frac{m(i)}{m(i+1)} &\leqslant \prod_{i=n-k}^{n-1} \frac{2(i+1)}{n-i} d^{2\log_{C_1(\varepsilon)} \frac{2\varepsilon n+1}{n-i} + 2} && \text{(Section 5.1.3)} \\ &\leqslant \frac{2^k n^k d^{2k}}{k!} d^{2\sum_{i=1}^{\varepsilon n} \log_{C_1(\varepsilon)} \frac{2\varepsilon n+1}{i}} \\ &\leqslant (2ed^2 n/k)^k d^{2\log_{C_1(\varepsilon)} \frac{(2k+1)^k}{k!}} \leqslant (2ed^2 n/k)^k d^{(4k+2)/\ln C_1(\varepsilon)}, \end{aligned}$$

where in the second to last inequality we used Theorem 5.11 and in the last inequality we used $\frac{(2k+1)^k}{k!} = \frac{k^k}{k!}(2+1/k)^k \leqslant e^{2k+1}$. Plugging $k = \varepsilon n$ into the above inequality proves the claim. □

Proof of Theorem 5.4. Using Lemmas 5.23 and 5.25 we can write and using for $\varepsilon \leqslant 1/11$, $\ln(C_1(\varepsilon)) \geqslant 2$,

$$m(n) \geqslant \frac{e^{-\varepsilon n}(d/e)^{n(1-\varepsilon)}}{(2e/\varepsilon)^{\varepsilon n} d^{2\varepsilon n + (4\varepsilon n+2)/\ln C_1(\varepsilon)}} \geqslant \left(\frac{d}{e}\right)^n \left(\frac{\varepsilon}{2e^3 d^6}\right)^{\varepsilon n}$$

as desired. □

5.4.2 Sampling / Counting Perfect Matchings

As an immediate corollary of Section 5.1.3 we prove Theorem 5.3. In particular,

$$\frac{m(n-1)}{m(n)} \leqslant 2nd^{2\log_{C_1(\varepsilon)}(2\varepsilon n+1)+2} \tag{5.4}$$

So, it follows from the following theorem of [JS89] that for any $\delta > 0$ we can sample a perfect matching of G from a distribution μ of total variation distance δ of the uniform distribution in time $\mathrm{poly}(n^{\frac{\log d}{\log \varepsilon^{-1}}}, \log(1/\delta))$.

Theorem 5.26 (Jerrum and Sinclair [JS89, Thm 3.6])**.** *Let G be a graph with $2n$ vertices. There is a Markov chain with a uniform stationary distribution on the space n and $n-1$ matchings of G such that that mixes in time* $\mathrm{poly}(n, \frac{m(n-1)}{m(n)})$.

Furthermore, Jerrum and Sinclair [JS89, Thm 5.3] showed how to estimate the number of perfect matchings up to $1 \pm \delta$ multiplicative factor in time $\mathrm{poly}(n, 1/\delta, \frac{m(n-1)}{m(n)})$. So, plugging in Eq. (5.4) into their theorem also allows us to approximate the number of perfect matchings (up to $1 \pm \delta$ multiplicatively) in ε-expander regular graphs in time $\mathrm{poly}(n^{\frac{\log d}{\log \varepsilon^{-1}}}, 1/\delta)$.

5.5 A Non-regular Counter-example

In this subsection we construct an infinite family of *non-regular* strong spectral expanders that do not have any perfect matchings. This shows that the regularity assumption in Theorem 5.4 is necessary.

Lemma 5.27. *Given a d-regular graph $G = (V, E)$ with $2n$ vertices, there exists a graph $H = (V', E')$ with $2n+2$ vertices such that*

- *H does not have any perfect matchings.*
- *$\sigma_2(\tilde{A}_H) \leqslant \sigma_2(\tilde{A}_G) + \sqrt{5/d}$.*
- *H has $2n-1$ vertices of degree d, one vertex of degree $d+2$, and two vertices of degree 1.*

Proof. Say $V = \{v_1, \ldots, v_{2n}\}$. To construct H, we add two new vertices v_{2n+1}, v_{2n+2} and we connect both of them to v_{2n}. Clearly H has no perfect matchings. We abuse notation and extend the normalize adjacency matrix of G, $\tilde{A}_G$ by adding two all-zeros rows and two all-zeros columns. Clearly, only introduces two new zero eigenvalues, and the $\sigma_2(\tilde{A}_G)$ remains invariant. It follows by a simple calculation that

$$\|\tilde{A}_G - \tilde{A}_H\|_F^2 = 2(d-1) \cdot \left(\frac{1}{d(d+1)}\right)^2 + 4\left(\frac{1}{\sqrt{d+1}}\right)^2 \leqslant \frac{2}{d^3} + \frac{4}{d} \leqslant \frac{5}{d}.$$

Therefore, by Theorem 5.13, for any $1 \leqslant i \leqslant 2n+2$ we have

$$|\lambda_i(\tilde{A}_G) - \lambda_i(\tilde{A}_H)|^2 \leqslant \sum_{j=1}^{2n+2} |\lambda_j(\tilde{A}_G) - \lambda_j(\tilde{A}_H)|^2 \leqslant \|\tilde{A}_G - \tilde{A}_H\|_F^2 \leqslant 5/d.$$

Therefore we obtain $\sigma_2(\tilde{A}_H) \leqslant \sigma_2(\hat{A}_G) + \sqrt{5/d}$. □

Recall by the work Friedman [Fri08, Bor19] for $d \geqslant 3$ and sufficiently large n, there exists a d-regular $\left(\frac{2\sqrt{d-1}}{d} + o(1)\right)$-expander $G_{2n,d}$ on $2n$ vertices. Theorem 5.5 is immediate.

6 On approximability of the permanent of PSD matrices

6.1 Introduction

Given a matrix $A \in \mathbb{C}^{n \times n}$, the permanent of A is defined as

$$\mathrm{per}(A) = \sum_{\sigma \in \mathbb{S}_n} \prod_{i=1}^{n} A_{i,\sigma_i},$$

where the sum is over all permutations over n elements. It is well-known that the permanent of a matrix with non-negative entries can be approximated up to a $1+\varepsilon$-multiplicative factor using the MCMC method [JSV04]. Recently there has been significant interest in studying permanent of Hermitian PSD matrices because of close connections to quantum optics and Boson sampling. A folklore algorithm is to simply take the product of the entries of the main diagonal to get an $n!$-approximation.

A few years ago, [AGOS17] obtained the first (deterministic) simply exponential approximation algorithms with approximation factor $e^{(\gamma+1)n}$. The algorithm proposed in [AGOS17] uses a basic SDP relaxation for the problem; many experts expected that perhaps by using higher-level SDP relaxations one can improve the approximation factor. Later on, several groups attempted to improve the approximation factor (see e.g., [Bar20]), but for the general case, only subexponential improvements to the approximation ratio were found [YP21, YP22]. Very recently, Meiburg showed that contrarily to the permanent of non-negative matrices, it is NP-Hard to approximate the permanent of a PSD matrix within a factor of $e^{-n^{1-\varepsilon}}$ for any $\varepsilon > 0$ [Mei23]. So, the MCMC method falls short of providing a $1+\varepsilon$-approximation for PSD permanents.

It remained an open problem if, perhaps by using randomness or higher level SDP relaxations, one can obtain an $e^{-\varepsilon n}$ approximation factor for ε arbitrarily small, or at the very least whether the $\gamma + 1$ factor in the exponent can be improved to a smaller constant. We answer both these questions in our work.

Our first result is an exponential improvement on the $e^{-(\gamma+1)n}$ approximation algorithm mentioned above.

Theorem 6.1 (Main Algorithmic Result)**.** *There is a deterministic polynomial time $e^{-(\gamma+0.9999)n}$-approximation algorithm for the permanent of a Hermitian PSD matrix $A \in \mathbb{C}^{n \times n}$.*

Our second result is the first exponential hardness of approximation for this problem. As a corollary of a general hardness of approximation result we prove (see Theorem 6.5 below), we show the following:

Theorem 6.2 (Main Hardness Result)**.** *For all $\varepsilon > 0$, it is* NP*-hard to approximate the permanent of a Hermitian PSD matrix $A \in \mathbb{C}^{n \times n}$ within a factor $e^{-(\gamma-\varepsilon)n}$.*

In particular, the above theorem shows that assuming NP $\neq$ RP *even using randomness* the approximation factor of [AGOS17, YP21] cannot be improved by more than a factor of e^n.

Maximizing Product of Linear Forms. Our hardness techniques also apply to an optimization problem that happens to be related to the permanent of PSD matrices, called the "maximizing product of linear forms" problem, studied by Yuan and Parrilo [YP22, YP21]: Given a matrix $V \in \mathbb{C}^{n\times d}$ with rows $v_1, \ldots, v_n \in \mathbb{C}^d$, define

$$r(V) := \max_{x\in\mathbb{C}^n:\|x\|_2=1} \prod_{i=1}^{n} |\langle x, v_i\rangle|^2. \tag{6.1}$$

They design a polynomial time $O(e^{-\gamma n\cdot(1-o(1))})$-approximation algorithm for $r(V)$ using semidefinite programming, where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. They also prove APX-hardness for this problem, and raise an open problem of finding the true approximability of this problem. Recently, Meiberg studied an equivalent problem under the name Approximate Quantum Maximum Likelihood Estimation, and showed NP-hardness of approximating it to within any constant factor [Mei23]. Our main technical hardness result, Theorem 6.5, immediately implies that the maximizing product of linear forms problem (and the Approximate Quantum Maximum Likelihood Estimation problem) does not admit a $e^{-\gamma n(1+\varepsilon)}$-approximation for any constant $\varepsilon > 0$, answering the question of Yuan and Parrilo up to sub-exponential factors in n.

6.1.1 Technical Contributions

Algorithmic results

In this part, we show the main ideas behind Theorem 6.1. We start by presenting algorithms used by previous work. Let $A = VV^\dagger$ be an $n \times n$ PSD matrix where $v_1, \ldots, v_n \in \mathbb{C}^n$ are the rows of V. Previous work ([AGOS17, YP22]) showed that the value of the following SDP gives a $e^{-(\gamma+1)n}$ approximation to $\mathrm{per}(A)$:

$$\mathrm{SDP}(V) := \max_{X\succeq 0, \mathrm{tr}(X)=n} \prod_{i\in[n]} v_i^{\mathcal{D}} X v_i.$$

SDP(V) might seem completely unrelated to the definition of per(A), but we remark that their relationship is a lot more clear when per(A) is rewritten using Wick's formula (Lemma 6.23). Notice that the objective function of SDP(V) is log-concave, so it can be optimized in polynomial time. It turns out that upon solving SDP(V), we can reduce to the case that the maximizer X^* of SDP(V) satisfies $v_i^{\mathcal{D}} X^* v_i = 1$ for all $i \in [n]$ (see Eq. (6.6)). This property simplifies matters enough that we will assume it for the rest of this subsection.

The above property implies that $A \preceq I$ (see Claim 6.25), so we immediately get $\mathrm{per}(A) \leqslant 1$. Conversely, [AGOS17, YP22] prove that

$$\mathrm{per}(A) \geqslant \frac{n!}{n^n} \cdot r(V) \geqslant e^{-n} \cdot r(V). \tag{6.2}$$

Noticing that SDP(V) is a semidefinite relaxation of $r(V)$, a simple Gaussian rounding argument ([YP22, Lemma 4.3], Lemma 6.41) can be used to show

$$r(V) \geqslant e^{-\gamma n}\, \mathrm{SDP}(V) = e^{-\gamma n}. \tag{6.3}$$

Putting these together, one gets

$$e^{-(\gamma+1)n} \leqslant \mathrm{per}(A) \leqslant 1, \tag{6.4}$$

giving a $e^{-(\gamma+1)n}$ approximation factor.

We remark that both sides of the above inequality can be tight, in particular the upper bound is tight for the identity matrix and the lower bound is tight for a certain family of low rank projection matrices (see [AGOS17]). So one may expect that no improvement is possible along this line.

In our approach we exploit the fact these inequalities are tight for matrices of very different rank – the known tight examples for the upper/lower bounds have very high/low rank respectively. In order to make this intuition concrete, we will use $\mathrm{tr}(A)$ as a smooth analogue of rank. Our main technical results are improvements to both sides of Eq. (6.4).

Lemma 6.3 (Improved Upper Bound)**.** *Let $\varepsilon \in [0,1]$. Any matrix $0 \preceq A \preceq I$ with $\mathrm{tr}(A) \leqslant (1-\varepsilon)n$ satisfies*

$$\mathrm{per}(A) \leqslant \left(1 - \frac{\varepsilon^2}{20}\right)^n .$$

Lemma 6.4 (Improved Lower Bound)**.** *Let $VV^{\mathcal{D}} = A \preceq I$, and assume that the maximizer X^* of $\mathrm{SDP}(V)$ satisfies $v_i^{\mathcal{D}} X^* v_i = 1$ for all i. For any $0 \leqslant \beta \leqslant 1$,*

$$\mathrm{per}(A) \geqslant e^{-n} \cdot r(V) \geqslant e^{-(\gamma+1)n} \cdot \exp\left(n \cdot \left(\ln(1-\beta) + \frac{\beta}{1-\beta} \cdot \frac{\mathrm{tr}(A)}{n} - \frac{0.273\beta^2}{(1-\beta)^2} \cdot \frac{n}{\mathrm{tr}(A)}\right)\right).$$

Our proof of Lemma 6.3 is inspired by an identity for $\mathrm{per}(A)$ appearing in [Bar20]. Our proof of Lemma 6.4 is based on an improved rounding procedure and is more technical, so we provide a proof overview in Section 6.1.2. We can now state our algorithm.

Algorithm:. Given a PSD matrix $A = VV^\dagger$ where $v_1, \ldots, v_n$ are rows of V, first reduce to the case that the maximizer X^* of $\mathrm{SDP}(V)$ satisfies $v_i^\dagger X^* v_i = 1$ for all i (as described in Eq. (6.6)). Output $\left(1 - \frac{\varepsilon^2}{20}\right)^n$, where ε is defined by $\mathrm{tr}(A) = (1-\varepsilon)n$.

We will use this algorithm in our proof of Theorem 6.1, which is straightforward to analyze when equipped with Lemmas 6.3 and 6.4.

Hardness results In this part we highlight the main technical contributions behind Theorem 6.2. Our proof broadly consists of two steps:

1. Show that $r(V)$ does not admit a $e^{-\gamma n(1+\varepsilon)}$-approximation algorithm.
2. Give an approximation-preserving reduction from $r(V)$ to PSD permanents.

We start by elaborating on Item 1. In order to draw analogies to the existing hardness of approximation literature, we will first rephrase and generalize the optimization problem of $r(V)$. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ be a field. For a vector $x \in \mathbb{F}^n$ and $p \in \mathbb{R} - \{0\}$, define

$$\|x\|_p = \left(\mathop{\mathbb{E}}_i |x_i|^p\right)^{1/p} .$$

We will be particularly interested in the case that $p = 0$, for which we define $\|x\|_0 = \lim_{p\to 0}\|x\|_p$. It is not too hard to see that for any vector x, $\|x\|_0$ equals $\prod_{i\in[n]} |x_i|^{1/n}$, the geometric mean of the magnitude of the entries of x. Note that in the case that $p < 1$, $\|\cdot\|_p$ is not a norm and not convex, but we will nevertheless refer to it as the p-norm.

Given a matrix $A \in \mathbb{F}^{m\times n}$, the $p \to q$ "norm" of A is defined as

$$\|A\|_{p\to q} = \max_{x\in\mathbb{F}^n:\|x\|_p=1} \|Ax\|_q.$$

The connection of $\|A\|_{p\to q}$ to $r(V)$ is apparent: for any matrix $V \in \mathbb{C}^{n\times d}$, we have

$$r(V) = \|V\|_{2\to 0}^{2n}.$$

Over the last decade there has been significant interest in designing approximation algorithms or proving hardness of approximation for matrix $p \to q$ norms for $p, q \geqslant 1$ [BBH⁺12, BRS15, BGG⁺23]. Most notably, the $2 \to 4$ norm has been shown to be closely related to the Unique games and the small set expansion conjectures [BBH⁺12]. To the best of our knowledge, the problem is not well-studied when $q < 1$. We prove tight hardness of approximation (assuming P $\neq$ NP) for the $2 \to q$ norm when $-1 < q < 2$.

For $q > -1$ let

$$\gamma_{\mathbb{F},q} = \mathop{\mathbb{E}}_{g\sim_{\mathbb{F}}\mathcal{N}(0,1)}[|g|^q]^{1/q}$$

be the q-norm of a standard (real/complex) normal random variable. Bhattiprolu, Ghosh, Guruswami, Lee, Tulsiani [BGG⁺23] showed that for any $1 \leqslant q < 2$, and any $\varepsilon > 0$ it is NP-hard to approximate the $2 \to q$ norm of a real $m \times n$ matrix better than $\gamma_{\mathbb{R},p} + \varepsilon$, matching known semidefinite relaxation-based approximation algorithms [Ste05]. In our main theorem, we build on their techniques and we extend their result to all $-1 < q < 2$.

Theorem 6.5 (Main Technical Hardness Theorem)**.** *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. For all $-1 < q < 2$ and $\varepsilon > 0$, it is* NP*-hard to approximate $\|A\|_{2\to q}$ given a matrix $A \in \mathbb{F}^{m\times n}$ within a factor of $\gamma_{\mathbb{F},q} + \varepsilon$.*

For the sake of completeness, in Section 6.7 we write down a semidefinite relaxation of $\|A\|_{2\to q}$ for all $-1 < q < 2$ and prove that it gives a $\gamma_{\mathbb{F},q}$-approximation to $\|A\|_{2\to q}$, matching the above hardness result. As $r(V) = \|V\|_{2\to 0}^{2n}$, we also get that it is NP-hard to approximate $r(V)$ within a factor of $(\gamma_{\mathbb{C},0} + \varepsilon)^{2n} = e^{-\gamma n(1+\varepsilon)}$.

Next, we elaborate on Item 2 – an approximation preserving reduction from $r(V)$ to $\mathrm{per}(A)$. Our main observation is that the permanent of a highly rank-deficient $n \times n$ PSD matrix $A = VV^\dagger$ is essentially (up to subexponential error) the same as $r(V)$. This is a consequence of Wick's formula (Lemma 6.23), which allows us to view the permanent of a PSD matrix as a squared absolute moment of a complex multivariate Gaussian. As a result, we are able to use Theorem 6.5 to prove Theorem 6.2, which we do in Section 6.4.2.

6.1.2 Overview of the proof of Lemma 6.4

Let us start by explaining the proof of Eq. (6.3), which Lemma 6.4 improves upon. For any distribution $\mathcal{D}$ over $\mathbb{C}^n$ with $\mathbb{E}[\|x\|_2^2] = 1$, we have the bound

$$r(V) = \max_{\|x\|_2=1} \prod_{i\in[n]} |\langle v_i, x\rangle|^2 \geqslant \left(\frac{\mathbb{E}_{x\sim\mathcal{D}} \prod_{i\in[n]} |\langle v_i, x\rangle|^{2/n}}{\mathbb{E}_{x\sim\mathcal{D}} \|x\|_2^2}\right)^n = \mathop{\mathbb{E}}_{x\sim\mathcal{D}}\left[\prod_{i\in[n]} |\langle v_i, x\rangle|^{2/n}\right]^n.$$

Using Jensen's inequality on the RHS, we get

$$r(V) \geqslant \exp\left(\sum_{i\in[n]} \mathop{\mathbb{E}}_{x\sim\mathcal{D}} \ln |\langle v_i, x\rangle|^2\right). \tag{6.5}$$

The basic Gaussian rounding scheme picks $x \sim \mathcal{D} = \mathbb{C}\mathcal{N}(0, X^*)$ (see Definition 6.16 for a definition of the complex Gaussian distribution). Notice that for each i, $\langle v_i, x\rangle \sim \mathbb{C}\mathcal{N}(0, v_i^{\mathcal{D}} X^* v_i) = \mathbb{C}\mathcal{N}(0,1)$ by assumption of Lemma 6.4. Since $\mathbb{E}_{y\sim\mathbb{C}\mathcal{N}(0,1)} \ln |y|^2 = -\gamma$, we immediately get $r(V) \geqslant \exp(-n\gamma)$.

One can see that the analysis (in particular, the application of Jensen's inequality) is tight if $A = V = X^* = I$ for example. So to improve on Eq. (6.5), we must use a different rounding algorithm. Our first observation is that in the special case that $A = V = I$ where $v_i = e_i$, we can get the optimal lower bound by sampling independent Rademachers $s_1, \ldots, s_n \sim \{\pm 1\}$, and setting $x = \sum_{i\in[n]} s_i v_i$. With this choice, $\mathbb{E}_x \ln |\langle v_i, x\rangle|^2 = \mathbb{E}_x \ln 1 = 0$, implying $r(V) \geqslant 1$.

One could try to use a similar rounding scheme in the more general case of Lemma 6.4, i.e., when A is close to I in the sense that $\mathrm{tr}(A) \geqslant (1-\varepsilon)n$. Unfortunately this strategy ends up failing, as when the v_i's are not exactly orthogonal, there could be a nonzero probability that $\langle v_i, x\rangle = 0$, which would imply $\mathbb{E}_x \ln |\langle v_i, x\rangle|^2 = -\infty$. Note that if A is close to I, this singularity is a very small probability event for most of the vectors v_i, so it is natural to try to avoid it by adding some noise to x. We do this by interpolating between the two rounding schemes. We pick a parameter $0 < \beta < 1$, and set $x = \sqrt{1-\beta}\, g + \sqrt{\beta} \sum_{i\in[n]} s_i v_i$, up to some normalization, where $g \sim \mathbb{C}\mathcal{N}(0, X^*)$.

On the technical side, this interpolation helps us analyze $\mathbb{E}_x \ln |\langle v_i, x\rangle|^2$ in terms of tractable quantities. We use a sharp bound on the expected log of the magnitude squared of a noncentral complex Gaussian (see Lemma 6.21): for any $c \in \mathbb{C}$,

$$\mathop{\mathbb{E}}_{g\sim\mathbb{C}\mathcal{N}(0,1)} \ln |g + c|^2 \geqslant -\gamma + |c|^2 - \frac{|c|^4}{4}.$$

As a result of this inequality, when β is bounded away from 1, we can effectively bound $\mathbb{E}_x \ln |\langle v_i, x\rangle|^2$ using only the second and fourth moments of the random variable $\sum_{i\in[n]} s_i v_i$, which are tractable.

6.1.3 Overview of the proof of Theorem 6.5

As alluded to in Section 6.1.1, prior to our work, optimal hardness results for the $2 \to q$ norm are already established for $q \geqslant 1$. Our first observation is that these results [GRSW16, BRS15, BGG+23] can be extended to all $-1 < q < 2$ (see Theorem 6.28), or even more generally, to 2-concave f-means (see Definition 6.9).

In particular, one can deduce the following theorem.

Theorem 6.6 (Informal version of Theorem 6.28). *Let $q < 2$. Assume that there is a family $\{E_k\}$ of $k \times d_k$ gadget matrices such that $\|E_k\|_{2\to q} = 1$, but for all "smooth" unit vectors x ($\|x\|_\infty \ll 1$), $\|E_k x\|_q \leqslant \gamma$. Then for all $\varepsilon > 0$, it is NP-Hard to distinguish between the following two cases given a matrix $A : \mathbb{C}^m \to \mathbb{C}^n$ with $\|A\|_{2\to 2} \leqslant 1$.*

1. *Completeness: $\|A\|_{2\to q} = 1$, or*

2. *Soundness:* $\|A\|_{2\to q} \leqslant \gamma + \varepsilon$.

It remains to construct an appropriate family of gadget matrices $\{E_k\}$. We will use the following family, which was suggested in [BRS15]. For $k \geqslant 1$, define $E_k^{(\mathbb{C})} \in \mathbb{C}^{4^k \times k}$ as the matrix whose rows consist of the members of $\frac{1}{\sqrt{k}} \cdot \{-1, +1, -i, +i\}^k$ ordered arbitrarily.

It remains to show that the matrices $E_k^{(\mathbb{C})}$ satisfy the requirements of Theorem 6.6 with $\gamma \approx \gamma_{\mathbb{C},q}$. By construction, $\|E_k^{(\mathbb{C})}\|_{2\to q} = 1$.

To prove this, in Section 6.6 we prove a Berry-Esseen type result for test functions of the form $|x|^q$ for $q \neq 0$ and $\log|x|$ otherwise, applied to a sum of independent random variables. In particular, the special case of interest to us (for Theorem 6.2) is $q = 0$. In that case, the test function is $\log|x|$ which has a singularity at $x = 0$, but we are nevertheless one can bound the right hand side of the lemma below by an arbitrarily small quantity as $\delta \to 0$.

Lemma 6.7 (Informal version of Lemma 6.30)**.** *Let $-1 < q < 2$ and $0 < \delta < 1$. For all "smooth" unit vectors x with $\|x\|_\infty \leqslant \delta$,*

$$f(\|E_k x\|_q) - \gamma_{\mathbb{C},q} \lesssim -\int_0^\delta \min(0, f(u)) + \delta \cdot \left(\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right),$$

where $f(x) = |x|^q$ for $q \neq 0$, and $f(x) = \log|x|$ for $q = 0$.

6.1.4 Future Directions

The most exciting open problem is to determine the correct approximability for PSD permanents. Improving the hardness result seems to be out of reach of current techniques, but our ideas provide a clear path to improving the algorithmic result. Any significant improvement to the algorithm along the lines of our ideas would require significantly better versions of Lemmas 6.3 and 6.4. In particular, Lemma 6.3 is currently the bottleneck to a better approximation ratio, specifically the $O(\varepsilon^2)$ dependence. We conjecture that it can be improved to $O(\varepsilon)$, which would yield better approximation ratios as a corollary.

Although we don't have concrete new applications of hardness of approximation of $\|A\|_{2\to q}$ for $q \neq 0$, we expect to find further applications of the machinery developed here in addressing counting and optimization of linear algebraic problems, e.g., in estimating mixed discriminant, sub-determinant maximization, Nash-welfare maximization, etc.

6.1.5 Paper Organization

In Section 6.2, we present preliminary definitions and results that we will use. In Section 6.3, we prove Theorem 6.1. In Section 6.4, we prove Theorems 6.2 and 6.5 (with some components of the proof appearing in Sections 6.5 and 6.6).

6.2 Preliminaries

6.2.1 Generalized Means and Norms

Although we mostly use p-norms that are defined using an expectation, it will be convenient to also define the counting version of 2-norm, which we denote as ℓ_2.

Definition 6.8 (ℓ_2-norm, Frobenius norm)**.** Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. For a vector $x \in \mathbb{F}^n$, define $\|x\|_{\ell_2} = \sqrt{\sum_{i\in[n]} |x_i|^2}$. For a matrix $A \in \mathbb{F}^{m\times n}$, define $\|A\|_F = \sqrt{\sum_{i,j\in[n]} |A_{i,j}|^2}$.

We work with a generalization of $\|\cdot\|_p$ using the framework of "f-means".

Definition 6.9. Let $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ be continuous and injective. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ be a field. For a vector $x \in \mathbb{F}^n$, define

$$[x]_f^f := \mathop{\mathbb{E}}_{i\sim[n]} f(|x_i|),$$

$$[x]_f := f^{-1}([x]_f^f) = f^{-1}(\mathop{\mathbb{E}}_{i\sim[n]} f(|x_i|).$$

We will refer to $[x]_f$ as the *f-mean* of x. More generally, for a random variable X over $\mathbb{F}$ define

$$[X]_f^f = \mathbb{E} f(|X|), \quad [X]_f = f^{-1}(\mathbb{E} f(|X|)).$$

For a matrix $A \in \mathbb{F}^{n\times d}$, define

$$[A]_{2\to f} := \max_{x\in\mathbb{F}^n} \frac{[Ax]_f}{\|x\|_2}.$$

f-means provide a convenient and unified way to talk about $\|\cdot\|_p$, even for the case of $p = 0$.

Observation 6.10 (Power Means)**.** For all $p \in \mathbb{R}$, define

$$f_p(x) = \begin{cases} x^p & \text{if } p > 0, \\ \log x & \text{if } p = 0, \\ -x^p & \text{if } p < 0. \end{cases}$$

Then, we have $[x]_{f_p} = \|x\|_p$ for any $p \in \mathbb{R}$.

We note that the observation would still hold if we used x^p instead of $-x^p$ in the third case, but it will be convenient for f_p to always be an increasing function.

We will mostly be concerned with f-means that are dominated by the 2-norm. This happens exactly when f is 2-concave:

Definition 6.11. A function $f : \mathbb{R}_{>0} \to \mathbb{R}$ is 2-concave if $x \to f(\sqrt{x})$ is concave.

Example 6.12. For any $p \leqslant 2$, f_p is 2-concave.

We make some useful observations about 2-concave functions.

Claim 6.13. Let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be a 2-concave increasing function. Then,

1. $[x]_f \leqslant \|x\|_2$ for any vector x. Equivalently, $[x]_f^f \leqslant f(\|x\|_2)$.

2. $f'(x) \leqslant \frac{f'(y)}{y} \cdot x$ for $0 < y \leqslant x$.

3. $f(x) - f(y) \leqslant \frac{f'(y)}{2y} \cdot x^2$ for $0 < y \leqslant x$.

Proof. 1. Using Jensen's inequality on the concave function $x \to f(\sqrt{x})$,

$$[x]_f^f = \mathbb{E}\, f(|x_i|) = \mathbb{E}\, f\left(\sqrt{|x_i|^2}\right) \leqslant f(\sqrt{\mathbb{E}\,|x_i|^2}) = f(\|x\|_2).$$

2. Follows from the fact that $f'(\sqrt{x}) = \frac{f(\sqrt{x})}{2\sqrt{x}}$ is increasing in x.

3. Integrating the above from y to x,

$$f(x) - f(y) \leqslant \frac{f'(y)}{y} \cdot \frac{(x^2 - y^2)}{2} \leqslant \frac{f'(y)}{2y} \cdot x^2.$$

□

One could ask when $[x]_f$ is a homogeneous function of x. It turns out that this is exactly when $[x]_f$ is a p-norm.

Lemma 6.14 ([HLP52]). *$[\cdot]_f$ is 1-homogeneous (that is, $[ax]_f = a[x]_f$ for all scalars a and vectors x) if and only if it equals $[\cdot]_{f_p} = \|\cdot\|_p$ for some $p \in \mathbb{R}$.*

We will require another simple fact about f-means.

Fact 6.15. *Let V be a matrix $V \in \mathbb{F}^{n\times d}$ and integer $k > 0$. Then,*

$$[V^{(k)}]_{2\to f} := \left[\begin{bmatrix} V \\ \vdots \\ V \end{bmatrix}\right]_{2\to f} = [V]_{2\to f},$$

where V is copied k times in the right hand side.

Proof. For any vector $x \in \mathbb{F}^d$, consider the two vectors $z = Vx$ and $z^{(k)} = V^{(k)}x$. Note that a uniformly random entry of z has the same distribution as a random entry of $z^{(k)}$, so $[z]_f = [z^{(k)}]_f$. The claim follows from the definition of $[\cdot]_{2\to f}$. □

6.2.2 Gaussians

We will consider both real and complex Gaussians.

Definition 6.16. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. For the $n \times n$ identity matrix I_n, $\mathbb{F}\mathcal{N}(0, I_n)$ is defined to be the distribution over vectors $x \in \mathbb{F}^n$ given by the density function

$$p_{\mathbb{F}}(x) = \begin{cases} (2\pi)^{-n/2} \cdot \exp\left(-\|x\|_{\ell_2}^2/2\right) & \text{if } \mathbb{F} = \mathbb{R}, \\ \pi^{-n} \exp(-\|x\|_{\ell_2}^2) & \text{if } \mathbb{F} = \mathbb{C}. \end{cases}$$

More generally, given a positive semidefinite covariance matrix $\Sigma = AA^{\mathcal{D}}$ for $A \in \mathbb{F}^{n \times d}$, define $\mathbb{F}\mathcal{N}(0, \Sigma)$ to be distributed as Ax, where $x \sim \mathbb{F}\mathcal{N}(0, I_d)$. We will sometimes use $\mathcal{N}$ to denote $\mathbb{R}\mathcal{N}$.

More concretely, a complex Gaussian $g \sim \mathbb{C}\mathcal{N}(0, 1)$ can be sampled by sampling its real and imaginary parts independently from $\mathcal{N}(0, 1/2)$. There are formulas for the moments of univariate real and complex Gaussians in terms of the Gamma function.

Definition 6.17. For any $p \in \mathbb{R}$ and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, define $\gamma_{\mathbb{F},p} = [g]_{f_p}$, where g is a random variable distributed as $\mathbb{F}\mathcal{N}(0, 1)$.

Fact 6.18. *For any $p \in (-1, \infty) - \{0\}$,*

$$\gamma_{\mathbb{R},p}^p = \mathop{\mathbb{E}}_{g \sim \mathcal{N}(0,1)}[|g|^p] = \frac{2^{p/2} \cdot \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}},$$

and for any $p \in (-2, \infty) - \{0\}$,

$$\gamma_{\mathbb{C},p}^p = \mathop{\mathbb{E}}_{g \sim CN(0,1)}[|g|^p] = \Gamma\left(\frac{p}{2} + 1\right).$$

In particular, this implies

$$\gamma_{\mathbb{R},1} = \sqrt{\frac{2}{\pi}}, \quad \gamma_{\mathbb{C},1} = \sqrt{\frac{\pi}{2}}, \quad \gamma_{\mathbb{R},0} = \lim_{p \to 0} \gamma_{\mathbb{R},p} = \sqrt{\frac{e^{-\gamma}}{2}}, \quad \gamma_{\mathbb{C},0} = \lim_{p \to 0} \gamma_{\mathbb{C},p} = \sqrt{e^{-\gamma}}.$$

Fact 6.19 (Moment Generating Function of $|g|^2$). *Let $g \sim \mathbb{C}\mathcal{N}(0, 1)$. For any $t < 1$, $\mathbb{E}[e^{t|g|^2}] = (1 - t)^{-1}$.*

We will need sharp bounds on the expected value of $\ln |g + c|^2$ for Gaussian g and fixed c. First, we prove an estimate on the exponential integral function.

Fact 6.20. *For $x \geqslant 0$ it holds that*

$$(-x) = \int_{-\infty}^{-x} \frac{e^t}{t} dt = \gamma + \ln(x) + \sum_{k=1}^{\infty} \frac{(-x)^k}{k \cdot k!} \leqslant \gamma + \ln(x) - x + \frac{x^2}{4}.$$

Proof. The identity is due to Equation 5.1.11 in [AS48, **?**]. For the inequality, we must show that the function

$$f(x) = \sum_{k=3}^{\infty} \frac{(-x)^k}{k \cdot k!}$$

is nonpositive for $x \geqslant 0$. To do this, observe first that $f(0) = 0$, and

$$\begin{aligned} f'(x) &= \sum_{k=3}^{\infty} \frac{(-1)^k \cdot x^{k-1}}{k!} \\ &= \frac{1}{x} \sum_{k=3}^{\infty} \frac{(-x)^k}{k!} \\ &= \frac{e^{-x} - \left(1 - x + \frac{x^2}{2}\right)}{x} \\ &\leqslant 0. && (e^{-x} \leqslant 1 - x + x^2/2 \text{ for } x \geqslant 0) \end{aligned}$$

Therefore, $f(x) \leqslant 0$ for $x \geqslant 0$. □

Lemma 6.21. *Let $c \in \mathbb{C}$. Then, $\mathbb{E}_{g \sim \mathbb{C}\mathcal{N}(0,1)}[\ln |g + c|^2] \geqslant -\gamma + |c|^2 - |c|^4/4$.*

Proof. Define $x = |c|^2$. By [Mos20, Eqn. 35, Thm. 1] we have the identity

$$\mathop{\mathbb{E}}_{g \sim \mathbb{C}\mathcal{N}(0,1)}[\ln |g + c|^2] = - \ln(x) - (x).$$

By Fact 6.20, this is at least $-\gamma + x - x^2/4$, as desired. □

6.2.3 Permanent

For a matrix $A \in \mathbb{C}^{n \times n}$, its permanent is defined as

$$\mathrm{per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)}.$$

On the domain of positive semidefinite matrices, the permanent has some nice properties. For example, it is monotone w.r.t. the Loewner order.

Lemma 6.22 (e.g., [AGOS17])**.** *If $A, B \in \mathbb{C}^{n \times n}$ are hermitian and $A \succeq B \succeq 0$, then*

$$\mathrm{per}(A) \geqslant \mathrm{per}(B).$$

Proof Sketch. The statement of the lemma follows, because $A \succeq B \succeq 0$ implies that $A^{\otimes n} \succeq B^{\otimes n} \succeq 0$. So, if 1_{S_n} is the indicator vector of all permutations in $\mathbb{R}^{n \otimes n}$,

$$\mathrm{per}(A) = \frac{1}{n!} 1_{S_n}^{\mathcal{D}} A^{\otimes n} 1_{S_n} \geqslant \frac{1}{n!} 1_{S_n}^{\mathcal{D}} B^{\otimes n} 1_{S_n} = \mathrm{per}(B)$$

as desired. □

Lemma 6.23. *For any PSD matrix $VV^{\mathcal{D}}$ with $V \in \mathbb{R}^{n \times d}$, we have*

$$\mathop{\mathbb{E}}_{x \sim \mathcal{N}(0,I)} \left[\prod_{i \in [n]} |\langle v_i, x \rangle|^2 \right] = c_{n,d}^{\mathbb{R}} \cdot \mathop{\mathbb{E}}_{x \in \mathbb{R}^d, \|x\|_2 = 1} \left[\prod_{i \in [n]} |\langle v_i, x \rangle|^2 \right].$$

For any PSD matrix $VV^{\mathcal{D}}$ with $V \in \mathbb{C}^{n\times d}$, we have

$$\operatorname{per}(VV^{\mathcal{D}}) = \mathop{\mathbb{E}}_{x\sim\mathbb{C}\mathcal{N}(0,I)}\left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right] = c^{\mathbb{C}}_{n,d}\cdot \mathop{\mathbb{E}}_{x\in\mathbb{C}^d,\|x\|_2=1}\left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right].$$

Here, $v_1, \ldots, v_n$ are the rows of V. The proportionality constants above are defined by

$$c^{\mathbb{R}}_{n,d} = \frac{\Gamma(n+d/2)}{\Gamma(d/2)\cdot(d/2)^n}, \qquad c^{\mathbb{C}}_{n,d} = \frac{(d+n-1)!}{(d-1)!\cdot d^n}.$$

Proof. The first equality in the second conclusion follows from Isserlis' theorem/Wick's formula (see 3.1.4 in [Bar16]).

We prove the other equality in the real case, but the complex case can be proved similarly. Observe that

$$\begin{aligned}\mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,I)}\left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right] &= \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,I)}\|x\|_2^{2n}\cdot\left[\prod_{i\in[n]}\left|\left\langle v_i, \frac{x}{\|x\|_2}\right\rangle\right|^2\right] \\ &= \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,I)}\|x\|_2^{2n}\cdot \mathop{\mathbb{E}}_{x\in\mathbb{R}^n,\|x\|_2=1}\left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right] \\ &= d^{-n}\mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,I)}\|x\|_{\ell_2}^{2n}\cdot \mathop{\mathbb{E}}_{x\in\mathbb{R}^n,\|x\|_2=1}\left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right]\end{aligned}$$

The first identity uses that for $x\sim\mathcal{N}(0,I)$, $\|x\|_2$ is independent from $x/\|x\|_2$. The second identity uses that fact that if $x\sim\mathcal{N}(0,I)$, then $x/\|x\|_2$ is distributed uniformly on a sphere of radius $\|x\|_2$. To conclude the proof, notice that $\mathbb{E}_{x\sim\mathcal{N}(0,I)}\|x\|_{\ell_2}^{2n}$ is the n^{th} moment of a chi-squared random variable with d-degrees of freedom, which is $2^n\frac{\Gamma(n+d/2)}{\Gamma(d/2)}$. □

We will also require the following formula for the permanent of the sum of two matrices.

Lemma 6.24 ([Per12, Page 2]). *For any two matrices $A, B \in \mathbb{C}^{n\times n}$,*

$$\operatorname{per}(A+B) = \sum_{S,T\subseteq[n],|S|=|T|}\operatorname{per}(A_{S,T})\cdot\operatorname{per}(B_{\bar{S},\bar{T}}).$$

When $A = I$, this simplifies to

$$\operatorname{per}(I+B) = \sum_{S\subseteq[n]}\operatorname{per}(B_{S,S}).$$

Above, $A_{S,T}$ is the $|S|\times|T|$ submatrix of A containing rows only in S and columns only in T.

6.3 Algorithm

We start by expanding on the basic setup of the algorithms of [AGOS17, YP22], which we briefly introduced in Section 6.1.1. After this, we will show how Lemmas 6.3 and 6.4 imply Theorem 6.1. Later on, in Sections 6.3.1 and 6.3.2 respectively, we prove Lemmas 6.3 and 6.4.

Let $A = VV^{\mathcal{D}}$ be the PSD matrix whose permanent we wish to compute. Let $v_1, \ldots, v_n \in \mathbb{C}^n$ be the rows of V, so by Lemma 6.23,

$$\operatorname{per}(A) = \mathop{\mathbb{E}}_{x \sim \mathbb{C}\mathcal{N}(0,I)} \left[\prod_{i \in [n]} |\langle x, v_i \rangle|^2 \right].$$

Recall the log-concave maximization problem $\mathrm{SDP}(V)$ we associated with this problem:

$$\mathrm{SDP}(V) = \max_{X : X \succeq 0, \operatorname{tr}(X) = n} \prod_{i \in [n]} v_i^{\mathcal{D}} X v_i.$$

Let X^* be the optimal solution to $\mathrm{SDP}(V)$. Note that X^* can be found efficiently. It will be convenient to make a simplification to our problem. We will replace the matrix A by $\tilde{A} = D^{-1/2} A D^{-1/2}$, where D is a positive semidefinite diagonal matrix defined as $D_{i,i} = v_i^{\mathcal{D}} X^* v_i$. Since D is diagonal,

$$\operatorname{per}(A) = \operatorname{per}(\tilde{A}) \cdot \operatorname{per}(D) = \operatorname{per}(\tilde{A}) \cdot \mathrm{SDP}(V),$$

so it suffices to approximate $\operatorname{per}(\tilde{A})$ instead of $\operatorname{per}(A)$. Writing $\tilde{A} = \tilde{V}\tilde{V}^{\dagger}$ for $\tilde{V} = D^{-1/2}V$, we can see that the objective functions of $\mathrm{SDP}(V)$ and $\mathrm{SDP}(\tilde{V})$ are positive scalar multiples of each other, so $\mathrm{SDP}(\tilde{V})$ is also maximized by X^*. Note that $\tilde{A}$ enjoys the additional property $\tilde{v}_i^{\mathcal{D}} X^* \tilde{v}_i = 1$ for all $i \in [n]$, where $\tilde{v}_i = D^{-1/2} v_i$. Replacing A by $\tilde{A}$, we will henceforth assume that the maximizer X^* of $\mathrm{SDP}(V)$ satisfies

$$v_i^{\mathcal{D}} X^* v_i = 1 \text{ for all } i \in [n]. \tag{6.6}$$

In particular, this implies $\mathrm{SDP}(V) = 1$. Under this assumption, A satisfies an important property.

Claim 6.25. We have $A \preceq I$.

Proof. Let $f(X) = \prod_{i \in [n]} v_i^{\mathcal{D}} X v_i$ be the objective function of $\mathrm{SDP}(V)$. We can compute

$$\nabla(\ln f)(X) = \sum_{i \in [n]} \frac{v_i v_i^{\mathcal{D}}}{v_i^{\mathcal{D}} X v_i}.$$

In particular, by Eq. (6.6), $\nabla(\ln f)(X^*) = \sum_{i \in [n]} v_i v_i^{\mathcal{D}} = V^{\mathcal{D}} V$.

The optimality conditions for X^* imply that for all symmetric matrices M with $\operatorname{tr}(M) = 0$ and $W_- \subseteq \operatorname{Range}(X^*)$ it holds that

$$\langle V^{\mathcal{D}} V, M \rangle = \langle \nabla(\ln f)(X^*), M \rangle \leqslant 0.$$

Here, W_- denotes the vector space spanned by the negative eigenvectors of M. Now, let $Q \succeq 0$ be an arbitrary PSD matrix, and set $M = Q - \frac{\operatorname{tr}(Q)}{n} X^*$. M satisfies both the conditions above, and therefore we have

$$0 \geqslant \langle V^{\mathcal{D}} V, M \rangle$$

$$
\begin{aligned}
&= \langle V^{\mathcal{D}}V, Q\rangle - \frac{\mathrm{tr}(Q)}{n}\langle V^{\mathcal{D}}V, X^*\rangle \\
&= \langle V^{\mathcal{D}}V, Q\rangle - \frac{\mathrm{tr}(Q)}{n}\sum_{i\in n} v_i^{\mathcal{D}} X^* v_i \\
&= \langle V^{\mathcal{D}}V, Q\rangle - \mathrm{tr}(Q). && \text{(by Eq. (6.6))}
\end{aligned}
$$

In other words, $\langle V^{\mathcal{D}}V, Q\rangle \leqslant \mathrm{tr}(Q)$ for all $Q \succeq 0$, implying $V^{\mathcal{D}}V \preceq I$. Therefore $A = VV^{\mathcal{D}} \preceq I$. □

Claim 6.25 immediately implies $\mathrm{per}(A) \leqslant 1$. In [AGOS17, YP22], the authors prove the complimentary inequalities

$$
\mathrm{per}(A) \geqslant \frac{n!}{n^n}\cdot r(V) \geqslant \exp(-\gamma n)\cdot\frac{n!}{n^n}\cdot \mathrm{SDP}(V) = \exp(-\gamma n)\cdot\frac{n!}{n^n} \gtrsim \exp(-(\gamma+1)n), \tag{6.7}
$$

and together, the two inequalities above provide a $e^{-(1+\gamma)n}$ approximation for $\mathrm{per}(A)$.

Recall Lemmas 6.3 and 6.4, which (under Claim 6.25) improve the above inequalities to

$$
e^{-(\gamma+1)n}\cdot\exp\left(n\cdot\ell\left(\frac{\mathrm{tr}(A)}{n}\right)\right) \leqslant \mathrm{per}(A) \leqslant \exp\left(n\cdot r\left(\frac{\mathrm{tr}(A)}{n}\right)\right). \tag{6.8}
$$

Here, $\ell(x) = \max_{0\leqslant\beta\leqslant1}\ln(1-\beta) + \frac{\beta x}{(1-\beta)} - \frac{0.273\beta^2}{(1-\beta)^2 x}$, and $r(x) = \ln\left(1-\frac{(1-x)^2}{20}\right)$. We are now ready to prove Theorem 6.1.

Proof of Theorem 6.1. Let $A = VV^{\mathcal{D}} \succeq 0$, where V has rows $v_1,\ldots,v_n$. Our algorithm will first solve $\mathrm{SDP}(V)$ and use Eq. (6.6) and Claim 6.25 to reduce to the case that $0 \preceq A \preceq I$ and $v_i^{\mathcal{D}} X^* v_i = 1$ for all i, where X^* is the optimal solution to $\mathrm{SDP}(V)$. We will then output $\exp\left(n\cdot r\left(\frac{\mathrm{tr}(A)}{n}\right)\right) = \left(1-\frac{\varepsilon^2}{20}\right)^n$.

Eq. (6.8) implies that the approximation factor of this algorithm is at least $e^{-(\gamma+1-\alpha)n}$, where α is the minimum value of $r(x)-\ell(x)$ over all $x\in[0,1]$. Write $\ell(x) \geqslant \ell'(x) := \max(0, \ln(1-\beta^*)+\frac{\beta^* x}{(1-\beta^*)}-\frac{0.273(\beta^*)^2}{(1-\beta^*)^2 x}$ for $\beta^* = 0.34$. One can numerically determine that $\alpha \geqslant \min_{0\leqslant x\leqslant 1} r(x)-\ell'(x) \geqslant 10^{-4}$. □

6.3.1 Proof of Lemma 6.3

We first prove an inequality that we will require. The proof of this inequality is inspired by an identity of Barvinok [Bar20].

Lemma 6.26. *For any matrix* $0 \preceq B \prec I$, $\mathrm{per}(I+B) \leqslant \det\left((I-B)^{-1}\right)$.

Proof. Write $B = VV^{\mathcal{D}}$ for $V \in \mathbb{C}^{\times n}$. Let $v_1,\ldots,v_n$ be the rows of V.

$$
\begin{aligned}
\mathrm{per}(I+B) &= \sum_{S\subseteq[n]} \mathrm{per}(B_{S,S}) && \text{(Lemma 6.24)} \\
&= \sum_{S\subseteq[n]} \mathop{\mathbb{E}}_{g\sim\mathbb{C}\mathcal{N}(0,I)}\left[\prod_{i\in S}|\langle v_i, g\rangle|^2\right] && \text{(Lemma 6.23)}
\end{aligned}
$$

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{g\sim\mathbb{C}\mathcal{N}(0,I)}\left[\prod_{i\in[n]}\left(1+|\langle v_i, g\rangle|^2\right)\right] \\
&\leqslant \mathop{\mathbb{E}}_{g\sim\mathbb{C}\mathcal{N}(0,I)}\left[\prod_{i\in[n]} e^{|\langle v_i, g\rangle|^2}\right] && (1+x\leqslant e^x \text{ for all } x) \\
&= \mathop{\mathbb{E}}_{g\sim\mathbb{C}\mathcal{N}(0,I)}\left[\exp\left(g^{\mathcal{D}}V^{\mathcal{D}}Vg\right)\right].
\end{aligned}
$$

Let $\sigma_1, \dots, \sigma_n$ be the eigenvalues of $V^{\mathcal{D}}V$. Since g is invariant under unitary transformations, we can rotate g into the eigenbasis of $V^{\mathcal{D}}V$ to get that

$$
\begin{aligned}
\mathrm{per}(I+B) &\leqslant \mathop{\mathbb{E}}_{g\sim\mathbb{C}\mathcal{N}(0,I)}\left[\exp\left(\sum_{i\in[n]}\sigma_i|g_i|^2\right)\right] \\
&= \prod_{i\in[n]}\mathop{\mathbb{E}}_{g\sim\mathbb{C}\mathcal{N}(0,1)}\left[e^{\sigma_i|g|^2}\right] && (\text{Independence of } g_i) \\
&= \prod_{i\in[n]}\frac{1}{1-\sigma_i}. && (\text{Fact 6.19})
\end{aligned}
$$

Noting that the eigenvalues of $V^{\mathcal{D}}V$ match those of $B = VV^{\mathcal{D}}$, this is equal to $\det\left((I-B)^{-1}\right)$. □

Now, we are ready to prove Lemma 6.3. Let $0 \preceq A \preceq I$ be a matrix with $\mathrm{tr}(A) \leqslant (1-\varepsilon)n$. Let $0 \leqslant \lambda_1 \leqslant \dots \leqslant \lambda_n \leqslant 1$ be the eigenvalues of A, and let $v_1, \dots, v_n$ be the corresponding eigenvectors. Let $t \in (1/2, 1]$ be a parameter we will set later, and let i_t be the smallest index i such that $\lambda_i > t$. For any parameter $t \in (1/2, 1]$, we can write

$$
A \preceq tI + \sum_{i\geqslant i_t}(\lambda_i - t)v_i v_i^{\mathcal{D}} = t\cdot\left(I + \sum_{i\geqslant i_t}\frac{\lambda_i - t}{t}v_i v_i^{\mathcal{D}}\right).
$$

Write $B = \sum_{i\geqslant i_t}\frac{\lambda_i - t}{t}v_i v_i^{\mathcal{D}}$. Since $t > 1/2$, $B \prec I$, so it satisfies the conditions of Lemma 6.26.

$$
\begin{aligned}
\mathrm{per}(A) &\leqslant t^n \cdot \mathrm{per}(I+B) && (\text{Lemma 6.22}) \\
&\leqslant \frac{t^n}{\det(I-B)}. && (\text{Lemma 6.26})
\end{aligned}
$$

We now pick $t = 1-\varepsilon/5$. For this choice of t, we must have $i_t \geqslant \frac{\varepsilon n}{2}$, since otherwise, $\mathrm{tr}(A) \geqslant t\cdot(n-i_t) \geqslant (1-\varepsilon/5)\cdot(1-\varepsilon/2)n > (1-\varepsilon)n$ contradicts the fact that $\mathrm{tr}(A) \leqslant (1-\varepsilon)n$. We compute

$$
\det(I-B) = \prod_{i\geqslant i_t}\left(1-\frac{\lambda_i - t}{t}\right) = \prod_{i\geqslant i_t}\left(2-\frac{\lambda_i}{t}\right) \geqslant \left(2-\frac{1}{t}\right)^{n-i_t}.
$$

Plugging in the definition of t and our lower bound on i_t, this is at least

$$
\left(2-\frac{1}{1-\varepsilon/5}\right)^{(1-\varepsilon/2)n} = \left(\frac{1-2\varepsilon/5}{1-\varepsilon/5}\right)^{(1-\varepsilon/2)n}.
$$

We now have our upper bound on $\mathrm{per}(A)$:

$$\mathrm{per}(A) \leqslant (1-\varepsilon/5)^n \cdot \left(\frac{1-\varepsilon/5}{1-2\varepsilon/5}\right)^{(1-\varepsilon/2)n} = \left(\frac{(1-\varepsilon/5)\cdot(1-\varepsilon/5)^{1-\varepsilon/2}}{(1-2\varepsilon/5)^{1-\varepsilon/2}}\right)^n .$$

To complete the proof, we use that $\frac{(1-\varepsilon/5)\cdot(1-\varepsilon/5)^{1-\varepsilon/2}}{(1-2\varepsilon/5)^{1-\varepsilon/2}} \leqslant 1-\frac{\varepsilon^2}{20}$ for all $\varepsilon \in [0,1]$.

6.3.2 Proof of Lemma 6.4

By Eq. (6.7), it suffices to prove a lower bound on $r(V)$. Let X^* be the optimal solution to $\mathrm{SDP}(V)$. Consider the following randomized rounding scheme to a solution of $r(V)$: sample $g \sim \mathbb{C}\mathcal{N}(0, X^*)$ and $s_i \sim \{z \in \mathbb{C} : |z| = 1\}$ independently for all $i \in [n]$. Let $x = \sqrt{1-\beta}g + \sqrt{\frac{\beta n}{\mathrm{tr}(A)}} \sum_{i\in[n]} s_i v_i$. We will use the bound

$$r(V)^{1/n} = \max_{\|x\|_2^2=1} \prod_{i\in[n]} |\langle v_i, x\rangle|^{2/n} \geqslant \frac{\mathbb{E}_x[\prod_{i\in n} |\langle v_i, x\rangle|^{2/n}]}{\mathbb{E}_x[\|x\|_2^2]}.$$

First we compute the denominator.

$$\begin{aligned}
n \cdot \mathbb{E}[\|x\|_2^2] &= n \cdot \mathbb{E}[\|x\|_2^2] \\
&= \mathbb{E}\left[\left\|\sqrt{1-\beta}g + \sqrt{\frac{\beta n}{\mathrm{tr}(A)}} \sum_{i\in[n]} s_i v_i\right\|_{\ell_2}^2\right] \\
&= (1-\beta)\,\mathbb{E}[\|g\|_{\ell_2}^2] + \frac{\beta n}{\mathrm{tr}(A)}\,\mathbb{E}\left[\sum_{i,j} s_i s_j \langle v_i, v_j\rangle\right] + \sqrt{\frac{\beta n}{\mathrm{tr}(A)}(1-\beta)}\,\mathbb{E}\left[\left\langle g, \sum_i s_i v_i\right\rangle\right] \\
&= (1-\beta)\,\mathbb{E}[\|g\|_{\ell_2}^2] + \frac{\beta n}{\mathrm{tr}(A)} \sum_i \|v_i\|_{\ell_2}^2 && \text{(Independence)} \\
&= (1-\beta)\cdot \mathrm{tr}(X^*) + \frac{\beta n}{\mathrm{tr}(A)} \cdot \sum_{i\in[n]} \|v_i\|_{\ell_2}^2 && (g \sim \mathbb{C}\mathcal{N}(0, X^*), \text{ definition of } \|\cdot\|_{\ell_2}) \\
&= n. && (\mathrm{tr}(X^*) = n, \textstyle\sum_{i\in[n]} \|v_i\|_{\ell_2}^2 = \mathrm{tr}(VV^{\mathcal{D}}) = \mathrm{tr}(A))
\end{aligned}$$

So, $\mathbb{E}[\|x\|_2^2] = 1$. It remains to lower bound the numerator. We start by applying Jensen's inequality to get

$$\mathop{\mathbb{E}}_x\left[\prod_{i\in n} |\langle v_i, x\rangle|^{2/n}\right] \geqslant \exp\left(\frac{1}{n}\sum_{i\in[n]} \mathop{\mathbb{E}}_x[\ln|\langle v_i, x\rangle|^2]\right). \tag{6.9}$$

We will bound each of the terms inside the sum. Fix some $i \in [n]$, and let $y_i = \sqrt{\frac{n}{\mathrm{tr}(A)}} \sum_{j\in[n]} s_j \langle v_i, v_j\rangle$ and $z_i = \langle g, v_i\rangle$, so $\langle v_i, x\rangle = \sqrt{1-\beta}z_i + \sqrt{\beta}y_i$. Notice that $z_i \sim \mathbb{C}\mathcal{N}(0, v_i^{\mathcal{D}} X^* v_i) = \mathbb{C}\mathcal{N}(0,1)$ by assumption. Let us bound

$$\mathbb{E}[\ln|\langle v_i, x\rangle|^2] = \mathbb{E}[\ln|\sqrt{1-\beta}z_i + \sqrt{\beta}y_i|^2]$$

$$= \ln(1-\beta) + \mathbb{E}\left[\ln\left|z_i + \sqrt{\frac{\beta}{1-\beta}} y_i\right|^2\right]$$

$$\geqslant -\gamma + \ln(1-\beta) + \frac{\beta}{1-\beta}\mathbb{E}[|y_i|^2] - \frac{\beta^2}{4(1-\beta)^2}\mathbb{E}[|y_i|^4].$$

(Lemma 6.21, $z_i \sim \mathbb{C}\mathcal{N}(0,1)$ and is independent of y_i)

We bound the second and fourth moments of y_i using the below claim, whose proof we defer to Section 6.3.2.

Claim 6.27. For all $i \in [n]$,

$$\mathbb{E}[|y_i|^2] = \frac{n}{\operatorname{tr}(A)} v_i^{\mathcal{D}} V^{\mathcal{D}} V v_i,$$

$$\mathbb{E}[|y_i|^4] \leqslant \frac{1.09n^2}{\operatorname{tr}(A)^2}\|v_i\|_{\ell_2}^2.$$

Plugging in the bounds from Claim 6.27 and summing over all i, we get

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{E}_x[\ln|\langle v_i, x\rangle|^2] \geqslant -\gamma + \ln(1-\beta) + \frac{\beta}{(1-\beta)\operatorname{tr}(A)}\sum_{i\in[n]} v_i^{\mathcal{D}} V^{\mathcal{D}} V v_i - \frac{0.273\beta^2 n}{(1-\beta)^2\operatorname{tr}(A)^2}\sum_{i\in[n]}\|v_i\|_{\ell_2}^2$$

$$= -\gamma + \ln(1-\beta) + \frac{\beta}{1-\beta}\cdot\frac{\|A\|_F^2}{\operatorname{tr}(A)} - \frac{0.273\beta^2}{(1-\beta)^2}\cdot\frac{n}{\operatorname{tr}(A)}$$

$$\geqslant -\gamma + \ln(1-\beta) + \frac{\beta}{1-\beta}\cdot\frac{\operatorname{tr}(A)}{n} - \frac{0.273\beta^2}{(1-\beta)^2}\cdot\frac{n}{\operatorname{tr}(A)}$$

($\|A\|_F^2 \geqslant \frac{\operatorname{tr}(A)^2}{n}$ by Jensen's inequality)

This completes the proof of Lemma 6.4.

Proof of Claim 6.27

Proof. We can directly compute

$$\frac{\operatorname{tr}(A)}{n}\mathbb{E}[|y_i|^2] = \sum_{j,k\in[n]} \mathbb{E}[s_j\overline{s_k}]\cdot\langle v_i, v_j\rangle\overline{\langle v_i, v_k\rangle}$$

$$= \sum_{j\in[n]} |\langle v_i, v_j\rangle|^2 = v_i^{\mathcal{D}} V^{\mathcal{D}} V v_i \qquad (s_j \text{ is independent from } s_k \text{ for } j \neq k)$$

Similarly,

$$\frac{\operatorname{tr}(A)^2}{n^2}\mathbb{E}[|y_i|^4] = \sum_{j,k,l,m\in[n]} \mathbb{E}[s_j s_k \overline{s_l s_m}]\langle v_i, v_j\rangle\langle v_i, v_k\rangle\overline{\langle v_i, v_l\rangle\langle v_i, v_m\rangle}$$

$$
\begin{aligned}
&= \sum_{j\in[n]} |\langle v_i, v_j\rangle|^4 + 2\sum_{j\neq k} |\langle v_i, v_j\rangle\langle v_i, v_k\rangle|^2 \\
&= 2\left(\sum_{j\in[n]} |\langle v_i, v_j\rangle|\right)^2 - \sum_{j\in[n]} |\langle v_i, v_j\rangle|^4 \\
&= 2(v_i^{\mathcal{D}} V^{\mathcal{D}} V v_i)^2 - \sum_{j\in[n]} |\langle v_i, v_j\rangle|^4 \\
&\leqslant 2\|v_i\|_{\ell_2}^4 - \|v_i\|_{\ell_2}^8 && (V^{\mathcal{D}} V \preceq I, \text{ since } A = VV^{\mathcal{D}} \preceq I) \\
&\leqslant 1.09 \cdot \|v_i\|_{\ell_2}^2 && (x^2 - x^4 \leqslant 1.09x \text{ for } x \geqslant 0)
\end{aligned}
$$

The second equality is because $\mathbb{E}[s_j s_k \overline{s_l s_m}] = 0$ unless each index appears an equal number of times in $\{j, k\}$ and $\{l, m\}$. □

6.4 Hardness of Approximation

As mentioned in Section 6.1.1, we will first prove Theorem 6.5. Later on, in Section 6.4.2, we will use Theorem 6.5 to prove Theorem 6.2 using an approximation-preserving reduction to the permanent problem.

Our first result is a general inapproximability result for the f-mean version of $\|A\|_{2\to q}$ that is dependent on an appropriate family of gadgets $\{E_k\}$ as defined below.

Theorem 6.28. *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and let $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ be a continuous increasing 2-concave function such that $\lim_{x\to\infty} \frac{f(x)}{x^2} = 0$. Let $\delta, \gamma > 0$. Assume that for all k, there is a matrix $E_k : \mathbb{F}^k \to \mathbb{F}^{d_k}$ satisfying the following:*

1. $\|E_k\|_{2\to 2} = 1$.
2. *The entries of E_k have magnitude equal to $\frac{1}{\sqrt{k}}$.*
3. *for all vectors $x \in \mathbb{F}^k$ with $\|x\|_\infty \leqslant \delta \cdot \|x\|_{\ell_2}$, $[E_k x]_f \leqslant \gamma \cdot \|x\|_2$.*

Then for all $\varepsilon > 0$, it is NP-Hard to distinguish between the following two cases given a matrix $A : \mathbb{F}^m \to \mathbb{F}^n$ with $\|A\|_{2\to 2} \leqslant 1$.

1. *Completeness: $[A]_{2\to f} = 1$, or*
2. *Soundness: $[A]_{2\to f} \leqslant \gamma + \varepsilon$.*

The proof of the above theorem is in Section 6.4.1 and closely follows the arguments used in [BRS15, BGG+23]. In order to instantiate it, we will have to construct a family of gadgets $\{E_k\}$ that have $\|E_k\|_{2\to 2} = 1$, but at the same time have small $2 \to f$-norm when restricted to "smooth" vectors.

Definition 6.29. For $k \geqslant 1$, let us define $E_k^{(\mathbb{R})} \in \mathbb{R}^{2^k \times k}$ as the matrix whose rows consist of the members of $\frac{1}{\sqrt{k}} \cdot \{-1, +1\}^k$ ordered arbitrarily. Similarly, we define $E_k^{(\mathbb{C})} \in \mathbb{C}^{4^k \times k}$ as the matrix whose rows consist of the members of $\frac{1}{\sqrt{k}} \cdot \{-1, +1, -i, +i\}^k$ ordered arbitrarily.

Observe that these matrices are normalized so that $\|E_k^{(\mathbb{F})}\|_{2\to 2} = 1$. The following Lemma shows that condition 3 of Theorem 6.28 is satisfied with $\gamma \approx \gamma_{\mathbb{F},p}$.

Lemma 6.30. *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let f be an absolutely continuous 2-concave increasing function. Let $x \in \mathbb{F}^k$, and $E = E_k^{(\mathbb{F})}$. For all $0 < \delta < 1$, if $\|x\|_\infty \leqslant \delta \|x\|_{\ell_2}$ then*

$$\left[\frac{Ex}{\|x\|_2}\right]_f^f \leqslant [g]_f^f + C \cdot \left(-\int_0^{C\delta} \min(0, f(u)) + \delta \cdot \left(\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1) \right) \right),$$

where $g \sim \mathbb{F}\mathcal{N}(0,1)$ and $C > 0$ is a universal constant. In particular if $f = f_p$ for some $-1 < p < 2$, we have

$$[Ex]_{f_p} \leqslant \|x\|_2 \cdot (\gamma_{\mathbb{F},p} + \varepsilon_\delta),$$

where $\varepsilon_\delta \to 0$ as $\delta \to 0$.

We prove Lemma 6.30 in Section 6.5. The proof requires a Berry-Esseen type result for test functions of the form $f(\|.\|_2)$ applied to a sum of independent random vectors, which we prove in Section 6.6.

With these results in hand, we can now prove Theorem 6.5.

Proof of Theorem 6.5. We pick δ to be such that Lemma 6.30 implies $\|Ex\|_q \leqslant \|x\|_2 \cdot (\gamma_{\mathbb{F},q} + \varepsilon/2)$ for all x satisfying $\|x\|_\infty \leqslant \delta \|x\|_{\ell_2}$.

We apply Theorem 6.28 to the increasing 2-concave function $f = f_p$, gadget family $\{E_k^{(\mathbb{F})}\}$, and parameters δ, $\gamma = \gamma_{\mathbb{F},q} + \varepsilon/2$, and $\varepsilon/2$. By Lemma 6.30, the three conditions are satisfied, implying that it is NP-Hard to distinguish the case that $\|A\|_{2\to q} = 1$ and $\|A\|_{2\to q} \leqslant \gamma_{\mathbb{F},q} + \varepsilon$. □

6.4.1 Proof of Theorem 6.28

We will closely follow the arguments used in [BRS15, BGG+23]. The starting point of our reduction will be the following result implicit in [BRS15], which informally says that it is NP-hard to find a sparse vector in a subspace, according to a certain block-wise notion of sparsity.

Theorem 6.31 ([BRS15])**.** *For all $\varepsilon, \delta, \alpha > 0$ and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, there is a $k = \mathrm{poly}(1/\varepsilon, 1/\delta, 1/\alpha)$ such that given a subspace $W \subseteq \mathbb{F}^{n\times k}$ in the form of a projection matrix $P \in \mathbb{F}^{(n\times k)\times(n\times k)}$, it is* NP*-Hard to distinguish between the following:*

- *There is a vector $x \in W$ such that for all $i \in [n]$, the vector $x_i \in \mathbb{F}^k$ is in $\{e_1, \ldots, e_k\}$.*
- *For all vectors $x \in W$ with $\|x\|_{\ell_2}^2 = n$, the set*

$$S = \{i \in [n] : \|x_i\|_{\ell_2} \leqslant 1/\alpha, \|x_i\|_\infty \geqslant \delta\}$$

  has size at most εn.

Let $0 < \varepsilon', \alpha \leqslant 1$ be constant parameters depending on δ, γ, and ε that we will specify later. We prove hardness of the $[A]_{2\to f}$ problem by a reduction from the NP-Hard problem described in Theorem 6.31 with parameters ε', δ, and α. Theorem 6.31 implies that there is some $k = \mathrm{poly}(1/\varepsilon', 1/\delta, 1/\alpha)$ such that given a projection matrix $P \in \mathbb{F}^{(n\times k)\times(n\times k)}$ for a subspace $W \subseteq \mathbb{F}^{n\times k}$, it is NP-Hard to distinguish between the following:

- There is a vector $x \in W$ such that for all $i \in [n]$, the vector $x_i \in \mathbb{F}^k$ is in $\{e_1, \dots, e_k\}$.
- For all vectors $x \in W$ with $\|x\|_{\ell_2}^2 = n$, the set

  $$S = \{i \in [n] : \|x_i\|_{\ell_2} \leqslant 1/\alpha, \|x_i\|_\infty \geqslant \delta\}$$

  has size at most εn.

Our reduction will map the projection matrix $P \in \mathbb{F}^{(n\times k)\times(n\times k)}$ to the matrix $A = (I_n \otimes E_k) \cdot P$. Note that $A \in \mathbb{F}^{(n\times d_k)\times(n\times k)}$. To analyze the reduction, we must prove completeness and soundness.

Completeness

If there is a vector $x \in W$ such that for all $i \in [n]$, $x_i \in \{e_1, \dots, e_k\}$, we need to show $[A]_{2\to f} = 1$.

Indeed, we can consider the vector $z := Ax = (E_k \otimes I_n) \cdot Px = (E_k \otimes I_n)x$. We have for all $i \in [n]$, $z_i = E_k x_i$. Since x_i is a standard basis vector, z_i must be equal to some column of E_k. So by Assumption 2, all entries of z have magnitude $1/\sqrt{k}$, implying $[z]_f = f^{-1}(f(\frac{1}{\sqrt{k}})) = \frac{1}{\sqrt{k}}$. Therefore $[A]_{2\to f} \geqslant \frac{[z]_f}{\|x\|_2} = 1$.

On the other hand, $[A]_{2\to f} \leqslant \|A\|_{2\to 2} \leqslant \|P\|_{2\to 2} \cdot \|E_k\|_{2\to 2} \leqslant 1$ by Claim 6.13.

Soundness Assuming for all $x \in W$ with $\|x\|_{\ell_2}^2 \leqslant n$, the set

$$S = \{i \in [n] : \|x_i\|_{\ell_2} \leqslant 1/\alpha, \|x_i\|_\infty \geqslant \delta\}$$

has size at most $\varepsilon' n$, we need to show $[A]_{2\to f} \leqslant \gamma + \varepsilon$.

Let $y \in \mathbb{F}^{n\times k}$ be an arbitrary vector with $\|y\|_2 = 1$, and set $x = \frac{1}{k} \cdot Py$ and $z = Ay = k \cdot (E_k \otimes I_n)x$. Note that because P is a projection matrix, $\|x\|_{\ell_2}^2 = nk \cdot \|x\|_2^2 \leqslant n\|y\|_2^2 = n$, and $\|z\|_2 \leqslant \|y\|_2 = 1$. By virtue of the normalization on x, we have $\|x_i\|_{\ell_2} = \|z_i\|_2$ for each block $i \in [n]$.

We must show $[z]_f \leqslant \gamma + \varepsilon$. We will upper bound the contribution of different indices $i \in [n]$ to $[z]_f$ separately. To do this, define the following partition of $[n]$:

$$\begin{aligned} V_0 &:= S = \{i \in [n] : \|x_i\|_{\ell_2} \leqslant 1/\alpha, \|x_i\|_\infty \geqslant \delta\}, \\ V_1 &:= \{i \in [n] : \|x_i\|_{\ell_2} \leqslant \alpha, \|x_i\|_\infty \leqslant \delta\alpha\}, \\ V_2 &:= \{i \in [n] : \|x_i\|_{\ell_2} \geqslant \alpha, \|x_i\|_\infty \leqslant \delta\alpha\}, \\ V_3 &:= \{i \in [n] : \|x_i\|_{\ell_2} > 1/\alpha\}. \end{aligned}$$

For all $u \in \{0,1,2,3\}$, define $z^{(u)} \in \mathbb{F}^{V_u \times d_k}$ as the collection of $z_i \in \mathbb{F}^{d_k}$ for all $i \in V_u$. Note that

$$[z]_f^f = \sum_{u\in\{0,1,2,3\}} \frac{|V_u|}{|V|} \cdot [z^{(u)}]_f^f. \tag{6.10}$$

We will prove bounds on $[z^{(u)}]_f^f$ for $u \in \{0,1,2,3\}$.

For $u = 0$, Claim 6.13 applied to z_i implies

$$[z^{(0)}]_f^f = \mathop{\mathbb{E}}_{i\sim V_0} [z_i]_f^f \leqslant \mathop{\mathbb{E}}_{i\sim V_0} f(\|z_i\|_2) = \mathop{\mathbb{E}}_{i\sim V_0} f(\|x_i\|_{\ell_2}) \leqslant f(1/\alpha). \tag{6.11}$$

For $u = 1$, a similar application of Claim 6.13 implies

$$[z^{(1)}]_f^f = \mathop{\mathbb{E}}_{i\sim V_1}[z_i]_f^f \leqslant \mathop{\mathbb{E}}_{i\sim V_1} f(\|z_i\|_2) = \mathop{\mathbb{E}}_{i\sim V_1} f(\|x_i\|_{\ell_2}) \leqslant f(\alpha). \tag{6.12}$$

For $u = 2$, we have

$$[z^{(2)}]_f^f = \mathop{\mathbb{E}}_{i\sim V_2}[z_i]_f^f \underset{\text{Assumption 3}}{\leqslant} \mathop{\mathbb{E}}_{i\sim V_2}[f(\gamma \cdot \|z_i\|_2)] \underset{\textit{Claim 6.13}}{\leqslant} f(\gamma \cdot \|z^{(2)}\|_2) \underset{\|z\|_2\leqslant 1}{\leqslant} f\left(\gamma \cdot \sqrt{\frac{|V|}{|V_2|}}\right). \tag{6.13}$$

Finally, for $u = 3$,

$$\begin{aligned}
[z^{(3)}]_f^f &= \mathop{\mathbb{E}}_{i\sim V_3}[z_i]_f^f \\
&\underset{\textit{Claim 6.13}}{\leqslant} \mathop{\mathbb{E}}_{i\sim V_3} f(\|z_i\|_2) \\
&= \mathop{\mathbb{E}}_{i\sim V_3}\left[\|z_i\|_2^2 \cdot \frac{f(\|z_i\|_2)}{\|z_i\|_2^2}\right] \\
&\leqslant \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2} \cdot \mathop{\mathbb{E}}_{i\sim V_3} \|z_i\|_2^2 \\
&= \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2} \cdot \|z^{(3)}\|_2^2 \\
&\underset{\|z\|_2^2=1}{\leqslant} \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2} \cdot \frac{|V|}{|V_3|}.
\end{aligned} \tag{6.14}$$

Now we are equipped to bound Eq. (6.10).

$$\begin{aligned}
[z]_f^f &\leqslant \sum_{u\in\{0,1,2\}} \frac{|V_u|}{|V \setminus V_3|}[z^{(u)}]_f^f + \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2} && \text{(Eqs. (6.10) and (6.14))} \\
&\leqslant f\left(\sqrt{\sum_{u\in\{0,1,2\}} \frac{|V_u|}{|V \setminus V_3|}[z^{(u)}]_f^2}\right) + \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2} && \text{(Jensen's inequality for } x \to f(\sqrt{x})\text{)} \\
&\leqslant f\left(\sqrt{\sum_{u\in\{0,1,2\}} \frac{|V_u|}{(1-\alpha^2)|V|}[z^{(u)}]_f^2}\right) + \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2} && (|V_3| \leqslant \alpha^2 \cdot |V_0| \text{ by def of } V_3) \\
&\leqslant f\left(\sqrt{\frac{\varepsilon'/\alpha^2 + \alpha^2 + \gamma^2}{(1-\alpha^2)}}\right) + \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2} && \text{(Eqs. (6.11) to (6.13) and } V_0 \leqslant \varepsilon'|V|) \\
&\leqslant f\left(\frac{\gamma + \sqrt{\varepsilon'}/\alpha + \alpha}{\sqrt{1-\alpha^2}}\right) + \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2} && (\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b},\ f \text{ is monotone}) \\
&= f\left(\frac{\gamma + 2\alpha}{\sqrt{1-\alpha^2}}\right) + \sup_{w\geqslant 1/\alpha} \frac{f(w)}{w^2}. && (\text{Setting } \varepsilon' = \alpha^4)
\end{aligned}$$

Using the assumption that f is continuous and $\lim_{x\to\infty} f(x)/x^2 = 0$, we get that the limit of the right hand side as $\alpha \to 0$ is exactly $f(\gamma)$. Therefore, there exists some $\alpha > 0$ independent of n such that $[z]_f \leqslant \gamma + \varepsilon$. We choose α in our invocation of Theorem 6.31 accordingly, completing the proof that $[A]_{2\to f} \leqslant \gamma + \varepsilon$.

6.4.2 Proof of Theorem 6.2

In this subsection we prove Theorem 6.2. We use the following lemma which proves that the approximability of the permanent of highly rank-deficient $n \times n$ PSD matrices is essentially the same as the approximability of of the $2 \to 0$ norm.

Lemma 6.32. *Let $V \in \mathbb{C}^{n\times d}$. Then,*

$$c^{\mathbb{C}}_{n,d} \cdot \binom{n+d-1}{d}^{-1} \cdot \|V\|^{2n}_{2\to 0} \leqslant \mathrm{per}(VV^{\mathcal{D}}) \leqslant c^{\mathbb{C}}_{n,d} \cdot \|V\|^{2n}_{2\to 0},$$

where $c^{\mathbb{C}}_{n,d}$ is defined in Lemma 6.23.

Proof. Let $v_1, \ldots, v_n$ be the rows of V. For the upper bound, we can write

$$\begin{aligned}\mathrm{per}(VV^{\mathcal{D}}) &= \mathop{\mathbb{E}}_{x\sim\mathbb{C}N(0,I)} \prod_{i\in[n]} |\langle x, v_i\rangle|^2 \\ &= c_{n,d} \cdot \mathop{\mathbb{E}}_{\|x\|_2=1} \prod_{i\in[n]} |\langle x, v_i\rangle|^2 && \text{(Lemma 6.23)} \\ &\leqslant c_{n,d} \cdot \max_{\|x\|_2=1} \prod_{i\in[n]} |\langle x, v_i\rangle|^2 \\ &= c_{n,d} \cdot \|V\|^{2n}_{2\to 0}.\end{aligned}$$

Next we prove the lower bound. Let $z \in \mathbb{C}^d$ be a vector with $\|z\|_2 = 1$ vector maximizing $\|Vz\|_0 = \prod_{i\in[n]} |\langle z, v_i\rangle|^{1/n}$. We have $zz^{\mathcal{D}} \preceq \|z\|^2_{\ell_2} \cdot I = d \cdot I$, so $Vzz^{\mathcal{D}}V^{\mathcal{D}} \preceq d \cdot VV^{\mathcal{D}}$. By Lemma 6.22, this implies $\mathrm{per}(Vzz^{\mathcal{D}}V^{\mathcal{D}}) \leqslant d^n \cdot \mathrm{per}(VV^{\mathcal{D}})$. Since $Vzz^{\mathcal{D}}V^{\mathcal{D}}$ is rank 1, we can compute its permanent as $\mathrm{per}(Vzz^{\mathcal{D}}V^{\mathcal{D}}) = n! \cdot \prod_{i\in[n]} |\langle z, v_i\rangle|^2$.

$$\begin{aligned}\|Vz\|^{2n}_0 &= \max_{\|x\|_2=1} \prod_{i\in[n]} |\langle x, v_i\rangle|^2 \\ &= \frac{1}{n!} \cdot \mathrm{per}(Vzz^{\mathcal{D}}V^{\mathcal{D}}) \\ &\leqslant \frac{d^n}{n!} \cdot \mathrm{per}(VV^{\mathcal{D}}) \\ &= \binom{n+d-1}{d} \cdot c^{-1}_{n,d} \cdot \mathrm{per}(VV^{\mathcal{D}}).\end{aligned}$$

□

Proof of Theorem 6.2. We start from Theorem 6.5 for the case $\mathbb{F} = \mathbb{C}$ and $q = 0$, to get that it is NP-hard to approximate $\|A\|_{2\to 0}$ within a factor of $e^{-\gamma/2+\varepsilon/4}$ (recall that $\gamma_{\mathbb{C},0} = e^{-\gamma/2}$ by Fact 6.18).

We reduce the problem of approximating $\|A\|_{2\to 0}$ for $A \in \mathbb{C}^{n\times d}$ to approximating the permanent of the positive semidefinite matrix $B = A^{(k)}(A^{(k)})^{\mathcal{D}}$, where $A^{(k)} \in \mathbb{C}^{nk\times d}$ is as in Fact 6.15. By Lemma 6.32, $\mathrm{per}(B)$ is proportional to $\|A^{(k)}\|_{2\to 0}^{2nk}$ up to a multiplicative error of

$$\binom{nk+d-1}{d} \leqslant e^{\varepsilon kn/2},$$

which holds for $k = O(\frac{d}{n\varepsilon^2})$ and $\varepsilon > 0$ small enough. Note that the reduction is efficient because k is polynomial in the size of A.

By Fact 6.15, we have $\|A^{(k)}\|_{2\to 0}^{2nk} = \|A\|_{2\to 0}^{2nk}$ which is hard to approximate within a factor of $e^{-kn(\gamma-\varepsilon/2)}$. Therefore, it is NP-hard to approximate $\mathrm{per}(B)$ within a factor of $e^{-kn(\gamma-\varepsilon)}$, where $B \in \mathbb{C}^{kn\times kn}$. □

6.5 Proof of Lemma 6.30

In this subsection we prove Lemma 6.30. We use the following corollary which we will prove in Section 6.6.

Corollary 6.33. *Let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be an absolutely continuous 2-concave increasing function with $0 \in \overline{\mathrm{Range}(f)}$. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and let $z \in \mathbb{F}^n$ be a vector with $\|z\|_{\ell_2} = 1$ and $\|z\|_\infty \leqslant \delta$. For each i, let σ_i be an independently and uniformly sampled member of $\{-1, +1\}$ if $\mathbb{F} = \mathbb{R}$ and be an independently and uniformly sampled member of $\{-1, +1, -i, +i\}$ otherwise. Then there exists a universal $C > 0$ such that $Z = \sum_{i\in[n]} \sigma_i z_i$ satisfies*

$$[Z]_f^f \leqslant [g]_f^f + C \cdot \left(-\int_0^{C\delta} \min(0, f(u)) + \delta \cdot \left(f(2\sqrt{\log(1/\delta)}) + f'(1) \right) \right),$$

where $g \sim \mathbb{F}\mathcal{N}(0, 1)$.

Now we are ready to prove the main result of this subsection.

Proof of Lemma 6.30. Let $z = x/\|x\|_2$ and $y = Ez = Ex/\|x\|_2$. Let m be the number of rows of E. Note that for $i \in [m]$, we have

$$y_i = \sum_{j\in[n]} E_{i,j} \cdot z_j.$$

Let $Z_j = E_{i,j} \cdot z_j$ be the random variable where i chosen uniformly at random from $[m]$. We invoke Corollary 6.33 on $z/\sqrt{k}$. We verify its conditions: First, $\|\frac{z}{\sqrt{k}}\|_{\ell_2} = \|z\|_2 = 1$. Second, by definition of $E_k^{(\mathbb{F})}$ we can write $Z_j = \sigma_j \cdot \frac{z_j}{\sqrt{k}}$, where the random variables $\sigma_j \in \{+1, -1\}$ when $\mathbb{F} = \mathbb{R}$ and $\sigma_j \in \{+1, -1, +i, -i\}$ chosen independently and uniformly at random Lastly,

$$\frac{|z_j|}{\sqrt{k}} = \frac{|x_j|}{\sqrt{k}\|x\|_2} = \frac{|x_j|}{\|x\|_{\ell_2}} \leqslant \frac{\|x\|_\infty}{\|x\|_{\ell_2}} \leqslant \delta.$$

Now by invoking Section 6.6 on the random variables $Z_1, \ldots, Z_k$, we get

$$\left[\frac{Ex}{\|x\|_2}\right]_f^f = [y]_f^f = \mathop{\mathbb{E}}_{i\sim[m]} f(y_i) = \mathbb{E}\left[f\left(\sum_{j\in[n]} Z_j\right)\right] \leqslant [g]_f^f + C \cdot \left(-\int_0^{C\delta} \min(0, f(u)) + \delta \cdot \left(f(2\sqrt{\log(1/\delta)}) + f'(1)\right)\right),$$

as desired.

Now assume $f = f_p$ for some $-1 < p < 2$. By Lemma 6.14, $\|\cdot\|_p$ is homogeneous, and we get

$$\begin{aligned}[Ex]_f &= \|x\|_2 \cdot \left[\frac{Ex}{\|x\|_2}\right]_f \\ &\leqslant \|x\|_2 \cdot f^{-1}\left(\gamma_{\mathbb{F},p}^p + C \cdot \left(-\int_0^{C\delta} \min(0, f(u)) + \delta \cdot \left(f(2\sqrt{\log(1/\delta)}) + f'(1)\right)\right)\right)\end{aligned}$$

The second term is 0 if $p > 0$, otherwise it is equal to $\frac{\delta^{p+1}}{p+1}$. The third term is 2δ if $p < 0$, otherwise it is bounded by $4\delta(\log(1/\delta) + 1)$ for δ small enough. Therefore, the limit of the right hand side as $\delta \to 0$ is $\|x\|_2 \cdot \gamma_{F,p}$. □

6.6 Proof of Corollary 6.33

Corollary 6.33 follows directly as a special case of the following Lemma, for $k = 1$ if $\mathbb{F} = \mathbb{R}$ and for $k = 2$ if $\mathbb{F} = \mathbb{C}$.

lemmamain Let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be an absolutely continuous 2-concave increasing function with $0 \in \overline{\text{Range}(f)}$. Let $k \geqslant 1$ and let $X_1, \ldots, X_n$ are bounded independent random variables in $\mathbb{R}^k$ with $\mathbb{E}[X_i] = 0$, $\text{Cov}(\sum_i X_i) = I_k/k$, and $\|X_i\|_{\ell_2} \leqslant \delta_i$ such that $\sum_i \delta_i^2 \leqslant 1$ and $\delta_i \leqslant \delta$ for some $0 < \delta < 1$, then there exists $\eta_k > 0$ such that

$$\left[\Big\|\sum_{i\in[n]} X_i\Big\|_{\ell_2}\right]_f^f \leqslant [\|g\|_{\ell_2}]_f^f - e\int_0^{C_k\delta/e} \min(0, f(u))du + \delta \cdot \left(C_k \cdot \max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right),$$

where $g \sim N(0, I_k/k)$, and C_k is the constant in Theorem 6.34.

Before proving Section 6.6, we state some probabilistic tools we will require in the proof.

Theorem 6.34 (Multivariate Berry-Esseen [Ben05]). *Let $k \geqslant 1$ and let $X_1, \ldots, X_n$ be independent random variables in $\mathbb{R}^k$ satisfying $\mathbb{E}[X_i] = 0$ for $1 \leqslant i \leqslant n$. Define $X = X_1 + \cdots + X_n$ and suppose $\mathbb{E}[XX^T] = I_k$. Further let $g \sim \mathcal{N}(0, I_k)$. Then there exists $C_k > 0$ such that for all convex sets $U \subseteq \mathbb{R}^k$ it holds that*

$$|\,\mathbb{P}[g \in U] - \mathbb{P}[X \in U]| \leqslant C_k \cdot \mathop{\mathbb{E}}_{i\in[n]}[\|X_i\|_{\ell_2}^3].$$

In particular, for all $u \geqslant 0$ it holds that

$$|\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - \mathbb{P}[\|X\|_{\ell_2} \leqslant u]| \leqslant C_k \cdot \mathop{\mathbb{E}}_{i\in[n]}[\|X_i\|_{\ell_2}^3].$$

Theorem 6.35 (Multivariate Hoeffding [Pin, Theorem 3])**.** *Let $k \geqslant 1$ and let $X_1, \ldots, X_n$ be independent random variables in $\mathbb{R}^k$ satisfying $\mathbb{E}[X_i] = 0$ for $1 \leqslant i \leqslant n$. Further suppose $\|X_i\|_{\ell_2} \leqslant \delta_i$ almost surely for $\delta_1, \ldots, \delta_n > 0$. Define $X = X_1 + \cdots + X_n$.*

$$|\mathbb{P}[\|X\|_{\ell_2} \geqslant u]| \leqslant 2 \cdot e^{-u^2/2\sum_{i\in[n]}\delta_i^2}.$$

Fact 6.36. *Let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be an absolutely continuous function. If $\int_0^b f(u)$ is bounded then $\lim_{u\to 0} u f(u) = 0$.*

Claim 6.37. Let X be a random variable over $\mathbb{R}_{>0}$, and let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be an absolutely continuous function with $f(a) = 0$ for some $a \in \mathbb{R}_{>0}$. Then, if $\mathbb{E}[f(X)]$ exists,

$$\mathbb{E}[f(X)] = -\int_0^a f'(u)\,\mathbb{P}[X \leqslant u]du + \int_a^\infty f'(u)\,\mathbb{P}[X \geqslant u]du + \lim_{u\to 0} f(u)\,\mathbb{P}[X \leqslant u] - \lim_{u\to\infty} f(u)\,\mathbb{P}[X \geqslant u].$$

Proof. Let $\mu : \mathbb{R}_{>0} \to \mathbb{R}$ denote the PDF of the random variable X.

$$\begin{aligned}
\mathbb{E}[f(X)] &= \int_0^\infty f(u)d\mu(u) \\
&= \int_0^a f(u)d\mu(u) + \int_a^\infty f(u)d\mu(u) \\
&= f(u)\,\mathbb{P}[X \leqslant u]\Big|_0^a - \int_{I_1} f'(u)\,\mathbb{P}[X \leqslant u] - f(u)\,\mathbb{P}[X \geqslant u]\Big|_a^\infty + \int_{I_2} f'(u)\,\mathbb{P}[X \geqslant u] \\
&= \lim_{u\to 0} f(u)\,\mathbb{P}[X \leqslant u] - \lim_{u\to\infty} f(u)\,\mathbb{P}[X \geqslant u] - \int_{I_1} f'(u)\,\mathbb{P}[X \leqslant u] + \int_{I_2} f'(u)\,\mathbb{P}[X \geqslant u] \\
&\qquad (f(a) = 0.)
\end{aligned}$$

□

Fact 6.38. *Let $k \geqslant 1$ and let $g \sim N(0, I_k/k)$. Then for all $0 \leqslant u \leqslant 1$, $\mathbb{P}[\|g\|_{\ell_2} \leqslant u] \leqslant eu$.*

Proof. $k \cdot \|g\|_{\ell_2}^2$ is distributed as a Chi-squared random variable with k degrees of freedom. In [DG03, Lemma 2.2], the authors show that

$$\mathbb{P}[\|g\|_{\ell_2} \leqslant u] \leqslant (u^2 e^{1-u^2})^{k/2} \leqslant e \cdot u,$$

as desired. □

With these facts in hand, we are ready to prove Section 6.6.

Proof of Section 6.6. Define the random variable $X = \|\sum_i X_i\|_{\ell_2}$. We split the domain of f into $I_1 = f^{-1}([-\infty, 0])$ and $I_2 = f^{-1}([0, \infty])$, where $I_1 = [0, a]$ and $I_2 = [a, \infty]$. Since $0 \in \overline{\mathrm{Range}(f)}$, we have $f(a) = 0$. We use Claim 6.37 to write the left hand side as

$$\begin{aligned}
\mathbb{E}[f(X)] &= -\int_0^a f'(u)\,\mathbb{P}[X \leqslant u] + \int_a^\infty f'(u)\,\mathbb{P}[X \geqslant u] + \lim_{u\to 0} f(u)\,\mathbb{P}[X \leqslant u] - \lim_{u\to\infty} f(u)\,\mathbb{P}[X \geqslant u] \\
&\leqslant -\int_0^a f'(u)\,\mathbb{P}[X \leqslant u] + \int_a^\infty f'(u)\,\mathbb{P}[X \geqslant u]. \qquad (f(0) \leqslant f(a) = 0,\, f(\infty) \geqslant f(a) = 0.)
\end{aligned}$$

We bound each of the above integrals separately.

Claim 6.39. We have

$$\int_0^a f'(u)\,\mathbb{P}[X \leqslant u] \geqslant \int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] + e\int_0^{C_k\delta/e} \min(0, f(u)).$$

Claim 6.40. We have

$$\int_a^\infty f'(u)\,\mathbb{P}[X \geqslant u] \leqslant \int_a^\infty f'(u)\,\mathbb{P}[|g| \geqslant u] + \delta\cdot\left(C_k\cdot\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right).$$

We will complete the proof of Section 6.6 using these two claims, and prove the claims later.

$$\begin{aligned}
\mathbb{E}[f(X)] &\leqslant -\int_0^a f'(u)\,\mathbb{P}[X \leqslant u] + \int_a^\infty f'(u)\,\mathbb{P}[X \geqslant u] \\
&\leqslant -\int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - e\int_0^{C_k\delta/e} \min(0, f(u)) \\
&\quad + \int_a^\infty f'(u)\,\mathbb{P}[|g| \geqslant u] + \delta\cdot\left(C_k\cdot f(2\sqrt{\log(1/\delta)}) + 2f'(1)\right) \\
&= \mathbb{E}[f(\|g\|_{\ell_2})] - \int_0^{C_k\delta} \min(0, f(u)) + \delta\cdot\left(C_k\cdot\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right).
\end{aligned}$$

Here, the last equality is by applying Claim 6.37 on the random variable $\|g\|_{\ell_2}$:

$$\begin{aligned}
\mathbb{E}[\|g\|_{\ell_2}] = &-\int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] + \int_a^\infty f'(u)\,\mathbb{P}[|g| \geqslant u] \\
&+ \lim_{u\to 0} f(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - \lim_{u\to\infty} f(u)\,\mathbb{P}[\|g\|_{\ell_2} \geqslant u]
\end{aligned}$$

Observe that by Fact 6.36, the third term above is zero, and by Claim 6.13,

$$\lim_{u\to\infty} f(u)\,\mathbb{P}[\|g\|_{\ell_2} \geqslant u] \leqslant \lim_{u\to\infty}(f'(1)\cdot u^2 + f(1))\cdot\mathbb{P}[\|g\|_{\ell_2} \geqslant u] = 0.$$

This completes the justification of the last inequality. It remains to prove Claim 6.39 and Claim 6.40. We begin by writing an inequality which will be used in both proofs. By Theorem 6.34, for any $u \geqslant 0$,

$$\begin{aligned}
|\,\mathbb{P}[X \leqslant u] - \mathbb{P}[\|g\|_{\ell_2} \leqslant u]| &\leqslant C_k\cdot \mathop{\mathbb{E}}_{i\in[n]}[\|X_i\|_{\ell_2}^3] \\
&\leqslant C_k\cdot\sum_{i\in[n]}[\|X_i\|_{\ell_2}^2]\cdot\max_{i\in[n]}\|X_i\|_{\ell_2} \\
&\leqslant C_k\cdot\delta. \qquad (6.15)
\end{aligned}$$

Proof of Claim 6.39. For a parameter b with $0 \leqslant b \leqslant a$, we write

$$\int_{I_1} f'(u)\,\mathbb{P}[X \leqslant u] \geqslant \int_b^a f'(u)\,\mathbb{P}[X \leqslant u] \qquad (f'(u) \geqslant 0)$$

$$\begin{aligned}
&\geqslant \int_b^a f'(u)(\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - C_k\delta) && \text{(Eq. (6.15))}\\
&= \int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - \int_0^b f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - C_k\delta(f(a) - f(b))\\
&\geqslant \int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - e\int_0^b f'(u)u + C_k\delta f(b)\\
& && \text{(using } f(a) = 0\text{, and Fact 6.38)}\\
&= \int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - euf(u)\Big|_0^b + e\int_0^b f(u) + C_k\delta f(b)\\
&= \int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] - ebf(b) + e\int_0^b f(u) + C_k\delta f(b) && \text{(Fact 6.36)}\\
&= \int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] + e\int_0^{\min(a, C_k\delta/e)} f(u) && \text{(Setting } b = \min(a, C_k\delta/e))\\
&= \int_0^a f'(u)\,\mathbb{P}[\|g\|_{\ell_2} \leqslant u] + e\int_0^{C_k\delta/e} \min(f(u), 0).
\end{aligned}$$

The penultimate equality is because if $\min(a, C_k\delta/e) = a$, then $f(b) = 0$, and if $\min(a, C_k\delta/e) = C_k\delta/e$, then $bf(b) = C_k\delta f(b)/e$.

□

Proof of Claim 6.40. Applying Theorem 6.35 on X,

$$\mathbb{P}[X \geqslant u] \leqslant 2 \cdot e^{-u^2/2\sum\delta_i^2} \leqslant 2 \cdot e^{-u^2/2}. \tag{6.16}$$

For $\max(a, 1) < c < \infty$ we write

$$\begin{aligned}
\int_{I_2} f'(u)\,\mathbb{P}[X \geqslant u] &= \int_a^c f'(u)\,\mathbb{P}[X \geqslant u] + \int_c^\infty f'(u)\,\mathbb{P}[X \geqslant u]\\
&\leqslant \int_a^c f'(u)(\mathbb{P}[|g| \geqslant u] + C_k\delta) + \int_c^\infty 2f'(u)e^{-u^2/2} && \text{(Eq. (6.15), Eq. (6.16))}\\
&\leqslant \int_a^c f'(u)\,\mathbb{P}[|g| \geqslant u] + C_k\delta\cdot(f(c) - f(a)) + 2f'(1)\cdot\int_c^\infty ue^{-u^2/2}\\
& && \text{(using } f \text{ is 2-concave, Claim 6.13, and } c \geqslant 1)\\
&\leqslant \int_a^\infty f'(u)\,\mathbb{P}[|g| \geqslant u] + C_k\delta\cdot f(c) + 2f'(1)\cdot\int_c^\infty ue^{-u^2/2}\\
& && (f \text{ is increasing, } f(a) = 0)\\
&= \int_a^\infty f'(u)\,\mathbb{P}[|g| \geqslant u] + C_k\delta\cdot f(c) + 2f'(1)\cdot e^{-c^2/2} && (6.17)
\end{aligned}$$

Setting $c = \max(a, 2\sqrt{\log(1/\delta)})$, we get

$$\int_{I_2} f'(u)\,\mathbb{P}[X \geqslant u] \leqslant \int_a^\infty f'(u)\,\mathbb{P}[|g| \geqslant u] + \delta\cdot\Big(C_k\cdot\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\Big). \tag{6.18}$$

The inequality above is by bounding $e^{-c^2/2} \leqslant \delta$, and if $\max(a, 2\sqrt{\log(1/\delta)}) = a$, then $f(c) = 0$. □

□

6.7 Algorithm for approximating $2 \to q$ norm

Let $q < 2$. Given a matrix $A \in \mathbb{F}^{n\times d}$ with rows $\{a_i\}_{i\in[n]}$, we write an expression for $\|A\|_{2\to q}$:

$$\|A\|_{2\to q} = f_q^{-1}\left(\max_{x\in\mathbb{F}^d, \|x\|_2=1}\left[\sum_{i\in[n]} f_q\left(|\langle a_i, x\rangle|\right)\right]\right) = f_q^{-1}\left(\max_{x\in\mathbb{F}^d, \|x\|_2=1}\left[\sum_{i\in[n]} f_q\left(\sqrt{a_i^{\mathcal{D}} x x^{\mathcal{D}} a_i}\right)\right]\right).$$

Notice that $xx^{\mathcal{D}}$ is a rank-1 PSD matrix with $\mathrm{tr}(xx^{\mathcal{D}}) = x^{\mathcal{D}}x = \|x\|_{\ell_2} = d \cdot \|x\|_2 = d$. We can relax the rank 1 constraint to obtain a relaxation of $\|A\|_{2\to q}$ which we denote by $\mathrm{SDP}_{2\to q}(A)$:

$$\mathrm{SDP}_{2\to q}(A) := f_q^{-1}\left(\max_{X\succeq 0, \mathrm{tr}(X)=d}\left[\sum_{i\in[n]} f_q\left(\sqrt{a_i^{\mathcal{D}} X a_i}\right)\right]\right),$$

where the maximum is taken over all matrices in $\mathbb{F}^{d\times d}$. Note that the objective function being maximized is concave for all $q \leqslant 2$ because the function $X \to a_i^{\mathcal{D}} X a_i$ is linear and f_q is 2-concave.

Lemma 6.41. *For any matrix $A \in \mathbb{F}^{n\times d}$ and any $-1 < q \leqslant 2$,*

$$\|A\|_{2\to q} \leqslant \mathrm{SDP}_{2\to q}(A) \leqslant \gamma_{\mathbb{F},q}^{-1} \cdot \|A\|_{2\to q}.$$

Proof. The first inequality is because $\mathrm{SDP}_{2\to q}(A)$ is a relaxation of $\|A\|_{2\to q}$. For the second inequality, we describe a rounding procedure for the relaxation.

Let $X^* \in \mathbb{F}^{d\times d}$ be the optimal solution to $\mathrm{SDP}_{2\to q}(A)$. We sample a vector $x \sim \mathbb{F}\mathcal{N}(0, X^*)$. Clearly,

$$\|A\|_{2\to q}^2 = \max_{x\in\mathbb{F}^n, x\neq 0} \frac{\|Ax\|_q^2}{\|x\|_2^2} \geqslant \frac{\mathbb{E}\,\|Ax\|_q^2}{\mathbb{E}\,\|x\|_2^2}. \tag{6.19}$$

First we calculate the denominator of the right hand side: $\mathbb{E}\,\|x\|_2^2 = \frac{1}{d} \cdot \mathrm{tr}(X) = 1$. For the numerator, we bound

$$\begin{aligned}
\mathbb{E}\,\|Ax\|_q^2 &= \mathbb{E}\left(f_q^{-1}\left(\sum_{i\in[n]} f_q(|\langle a_i, x\rangle|)\right)^2\right) \\
&\geqslant f_q^{-1}\left(\sum_{i\in[n]} \mathbb{E}\, f_q(|\langle a_i, x\rangle|)\right)^2 && \text{(Jensen's inequality, and } x \to f_q^{-1}(x)^2 \text{ is convex)} \\
&= f_q^{-1}\left(\sum_{i\in[n]} \mathop{\mathbb{E}}_{g\sim\mathbb{F}\mathcal{N}(0,1)} f_q\left(\sqrt{a_i^{\mathcal{D}} X^* a_i} \cdot |g|\right)\right)^2 && \text{(}\langle a_i, x\rangle \text{ is a Gaussian with variance } a_i^{\mathcal{D}} X^* a_i\text{)}
\end{aligned}$$

$$= f_q^{-1}\left(\mathop{\mathbb{E}}_{g\sim\mathbb{F}\mathcal{N}(0,1)} f_q(|g|)\right)^2 \cdot f_q^{-1}\left(\sum_{i\in[n]} f_q\left(\sqrt{a_i^{\mathcal{D}} X^* a_i}\right)\right)^2 \qquad \text{(Homogeneity)}$$

$$= \gamma_{\mathbb{F},q}^2 \cdot \mathrm{SDP}_{2\to q}(A)^2.$$

Together with Eq. (6.19), this implies $\|A\|_{2\to q} \geqslant \gamma_{\mathbb{F},q} \cdot \mathrm{SDP}_{2\to q}(A)$, completing the proof.

□

References

[ABBS10] Jacob Abernethy, Peter Bartlett, Niv Buchbinder, and Isabelle Stanton. A regularization approach to metrical task systems. In *Algorithmic Learning Theory*, ALT 2010. Springer, 2010.

[ADJ97] Jan Ambjørn, Bergfinnur Durhuus, and Thordur Jonsson. *Quantum geometry*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge, 1997. A statistical field theory approach.

[AGOS17] Nima Anari, Leonid Gurvits, Shayan Oveis Gharan, and Amin Saberi. Simply exponential approximation of the permanent of positive semidefinite matrices. In Chris Umans, editor, *FOCS*, pages 914–925. IEEE Computer Society, 2017.

[AL07] David Aldous and Russell Lyons. Processes on unimodular random networks. *Electron. J. Probab.*, 12:no. 54, 1454–1508, 2007.

[Ang03] O. Angel. Growth and percolation on the uniform infinite planar triangulation. *Geom. Funct. Anal.*, 13(5):935–974, 2003.

[AOV21] N. Anari, S. Oveis Gharan, and C. Vinzant. Log-concave polynomials i: Entropy and a deterministic approximation algorithm for counting bases of matroids. *Duke Mathematical Journal*, 2021. to appear.

[AS48] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1948.

[Bar96] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 184–193, 1996.

[Bar98] Martin T. Barlow. Diffusions on fractals. In *Lectures on probability theory and statistics (Saint-Flour, 1995)*, volume 1690 of *Lecture Notes in Math.*, pages 1–121. Springer, Berlin, 1998.

[Bar99] Alexander I. Barvinok. Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor. *Random Struct. Algorithms*, 14(1):29–61, 1999.

[Bar13] Alexander Barvinok. A bound for the number of vertices of a polytope with applications. *Combinatorica*, 33(1):1–10, 2013.

[Bar16] Alexander Barvinok. *Combinatorics and complexity of partition functions*, volume 30. Springer, 2016.

[Bar20] Alexander Barvinok. A remark on approximating permanents of positive definite matrices. arxiv, 2020.

[BBBT97] Yair Bartal, Avrim Blum, Carl Burch, and Andrew Tomkins. A polylog(n)-competitive algorithm for metrical task systems. In *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97, pages 711–719, New York, NY, USA, 1997. ACM.

[BBH+12] Boaz Barak, Fernando G.S.L. Brandao, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *STOC*, pages 307–326. ACM, 2012.

[BBM06] Yair Bartal, Béla Bollobás, and Manor Mendel. Ramsey-type theorems for metric spaces with applications to online problems. *J. Comput. System Sci.*, 72(5):890–921, 2006.

[BC12] Itai Benjamini and Nicolas Curien. Ergodic theory on stationary random graphs. *Electron. J. Probab.*, 17:no. 93, 20 pp., 2012.

[BC13] Itai Benjamini and Nicolas Curien. Simple random walk on the uniform infinite planar quadrangulation: subdiffusivity via pioneer points. *Geom. Funct. Anal.*, 23(2):501–531, 2013.

[BCL+18] Sébastien Bubeck, Michael B. Cohen, Yin Tat Lee, James R. Lee, and Aleksander Madry. k-server via multiscale entropic regularization. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 3–16, 2018.

[BCLL21] Sébastien Bubeck, Michael B. Cohen, James R. Lee, and Yin Tat Lee. Metrical task systems on trees via mirror descent and unfair gluing. *SIAM J. Comput.*, 50(3):909–923, 2021.

[BCN14] Niv Buchbinder, Shahar Chen, and Joseph (Seffi) Naor. Competitive analysis via regularization. In *Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pages 436–444, Philadelphia, PA, USA, 2014. Society for Industrial and Applied Mathematics.

[BCR23] Sébastien Bubeck, Christian Coester, and Yuval Rabani. The randomized -server conjecture is false! In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 581–594, 2023.

[Ben05] Vidmantas Bentkus. A lyapunov-type bound in rd. *Theory of Probability & Its Applications*, 49(2):311–323, 2005.

[Ben13] Itai Benjamini. *Coarse geometry and randomness*, volume 2100 of *Lecture Notes in Mathematics*. Springer, Cham, 2013. Lecture notes from the 41st Probability Summer School held in Saint-Flour, 2011, Chapter 5 is due to Nicolas Curien, Chapter 12 was written by Ariel Yadin, and Chapter 13 is joint work with Gady Kozma, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

[BGG+23] Vijay Bhattiprolu, Mrinal Kanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Inapproximability of matrix $p \to q$ norms. *SIAM Journal on Computing*, 52(1):132–155, 2023.

[BK02] Mario Bonk and Bruce Kleiner. Quasisymmetric parametrizations of two-dimensional metric spheres. *Invent. Math.*, 150(1):127–183, 2002.

[BKRS00] Avrim Blum, Howard Karloff, Yuval Rabani, and Michael Saks. A decomposition theorem for task systems and bounds for randomized server problems. *SIAM J. Comput.*, 30(5):1624–1661, 2000.

[BLMN05] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005.

[BLS92] Allan Borodin, Nathan Linial, and Michael E. Saks. An optimal on-line algorithm for metrical task system. *J. ACM*, 39(4):745–763, October 1992.

[BM86] Béla Bollobás and Brendan D McKay. The number of matchings in random regular graphs and bipartite graphs. *Journal of Combinatorial Theory, Series B*, 41(1):80–91, 1986.

[Bor19] Charles Bordenave. A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts. *Annales scientifiques de l'Ecole normale supérieure*, 2019.

[BP11] Itai Benjamini and Panos Papasoglu. Growth and isoperimetric profile of planar graphs. *Proc. Amer. Math. Soc.*, 139(11):4105–4111, 2011.

[BRS15] Jop Briët, Oded Regev, and Rishi Saket. Tight hardness of the non-commutative grothendieck problem. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1108–1122. IEEE, 2015.

[BS01] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13 pp., 2001.

[BSVV06] Ivona Bezáková, Daniel Stefankovic, Vijay V. Vazirani, and Eric Vigoda. Accelerating simulated annealing for the permanent and combinatorial counting problems. In *SODA*, pages 900–907. ACM Press, 2006.

[CKR01] Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 8–16, 2001.

[CL19] Christian Coester and James R. Lee. Pure entropic regularization for metrical task systems. In *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, pages 835–848, 2019.

[CS12] M. Chudnovsky and P Seymour. Perfect matchings in planar cubic graphs. *Combinatorica*, 32:403–424, 2012.

[DG03] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

[DG20] Jian Ding and Ewain Gwynne. The Fractal Dimension of Liouville Quantum Gravity: Universality, Monotonicity, and Bounds. *Comm. Math. Phys.*, 374(3):1877–1934, 2020.

[Ego81] G.P. Egorychev. The solution of van der waerden's problem for permanents. *Advances in Math.*, 42:299–305, 1981.

[EKK+11] Louis Esperet, František Kardoš, Andrew D. King, Daniel Král, and Serguei Norine. Exponentially many perfect matchings in cubic graphs. *Advances in Mathematics*, 227(4):1646–1664, 7 2011.

[EL20] F. Ebrahimnejad and J. R. Lee. On planar graphs of uniform polynomial growth. To appear, *Probab. Theory Related Fields.* Preprint at arXiv:math/2005.03139, 2020.

[Fal81] D. I. Falikman. Proof of the van der waerden's conjecture on the permanent of a doubly stochastic matrix. *Mat. Zametki*, 29(6):931–938, 1981. (in Russian).

[FM03] Amos Fiat and Manor Mendel. Better algorithms for unfair metrical task systems and applications. *SIAM Journal on Computing*, 32(6):1403–1422, 2003.

[Fri08] J. Friedman. A proof of alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910), 2008.

[FRT04] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. Syst. Sci.*, 69(3):485–497, 2004.

[GHS20] Ewain Gwynne, Nina Holden, and Xin Sun. A mating-of-trees approach for graph distances in random planar maps. *Probab. Theory Related Fields*, 177(3-4):1043–1102, 2020.

[GK10] David Gamarnik and Dmitriy Katz. A deterministic approximation algorithm for computing the permanent of a 0,1 matrix. *Journal of Computer and System Sciences*, 76(8):879–883, 2010.

[GN13] Ori Gurel-Gurevich and Asaf Nachmias. Recurrence of planar graph limits. *Ann. of Math. (2)*, 177(2):761–781, 2013.

[GRSW16] Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu. Bypassing ugc from some optimal geometric inapproximability results. *ACM Transactions on Algorithms*, 12(1):1–25, 2016.

[Gur06] Leonid Gurvits. Hyperbolic polynomials approach to van der waerden/schrijver-valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications. In Jon M. Kleinberg, editor, *STOC*, pages 417–426. ACM, 2006.

[HLP52] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge university press, 1952.

[JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM journal on computing*, 18(6):1149–1178, 1989.

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

[JV96] Mark Jerrum and Umesh V. Vazirani. A mildly exponential approximation algorithm for the permanent. *Algorithmica*, 16(4/5):392–401, 1996.

[KM08] Takashi Kumagai and Jun Misumi. Heat kernel estimates for strongly recurrent random walk on random media. *J. Theoret. Probab.*, 21(4):910–935, 2008.

[Kri05] M. A. Krikun. Uniform infinite planar triangulation and related time-reversed critical branching process. *J. Math. Sci.*, 131:5520–5537, 2005.

[Lee17] James R. Lee. Conformal growth rates and spectral geometry on distributional limits of graphs. Preprint at arXiv:math/1701.01598, 2017.

[LP86] L. Lovász and M.D. Plummer. Matching theory. 1986.

[LP13] J. R. Lee and Yuval Peres. Harmonic maps on amenable groups and a diffusive lower bound for random walks. *Ann. Probab.*, 41(5):3392–3419, 2013.

[LP16] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*. Cambridge University Press, New York, 2016. Preprint at `http://pages.iu.edu/~rdlyons/`.

[LSW00] N. Linial, A. Samorodnitsky, and A. Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, 2000.

[LT79] Richard J. Lipton and Robert Endre Tarjan. A separator theorem for planar graphs. *SIAM J. Appl. Math.*, 36(2):177–189, 1979.

[Mei23] A Meiburg. Inapproximability of positive semidefinite permanents and quantum state tomography. *Algorithmica*, 85:3828–3854, 2023.

[Mos20] Stefan M Moser. Expected logarithm and negative integer moments of a noncentral χ 2-distributed random variable. *Entropy*, 22(9):1048, 2020.

[MTTV97] Gary L. Miller, Shang-Hua Teng, William Thurston, and Stephen A. Vavasis. Separators for sphere-packings and nearest neighbor graphs. *J. ACM*, 44(1):1–29, 1997.

[MTTV98] Gary L. Miller, Shang-Hua Teng, William Thurston, and Stephen A. Vavasis. Geometric separators for finite-element meshes. *SIAM J. Sci. Comput.*, 19(2):364–386, 1998.

[Per12] Jerome K Percus. *Combinatorial methods*, volume 4. Springer Science & Business Media, 2012.

[Pin] Iosif Pinelis. An approach to inequalities for the distributions of infinite-dimensional martingales. *Progress in Probability*, 30:128–134.

[PRS94] Serge Plotkin, Satish Rao, and Warren D. Smith. Shallow excluded minors and improved graph decompositions. In *Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (Arlington, VA, 1994)*, pages 462–470, New York, 1994. ACM.

[RSZ16] Mark Rudelson, Alex Samorodnitsky, and Ofer Zeitouni. Hafnians, perfect matchings and gaussian matrices. *Ann. Probab.*, 44(4):2858 – 2888, 7 2016.

[RW94] Robert W. Robinson and Nicholas C. Wormald. Almost all regular graphs are hamiltonian. *Random Structures & Algorithms*, 5(2):363–374, 1994.

[Sch98] A. Schrijver. Counting 1-factors in regular bipartite graphs. *Journal of Combinatorial Theory*, B 72:122–135, 1998.

[Sei99] Steve Seiden. Unfair problems and randomized algorithms for metrical task systems. *Inf. Comput.*, 148(2):219–240, February 1999.

[ST07] Daniel A. Spielman and Shang-Hua Teng. Spectral partitioning works: Planar graphs and finite element meshes. *Linear Algebra and its Applications: Special Issue in honor of Miroslav Fiedler*, 421(2–3):284–305, March 2007.

[Ste05] Daureen Steinberg. Computation of matrix norms with applications to robust optimization. *Research thesis, Technion-Israel University of Technology*, 2, 2005.

[Tan84] R Michael Tanner. Explicit concentrators from generalized n-gons. *SIAM Journal on Algebraic Discrete Methods*, 5(3):287–293, 1984.

[Ten98] Shang-Hua Teng. Combinatorial aspects of geometric graphs. *Computational Geometry*, 9(4):277–287, 1998.

[YP21] Chenyang Yuan and Pablo A. Parrilo. Semidefinite relaxations of products of nonnegative forms on the sphere. CoRR abs/2102.13220, 2021.

[YP22] Chenyang Yuan and Pablo A. Parrilo. Maximizing products of linear forms, and the permanent of positive semidefinite matrices. *Math. Program.*, 193(1):499–510, 2022.