Sparsest Cut on quotients of the hypercube

[Preliminary draft]

Alexandra Kolla James R. Lee

Abstract

We present a simple construction and analysis of an $\Omega(\log \log N)$ integrality gap for the well-known Sparsest Cut semi-definite program (SDP). This holds for the uniform demands version (i.e. edge expansion). The same quantitative gap was proved earlier by Devanur, Khot, Saket, and Vishnoi [STOC 2006], following an integrality gap for non-uniform demands due to Khot and Vishnoi [FOCS 2005]. These previous constructions involve a complicated SDP solution and analysis, while our gap instance, vector solution, and analysis are somewhat simpler and more intuitive.

Furthermore, our approach is rather general, and provides a variety of different gap examples derived from quotients of the hypercube. It also illustrates why the lower bound is stuck at $\Omega(\log \log N)$, and why new ideas are needed in order to derive stronger examples.

1 Introduction

Certainly the notion of graph expansion plays a central role in the modern theory of computation. Moreover, given an input graph $G = (V, E)$, the computational problem of computing the least expanding set in G, or the extent to which G is an expander, is a fundamental one in algorithm design. If we let $E(S, \bar{S})$ denote the set of edges between $S \subseteq V$ and its complement and define

$$\Phi(G) = \min \left\{ \frac{|E(S, \bar{S})|}{|S||\bar{S}|} : S \subseteq V \right\},$$

then calculating $\Phi(G)$ (and the set which achieves the minimum) if the well-known *uniform* Sparsest Cut problem. Since the problem is NP-hard, much recent work has focused on approximating $\Phi(G)$.

The first such algorithm, due to Leighton and Rao [?], achieved an $O(\log N)$-approximation, where $N = |V|$, and was based on a linear programming relaxation that computes an all-pairs multi-commodity flow in G. Later, Linial, London, and Rabinovich [?], and Aumann and Rabani [?], found a connection between rounding this linear programming (and its generalizations) and the problem of embedding finite metric spaces into L_1.

Around this time, a natural semi-deifnite programming (SDP) relaxation was proposed. This relaxation can be written succinctly as

$$\mathsf{SDP}(G) = \min \left\{ \frac{\sum_{uv \in E} \|x_u - x_v\|^2}{\sum_{u,v \in V} \|x_u - x_v\|^2} : \|x_u - x_v\|^2 \leq \|x_u - x_w\|^2 + \|x_w - x_v\|^2 \ \forall u, v, w \in V \right\},$$

where the minimum ranges over all vectors $\{x_u\}_{u \in V} \subseteq \mathbb{R}^{N-1}$. The latter constraints are referred to alternatively as the "negative-type inequalities," the "ℓ_2^2 inequalities," or the "squared triangle inequalities," and the geometric constraints they place on the solution are still poorly understood.

In fact, Goemans and Linial [?, ?] conjectured that the integrality gap of this relaxation is only $O(1)$ (in fact, they conjectured that a more general "non-uniform" version of the problem satisfied this bound). In a seminal work of Arora, Rao, and Vazirani [?], it was shown that the integrality gap is at most $O(\sqrt{\log N})$, but the question of lower bounds on the integrality gap remained open, largely because of the difficulty of producing interesting systems of vectors that satisfied the ℓ_2^2 inequalities.

Finally, in a remarkable paper, Khot and Vishnoi [?] disproved the non-uniform Goemans-Linial conjecture using a connection with the Unique Games conjecture [?]. A year later, Devanur, Khot, Saket, and Vishnoi [?] showed how one can obtain a gap for the *uniform* version defined above. Their quantitative lower bound is $\Omega(\log \log N)$, and the exponential gap between this and the $O(\sqrt{\log N})$ upper bound still remains.

Problematically, both the constructions of [?] and [?] are shrouded in mystery. The construction and analysis have often been referred to as "difficult," "impenetrable," "extremely technical," and "magic" (the last description coming from the authors themselves). The goal of the present work is to present a simple, self-contained construction and analysis of an $\Omega(\log \log N)$ integrality gap. Our inputs instances, vector solutions, and analysis are all simpler and more intuitive than their counterparts in [?] and [?].

It is difficult to overestimate the importance of the Sparsest Cut problem, the preceding SDP, and its place in the larger theory of approximation algorithms. We mention, first of all, that the algorithm and analysis of [**?**] drove a huge wave of new results in approximation algorithms. Furthermore, the Sparsest Cut problem and the analysis of this SDP were some of the primary driving forces in the field of metric embeddings, and led to a number of beautiful results and connections. The SDP combines the flow-based constraints of the Leighton-Rao LP, together with the second (Laplacian) eigenvalue bound used in spectral partitioning (see [**?**] and also Section C), and in this sense represents a new frontier in algorithm design.

Finally, we mention that the uniform Sparsest Cut problem is still very poorly understood from the standpoint of approximation algorithms. It is known to be hard to approximate within $1+\epsilon_0$, for some small constant $\epsilon_0 > 0$, unless NP has subexponential-time algorithms [**?**], but no better lower bound is known, even assuming the unique games conjecture. On the other hand, as we previously mentioned, the best upper bound is $O(\sqrt{\log N})$.

1.1 Outline, and an intuitive overview

Our gap instances are simply quotients of the standard hypercube—which we will represent by $Q_n = \{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$—under some action by permutations of the coordinates. The sparsity of cuts in these graphs was studied by Khot and Naor [**?**], and those authors also suggested them as a possible source for integrality gaps.

For instance, consider the cyclic shift operator $\sigma(x_1, x_2, \ldots, x_n) = (x_2, \ldots, x_n, x_1)$, and define the quotient metric

$$d(u,v) = \min\left\{\|u - \sigma^i v\|_1 : i = 0, 1, \ldots, n-1\right\},$$

which is clearly σ-invariant, i.e. $d(u,v) = d(\sigma u, v) = d(u, \sigma v)$, and hence actually a metric on the orbits of Q_n under the action of σ. It is straightforward to verify that d satisfies the triangle inequality.

Our approach is simply to define vectors $\{x_u\}_{u\in Q_n}$ such that $\|x_u - x_v\|^2 \approx d(u,v)$ holds for all $u, v \in \mathcal{P}$, where $\mathcal{P}$ is a certain "pseudorandom" subset of Q_n, and $|Q_n \setminus \mathcal{P}| = o(|Q_n|)$. We use this connection (and the fact that d is a metric) to prove the triangle inequalities for $\{x_u\}_{u\in\mathcal{P}}$. We then map all the points of $Q_n \setminus \mathcal{P}$ to some fixed x_{u_0} for $u_0 \in \mathcal{P}$. Being such a small fraction of points, their contribution to the SDP is inconsequential.

For cyclic shifts, our vector solution is essentially the following,

$$x_u = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\sigma^i u)^{\otimes t}, \tag{1}$$

for some small $t = O(1)$ (see Section 3 for a more detailed overview). In general, we simply average over the action of a group, and take small tensor powers (see Section 2 for a review of tensor products).

Now, our $\mathcal{P}$ is essentially the set of points whose orbits are not too self-correlated, e.g. points $u \in Q_n$ with $\langle u, \sigma^i u\rangle \le n^{-1/3}$, say, for every $i \in \{1, 2, \ldots, n-1\}$. To show that $d(u,v) \approx \|x_u - x_v\|^2$ for $u, v \in \mathcal{P}$, we will assume that $\|x_u\| = 1$ for every $u \in \mathcal{P}$ (this is almost true, by virtue of the definition of $\mathcal{P}$). In this case, it suffices to prove that $1 - \langle x_u, x_v\rangle \approx 1 - \lambda(u,v)$, where

$$\lambda(u,v) = \max\left\{\langle u, \sigma^i v\rangle : i = 0, 1, \ldots, n-1\right\}$$

is the associated "quotient inner product."

To see that this holds, we write

$$\langle x_u, x_v \rangle = \sum_{i=0}^{n-1} \langle u, \sigma^i v \rangle^t. \tag{2}$$

Now, if $\lambda(u,v) \geq 1-\delta$, then $\langle x_u, x_v \rangle \geq (1-\delta)^t \geq 1-\delta t$. On the other hand, if $\langle x_u, x_v \rangle \geq 1-\delta$, we need to find a single $i \in [n]$ for which $\langle u, \sigma^i v \rangle \approx \langle x_u, x_v \rangle$. Since we are taking t^{th} powers in (2), any small inner products $\langle u, \sigma^j v \rangle$ are dampened out. But if there were two distinct indices i, j for which $\langle u, \sigma^i v \rangle$ and $\langle u, \sigma^j v \rangle$ were both moderately large, then $\langle u, \sigma^{i-j} u \rangle$ would also be large, which doesn't happen because $u \in \mathcal{P}$. Hence $\langle x_u, x_v \rangle$ can only be close to 1 if the contribution comes almost entirely from one shift. This matching property is precisely what yields the triangle inequalities.

Outline. A more precise version of this argument for cyclic shifts is presented in Section 3, while the full argument (and for general quotients) is given in Section 4. In Section A.1, we discuss why vector solutions like (1) are probably insufficient for going beyond a gap of $\Omega(\log \log N)$. It is suggested that the reader first review Section 2 for some definitions and terminology.

In Section B, we consider group actions where the groups are quite large (e.g. $\exp(\sqrt{n})$) so that (1) will no longer work, but a different embedding succeeds in giving a valid vector solution. Unfortunately, it is also fairly easy to see that this is example has an integrality gap of $O(\log \log N)$, but the technique may be useful for future constructions. Finally, in Section C, we discuss the SDP dual and give some open questions whose resolution would further simplify integrality gap constructions.

2 Preliminaries

We first discuss some preliminary notions and theorems that will be used throughout the paper.

Asymptotic notation. For expressions A and B, we will use the notation $A \lesssim B$ to denote $A = O(B)$, and $A \approx B$ to denote the conjunction of $A \lesssim B$ and $A \gtrsim B$.

Sparsity of graphs. We will consider undirected graphs $G = (V, E)$ where every edge (u, v) has a non-negative weight $w(u, v)$. For any subset $E' \subseteq E$ of edges, we write $w(E') = \sum_{e \in E'} w(e)$. For two sets $S, T \subseteq V$, we write $E(S, T)$ for the set of edges with one endpoint in S and one in T.

For a subset $S \subseteq V$, we use

$$\Phi(S) = \frac{w(E(S, \bar{S}))}{|S||\bar{S}|}$$

to denote the *sparsity of* S. We then write $\Phi(G) = \min_{S \subseteq V} \Phi(S)$ for the *sparsest cut value of* G.

We will be particularly interested in graphs derived from the (unweighted) n-dimensional hypercube $Q_n = \left\{ \frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}} \right\}^n$. We will use Q_n to denote the set of vertices in the n-cube, and $E(Q_n)$ to denote the set of edges. The classical discrete isoperimetric inequality shows that if we write $S_i = \{x \in Q_n : x_i < 0\}$, then for every $i \in [n]$,

$$\Phi(Q_n) = \Phi(S_i) = \frac{4|E(S_i, \bar{S}_i)|}{|Q_n|^2} \approx |Q_n|^{-1}.$$

A well-known theorem of Kahn, Kalai, and Linial [**?**] then asserts the following.

Theorem 2.1 (KKL Theorem)**.** *For any $S \subseteq Q_n$, there exists an $i \in [n]$ for which*

$$\frac{|E(S,\bar{S}) \cap E(S_i,\bar{S}_i)|}{|S||\bar{S}|} \gtrsim \frac{\log n}{n}\,\Phi(Q_n).$$

Weighted "quotients" of the cube. Let Γ be any group acting on $[n] = \{1, 2, \ldots, n\}$ by permutations. We can naturally extend Γ to act on Q_n via $\pi(x_1, \ldots, x_n) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$ for any $\pi \in \Gamma$. For an element $u \in Q_n$, we use Γu to denote the Γ-orbit of u. We refer to a subset $S \subseteq Q_n$ as Γ-invariant if $\Gamma S = S$.

We define a weighted graph Q_n/Γ as follows. The vertices are simply those of Q_n, and the edges are $E(Q_n) \cup E'$, where $E' = \{(u, v) : u \in \Gamma v\}$. We define

$$w(e) = \begin{cases} 1 & e \in E(Q_n) \\ 2^{2n} & e \in E'. \end{cases}$$

The point of this choice is to ensure that $\Phi(Q_n/\Gamma) = \Phi(S)$ is always achieved by a Γ-invariant set S, since separating any Γ-orbit involves cutting an edge of very large value. (Note that, because we are only using weights which are polynomial in the graph size, our gap examples can easily be made unweighted.)

We recall that Γ is said to act *transitively* on $[n]$ if for every $i, j \in [n]$, there exists a permutation $\pi \in \Gamma$ with $\pi(i) = j$. From Theorem 2.1, one can easily derive the following.

Theorem 2.2 (Transitive actions)**.** *If Γ acts transitively on $[n]$, then $\Phi(Q_n/\Gamma) \gtrsim \Phi(Q_n)\log n$.*

Proof. We know that $\Phi(Q_n/\Gamma) = \Phi(S)$ for some Γ-invariant set S. By Theorem 2.1, there exists an $i \in [n]$ for which

$$\frac{|E(S,\bar{S}) \cap E(S_i,\bar{S}_i)|}{|S||\bar{S}|} \gtrsim \frac{\log n}{n}\,\Phi(Q_n).$$

But for any other $j \in [n]$, there exists an action $\pi \in \Gamma$ with $\pi(i) = j$, hence

$$\frac{|E(\pi(S),\pi(\bar{S})) \cap E(S_j,\bar{S}_j)|}{|\pi(S)||\pi(\bar{S})|} = \frac{|E(S,\bar{S}) \cap E(S_i,\bar{S}_i)|}{|S||\bar{S}|},$$

implying that

$$\Phi(S) = \sum_{j=1}^{n} \frac{|E(S,\bar{S}) \cap E(S_j,\bar{S}_j)|}{|S||\bar{S}|} = n \cdot \frac{|E(S,\bar{S}) \cap E(S_i,\bar{S}_i)|}{|S||\bar{S}|} \gtrsim \Phi(Q_n)\log n.$$

$\square$

The Sparsest Cut SDP. Given a weighted graph $G = (V, E)$, we recall the standard SDP relaxation of Sparsest Cut,

$$\mathsf{SDP}(G) = \min\left\{ \frac{\sum_{uv\in E} w(u,v)\|x_u - x_v\|^2}{\sum_{u,v\in V}\|x_u - x_v\|^2} : \|x_u - x_v\|^2 \le \|x_u - x_w\|^2 + \|x_w - x_v\|^2 \ \forall u, v, w \in V \right\},$$

where the minimum is taken over all choices of vectors $\{x_u\}_{u\in V}$ lying in some finite-dimensional Euclidean space. It is well-known that $\mathsf{SDP}(Q_n) = \Phi(Q_n) \approx |Q_n|^{-1}$.

We say that a vector solution $\{x_u\}_{u\in Q_n}$ is Γ-invariant if $x_u = x_{\pi(u)}$ for all $u \in Q_n$ and $\pi \in \Gamma$. Observe that a Γ-invariant solution for the Sparsest Cut SDP on Q_n/Γ has value

$$\frac{\sum_{uv\in E(Q_n)} \|x_u - x_v\|^2}{\sum_{u,v\in Q_n} \|x_u - x_v\|^2},$$

since all elements of a Γ-orbit are mapped to the same vector.

Weak triangle inequalities and pseudometrics. For the sake of exposition, we will also define an "SDP value" for solutions satisfying a weak form of the triangle inequalities. We recall that for any set X, a non-negative, symmetric function $d : V \times V \to \mathbb{R}$ is called a *pseudometric on* V if it satisfies the triangle inequalities, i.e. $d(u, v) \le d(u, w) + d(w, v)$ for all $u, v, w \in V$, and additionally $d(u, u) = 0$ for all $u \in V$.

For any $\beta \ge 1$, let

$$\mathsf{SDP}_\beta(G) = \min \left\{ \frac{\sum_{uv\in E} w(u,v)\|x_u - x_v\|^2}{\sum_{u,v\in V} \|x_u - x_v\|^2} : d(u,v) \le \|x_u - x_v\|^2 \le \beta d(u,v) \right\},$$

where the minimum is over all choices of vectors $\{x_u\}_{u\in V}$, and additionally over all pseudometrics d on V. Observe that $\mathsf{SDP}(G) = \mathsf{SDP}_1(G)$. One might also note that the Arora-Rao-Vazirani algorithm [**?**], and all known analyses derived from it, only use the weaker $\mathsf{SDP}_{O(1)}$ inequalities.

Tensoring. We recall that for two vectors $x, y \in \mathbb{R}^k$ and $t \in \mathbb{N}$, we have the tensored vectors $x^{\otimes t}, y^{\otimes t} \in \mathbb{R}^{k^t}$ which satisfy $\langle x^{\otimes t}, y^{\otimes t}\rangle = \langle x, y\rangle^t$.

Finally, we need the following tail inequality.

Lemma 2.3 (Hoeffding bound)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables with $\mathbb{E}X_i = 0$ for every $i \in [n]$. Then,*

$$\Pr\left[\left|\sum_{i=1}^n X_i\right| \ge L\right] \le 2\exp\left(\frac{-L^2}{2\sum_{i=1}^n \|X_i\|_\infty^2}\right).$$

3 A simple example: Cyclic shifts

Consider the cyclic shift operator $\sigma : [n] \to [n]$ defined by $\sigma(i) = (i + 1) \bmod n$, and let $\Gamma = \{\sigma^0, \sigma^1, \ldots, \sigma^{n-1}\}$ be the group of permutations generated by σ. By Theorem 2.2, we have $\Phi(Q_n/\Gamma) \gtrsim \Phi(Q_n)\log n$. On the other hand, we will now show that the "weak" SDP value of Q_n/Γ is approximately $\mathsf{SDP}(Q_n)$, thus exhibiting a (weak) SDP gap of $\Omega(\log n) = \Omega(\log\log |Q_n|)$. This will illustrate the main ideas behind our proof for general quotients, and the true SDP value will be analyzed in the next section.

Theorem 3.1. *For $n \in \mathbb{N}$, $\mathsf{SDP}_{16}(Q_n/\Gamma) \lesssim \mathsf{SDP}(Q_n)$.*

Proof. For every $u \in Q_n$, we define the vector

$$x_u = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\sigma^i u)^{\otimes 8},$$

and put $\widetilde{x_u} = x_u/\|x_u\|$. Observe that

$$\langle x_u, x_v \rangle = \frac{1}{n} \sum_{i,j=0}^{n-1} |\langle \sigma^i u, \sigma^j v \rangle|^8 = \sum_{i=0}^{n-1} |\langle u, \sigma^i v \rangle|^8. \tag{3}$$

We now define a subset of "pseudorandom" vertices of Q_n whose orbits under Γ are not too self-correlated,

$$\mathcal{P} = \left\{ u \in Q_n : \sum_{i=0}^{n-1} |\langle u, \sigma^i u \rangle|^6 \le 1 + \frac{1}{4n} \right\}.$$

Note that, by Cauchy-Schwarz, for $u, v \in \mathcal{P}$, we have

$$\sum_{i=0}^{n-1} |\langle u, \sigma^i v \rangle|^6 \le \sqrt{\sum_{i=0}^{n-1} |\langle u, \sigma^i u \rangle|^6} \sqrt{\sum_{i=0}^{n-1} |\langle v, \sigma^i v \rangle|^6} \le 1 + \frac{1}{4n}. \tag{4}$$

(To see this, observe that $\sum_{i=0}^{n-1} |\langle u, \sigma^i v \rangle|^6$ is an inner product, as in (3).)

Most vertices are pseudorandom. For any $u \in Q_n$, we can write

$$\langle u, \sigma u \rangle = \sum_{\substack{1 \le i \le n \\ i \text{ even}}} u_i u_{\sigma(i)} + \sum_{\substack{1 \le i \le n \\ i \text{ odd}}} u_i u_{\sigma(i)} = T + T',$$

where each u_i appears exactly once in each of the sums T and T'. It is easy to see that a similar decomposition holds for $\langle u, \sigma^i u \rangle$ for any $i \in \{1, 2, \ldots, n-1\}$.

Therefore by Lemma 2.3, we have

$$\Pr_{u \in Q_n} \left[|\langle u, \sigma^i u \rangle| \ge 2t/\sqrt{n} \right] \le \Pr\left[|T| \ge t/\sqrt{n} \right] + \Pr\left[|T'| \ge t/\sqrt{n} \right] \le 4e^{-t^2/2}, \tag{5}$$

since each of T and T' is a sum of i.i.d. uniform elements of $\{\pm \frac{1}{n}\}$. Setting $t = n^{1/3}/2$ and taking a union bound over $i = 1, 2, \ldots, n-1$ yields

$$\Pr_{u \in Q_n} \left[\sum_{i=0}^{n-1} |\langle u, \sigma^i v \rangle|^6 > 1 + \frac{1}{4n} \right] \le 4ne^{-n^{2/3}/8} \le n^{-2}, \tag{6}$$

for n sufficiently large, hence $|\mathcal{P}| \ge |Q_n|(1 - n^{-2})$.

The SDP value. Fix some $u_0 \in \mathcal{P}$. Our final SDP solution will consist of the vectors $\{x'_u\}_{u \in Q_n}$ with $x'_u = \widetilde{x_u}$ for $u \in \mathcal{P}$ and $x'_u = \widetilde{x_{u_0}}$ otherwise. Thus we will only need to verify the weak triangle inequalities for $\{\widetilde{x_u}\}_{u \in \mathcal{P}}$. It is clear that our proposed SDP solution is Γ-invariant.

For an edge $(u, v) \in E(Q_n)$, using (3), we have

$$\langle x_u, x_v \rangle \ge |\langle u, v \rangle|^8 = \left(1 - \frac{2}{n}\right)^8 \ge 1 - \frac{16}{n}.$$

Hence for $u, v \in \mathcal{P}$ with $(u, v) \in E(Q_n)$, we have $\|\widetilde{x_u} - \widetilde{x_v}\|^2 = O(1/n)$. In particular,

$$\sum_{(u,v) \in E(Q_n)} \|x'_u - x'_v\|^2 \lesssim \frac{|E(Q_n)|}{n} + 4|E(Q_n \setminus \mathcal{P})| \lesssim \frac{|E(Q_n)|}{n}, \tag{7}$$

since $|Q_n \setminus \mathcal{P}| \leq |Q_n|/n^2$.

On the other hand, if we choose $u, v \in Q_n$ at random, then for any $i \in [n]$, using Lemma 2.3,

$$\Pr_{u,v \in Q_n}\left[|\langle u, \sigma^i v\rangle| \geq t/\sqrt{n}\right] \leq 2e^{-t^2/2}.$$

Setting $t \approx \sqrt{\log n}$ and taking a union bound over all $i \in [n]$ shows that for n sufficiently large, $\Pr_{u,v\in Q_n}[|\langle x_u, x_v\rangle| \geq \frac{1}{4}] \leq \frac{1}{2}$. In particular,

$$\sum_{u,v\in Q_n} \|x'_u - x'_v\|^2 \geq \sum_{u,v\in\mathcal{P}} \|\widetilde{x_u} - \widetilde{x_v}\|^2 \approx \sum_{u,v\in\mathcal{P}} \|x_u - x_v\|^2 \gtrsim |\mathcal{P}|^2 \gtrsim |Q_n|^2.$$

Combining the preceding line with (7) shows that the value of the potential SDP solution $\{x'_u\}_{u\in Q_n}$ is $O(|Q_n|^{-1}) = O(\mathsf{SDP}(Q_n))$.

Verifying the weak triangle inequalities. We are thus left to verify the weak triangle inequalities for $\{\widetilde{x_u}\}_{u\in\mathcal{P}}$. To this end, we will define a cyclic shift-invariant metric d on Q_n and then show that for $u, v \in \mathcal{P}$, we have $d(u, v) \approx \|\widetilde{x_u} - \widetilde{x_v}\|^2$.

Let $\lambda(u, v) = \max\{|\langle u, \sigma^i v\rangle| : i \in [n]\}$ and put $d(u, v) = 1 - \lambda(u, v)^8$. It is clear that $d(u, v) = d(\sigma u, v) = d(u, \sigma v)$. Next, observe that for any $u, v, w \in Q_n$, we have

$$1 + \langle u, v\rangle \geq \langle u, w\rangle + \langle v, w\rangle,$$

since the inequality $1 + xy \geq xz + yz$ for $x, y, z \in \{-1, 1\}$ is straightforward to verify. Observing that $u^{\otimes 8}, v^{\otimes 8}, w^{\otimes 8} \in Q_{n^8}$, it follows that

$$1 + |\langle u, v\rangle|^8 \geq |\langle u, w\rangle|^8 + |\langle v, w\rangle|^8. \tag{8}$$

Now suppose that $i, j \in \mathbb{N}$ are such that $\lambda(u, w) = |\langle \sigma^i u, w\rangle|$ and $\lambda(v, w) = |\langle \sigma^j v, w\rangle|$. In that case, we have

$$\begin{aligned} 1 + \lambda(u, v)^8 &\geq 1 + |\langle \sigma^i u, \sigma^j v\rangle|^8 \\ &\geq |\langle \sigma^i u, w\rangle|^8 + |\langle \sigma^j v, w\rangle|^8 \\ &= \lambda(u, w)^8 + \lambda(v, w)^8, \end{aligned}$$

where the second inequality is simply (8). Rearranging shows that the preceding inequality is precisely $d(u, v) \leq d(u, w) + d(v, w)$, i.e. that d satisfies the triangle inequality.

We are thus left to show that $1 - \lambda(u, v)^8 \approx 1 - \langle \widetilde{x_u}, \widetilde{x_v}\rangle$ for $u, v \in \mathcal{P}$. If $\lambda(u, v) = 1$, then both expressions are 0, so we may assume that $\lambda(u, v) \neq 1$. One direction is easy: Using the fact that if $\lambda(u, v) \neq 1$, then $\lambda(u, v)^8 \leq \lambda(u, v) \leq 1 - \frac{2}{n}$, we have

$$\begin{aligned} 1 - \langle \widetilde{x_u}, \widetilde{x_v}\rangle &\leq 1 - (1 + \tfrac{1}{4n})^{-1}\langle x_u, x_v\rangle \\ &\leq 1 - (1 + \tfrac{1}{4n})^{-1}\lambda(u, v)^8 \\ &\leq 1 - (1 - \tfrac{1}{4n})\lambda(u, v)^8 \\ &\leq 2\left[1 - \lambda(u, v)^8\right]. \end{aligned}$$

Now, the key to satisfying the (weak) triangle inequalities is the following simple calculation:

$$\langle \widetilde{x_u}, \widetilde{x_v}\rangle \leq \langle x_u, x_v\rangle = \sum_{i=0}^{n-1} |\langle u, \sigma^i v\rangle|^8 \leq \lambda(u, v)^2 \sum_{i=0}^{n-1} |\langle u, \sigma^i v\rangle|^6 \leq (1 + \tfrac{1}{4n})\lambda(u, v)^2,$$

where in the last inequality, we have used $u, v \in \mathcal{P}$. Thus assuming $\langle \widetilde{x_u}, \widetilde{x_v} \rangle = 1 - \delta$, we get

$$\lambda(u,v)^8 \geq \left((1-\delta)\left(1 - \frac{1}{4n}\right)\right)^4 \geq 1 - 4\left(\delta + \frac{1}{4n}\right),$$

but $\lambda(u,v) \leq 1 - \frac{2}{n}$, hence $\delta \geq \frac{1}{4n}$ so that $\lambda(u,v)^8 \geq 1 - 8\delta$, implying $1 - \lambda(u,v)^8 \leq 8(1 - \langle \widetilde{x_u}, \widetilde{x_v} \rangle)$. $\square$

4 General quotients

In the present section, we derive SDP solutions for "pseudorandom" subsets of general quotient constructions. Unlike the previous section, we will ensure that these solutions satisfy the full triangle inequalities.

4.1 Metrics and kernels

Fix a subgroup Γ acting on $[n]$ by permutations. We let $\psi_\Gamma = \max\{|\Gamma u| : u \in Q_n\}$ be the maximum size of any Γ-orbit. For $u, v \in Q_n$, we define

$$\lambda(u,v) = \max_{\pi \in \Gamma} |\langle u, \pi v \rangle|,$$

and for every $t \in \mathbb{N}$,

$$\alpha_t(u,v) = \sum_{\pi \in \Gamma} |\langle u, \pi v \rangle|^{2t},$$

and

$$\overline{\alpha_t}(u,v) = \frac{\alpha_t(u,v)}{\sqrt{\alpha_t(u,u)\,\alpha_t(v,v)}}.$$

Finally, we define two distance functions on Q_n corresponding to λ and α_t, respectively. For $s, t \in \mathbb{N}$, define

$$\begin{aligned} \rho_{s,t}(u,v) &= 1 - \left(\frac{1}{2} + \frac{1}{2}\lambda(u,v)^{2t}\right)^s \\ K_{s,t}(u,v) &= 1 - \left(\frac{1}{2} + \frac{1}{2}\overline{\alpha_t}(u,v)\right)^s. \end{aligned}$$

Lemma 4.1. *For every $t \in \mathbb{N}$, both α_t and $\overline{\alpha_t}$ are positive semi-definite kernels on Q_n. For every $s \in \mathbb{N}$, the same is true for $(u,v) \mapsto \left(\frac{1}{2} + \frac{1}{2}\overline{\alpha_t}(u,v)\right)^s$.*

Proof. If we define $f : Q_n \to \mathbb{R}^{n^{2t}}$ by $f(u) = |\Gamma|^{-1/2} \sum_{\pi \in \Gamma} (\pi u)^{\otimes 2t}$ then $\alpha_t(u,v) = \langle f(u), f(v) \rangle$ and $\overline{\alpha_t}(u,v) = \left\langle \frac{f(u)}{\|f(u)\|_2}, \frac{f(v)}{\|f(v)\|_2} \right\rangle$. For the final implication, note that the sum of two PSD kernels is PSD, and also a positive integer power of a PSD kernel is PSD. $\square$

From Lemma 4.1 and the fact that $0 \leq \overline{\alpha_t}(u,v) \leq 1$ for all $u, v \in Q_n$, one verifies that $K_{s,t}$ is a negative-definite kernel on Q_n, i.e. there exists a system of (unit) vectors $\{x_u\}_{u \in Q_n}$ such that $\|x_u - x_v\|^2 = K_{s,t}(u,v)$.

It is clear that both functions $\rho_{s,t}$ and $K_{s,t}$ are Γ-invariant in both coordinates. We will now show that $\rho_{s,t}$ is a metric. In Section 4.2, we will show that $K_{s,t}(u,v) \approx \rho_{s,t}(u,v)$ for "pseudorandom" $u, v \in Q_n$. This will motivate our analysis of the metrical properties of $K_{s,t}$ in Section A.

Lemma 4.2. *If $0 \le a \le b \le c \le 1$ and $1+a \ge b+c$, then for any $r \ge 1$, $a^r - b^r - c^r \ge a - b - c$. In particular, for any $a,b,c \in [0,1]$, $1+a \ge b+c$ implies $1+a^r \ge b^r + c^r$.*

Proof. We may assume that $a \neq 1$. In this case, write b and c as a convex combination of a and 1 as follows: $b = \frac{1-b}{1-a}a + (1 - \frac{1-b}{1-a})$ and $c = \frac{1-c}{1-a}a + (1 - \frac{1-c}{1-a})$. Now, using the fact that $x - x^r$ is concave for $x \in [0,1]$ and $r \ge 1$, write

$$(b - b^r) + (c - c^r) \ge \frac{1-b}{1-a}(a - a^r) + \frac{1-c}{1-a}(a - a^r) \ge \frac{2-b-c}{1-a}(a - a^r) \ge a - a^r,$$

where the final inequality follows from $1 + a \ge b + c$. To verify the second claim of the lemma, note that if $a > b$ or $a > c$, then $1 + a^r \ge b^r + c^r$ holds trivially. □

Corollary 4.3. *Let X be any set, $U : X \times X \to [0,1]$, and $s \ge 1$. If $D'(x,y) = 1 - (\frac{1}{2} + \frac{1}{2}U(x,y))^s$ is a pseudometric on X, then so is $D(x,y) = 1 - (\frac{1}{2} + \frac{1}{2}U(x,y))^{s'}$ for any $s' \ge s$.*

Proof. The triangle inequality for D on $x, y, z \in X$ reduces to verifying

$$1 + (\tfrac{1}{2} + \tfrac{1}{2}U(x,y))^{s'} \ge (\tfrac{1}{2} + \tfrac{1}{2}U(x,z))^{s'} + (\tfrac{1}{2} + \tfrac{1}{2}U(y,z))^{s'}.$$

Since $s' \ge s$, Lemma 4.2 implies that this reduces to the triangle inequality for D'. □

Lemma 4.4. *For every $s, t \in \mathbb{N}$, $\rho_{s,t}$ is a pseudometric on Q_n.*

Proof. By Corollary 4.3, it suffices to prove this for $\rho_{1,t}$. It's clear that for any $u \in Q_n$, $\rho_{1,t}(u,u) = 0$ because $\lambda(u,u) = 1$. Now fix $u, v, w \in Q_n$. The triangle inequality $\rho_{1,t}(u,v) \le \rho_{1,t}(u,w) + \rho_{1,t}(v,w)$ reduces to verifying

$$1 + \lambda(u,v)^{2t} \ge \lambda(u,w)^{2t} + \lambda(v,w)^{2t}. \tag{9}$$

Suppose that $\lambda(u,w) = |\langle \pi u, w\rangle|$ and $\lambda(v,w) = |\langle v, \pi' w\rangle|$. Then,

$$\begin{aligned} \lambda(u,v)^{2t} &\ge |\langle \pi u, \pi' v\rangle|^{2t} \\ &\ge |\langle \pi u, w\rangle|^{2t} + |\langle \pi' v, w\rangle|^{2t} - 1 \qquad (10) \\ &= \lambda(u,w)^{2t} + \lambda(v,w)^{2t} - 1, \end{aligned}$$

where (10) follows just as in (8). □

Before turning to the precise relationship between $K_{s,t}$ and $\rho_{s,t}$, we calculate $\rho_{s,t}(u,v)$ for edges and for random pairs in Q_n.

Lemma 4.5 (Edges)**.** *If $u, v \in E(Q_n)$, then $\rho_{s,t}(u,v) \le \frac{2st}{n}$.*

Proof. Observe that

$$\lambda(u,v)^{2t} \ge \left(1 - \frac{2}{n}\right)^{2t} \ge 1 - \frac{4t}{n},$$

hence $\rho_{s,t}(u,v) \le 1 - (1 - \frac{2t}{n})^s \le \frac{2st}{n}$. □

The next lemma is a straightforward application of Lemma 2.3 and a union bound.

Lemma 4.6 (Random pairs)**.** *Suppose that $u, v \in Q_n$ are chosen independently and uniformly at random. Then,*

$$\Pr\left[\lambda(u,v)^{2t} \geq L\right] \leq 2\psi_G \exp\left(\frac{-L^{1/t} n}{2}\right).$$

In particular, for any $s, t \in \mathbb{N}$, if $\psi_\Gamma \leq 2^{0.1n}$, then

$$\Pr[\rho_{s,t}(u,v) \geq \tfrac{1}{4}] \geq \Pr[\lambda(u,v)^{2t} \leq \tfrac{1}{2}] \geq \tfrac{1}{2}.$$

4.2 Pseudorandom orbits and $\rho_{s,t} \approx K_{s,t}$

For $r \in \mathbb{N}$, define

$$\mathcal{P}_r(\eta) = \{u \in Q_n : \alpha_r(u,u) \leq 1 + \eta\}$$

as the set of all elements whose Γ-orbits are not too self-correlated. Note that, by Cauchy-Schwarz, $u, v \in \mathcal{P}_r(\eta)$ implies $\alpha_r(u,v) \leq \sqrt{\alpha_r(u,u)\,\alpha_r(v,v)} \leq 1 + \eta$.

The next lemma is central. It says that if $\alpha_t(u,v)$ is large and u, v are pseudorandom, then the contribution to $\alpha_t(u,v)$ comes mainly from a single large "matching" term, i.e. u is strongly correlated with some element of Γv.

Lemma 4.7. *Let $t > r$ and $\delta \in [0,1]$. If $u, v \in \mathcal{P}_r(\eta)$ and $\alpha_t(u,v) \geq 1 - \delta$, then*

$$\lambda(u,v)^{2(t-r)} \quad \geq \quad 1 - \delta - \eta.$$

Proof. We have,

$$\alpha_t(u,v) \leq \lambda(u,v)^{2t-2r} \sum_{\pi \in \Gamma} |\langle u, \pi v\rangle|^{2r} = \lambda(u,v)^{2(t-r)} \alpha_r(u,v) \leq (1+\eta)\lambda(u,v)^{2(t-r)}.$$

It follows that $\lambda(u,v)^{2(t-r)} \geq \frac{1-\delta}{1+\eta} \geq 1 - \delta - \eta$. □

Theorem 4.8 (Weak triangle inequality for $K_{s,t}$)**.** *For every $r, s \in \mathbb{N}$ and $u, v \in \mathcal{P}_r(\frac{1}{4n})$,*

$$\rho_{s,2r}(u,v) \approx K_{s,2r}(u,v),$$

where the implicit constant is independent of the given parameters.

Proof. Let $\eta = \frac{1}{4n}$ and $t = 2r$, and suppose that $u, v \in \mathcal{P}_r(\eta)$. If $\lambda(u,v) = 1$, then $\overline{\alpha_t}(u,v) = 1$ as well, hence $\rho_{s,t}(u,v) = K_{s,t}(u,v)$.

Now suppose that $\lambda(u,v) \neq 1$. In that case,

$$\lambda(u,v)^{2t} \leq \left(1 - \tfrac{2}{n}\right)^{2t} \leq 1 - \tfrac{2}{n}. \tag{11}$$

Assume that $\overline{\alpha_t}(u,v) = 1 - \delta$ for some $\delta \in [0,1]$. Then, $\alpha_t(u,v) \geq \overline{\alpha_t}(u,v) \geq 1 - \delta$, so Lemma 4.7 implies that $\lambda(u,v)^{2t} \geq (1 - \delta - \eta)^2 \geq 1 - 2(\delta + \eta)$, and from (11), we conclude that $\delta \geq \frac{3}{4n}$. This, in turn, implies that $\eta \leq \delta/3$, which gives $\lambda(u,v)^{2t} \geq 1 - 3\delta$.

Finally, we observe that

$$\overline{\alpha_t}(u,v) \geq (1-\eta)\alpha_t(u,v) \geq (1 - \delta/3)\alpha_t(u,v) \geq (1 - \delta/3)\lambda(u,v)^{2t},$$

hence $\lambda(u,v)^{2t} \leq (1-\delta)(1+\delta/3) \leq 1 - \frac{2\delta}{3}$. We have thus shown that $1 - \lambda(u,v)^{2t}$ and $1 - \overline{\alpha_t}(u,v)$ are within an $O(1)$ factor for all $u, v \in \mathcal{P}_r(\eta)$. □

Verification of the full triangle inequalities occurs in Appendix A.

Acknowledgements

We thank Gil Kalai and Muli Safra for a variety discussions related to our project.

References

A Triangle inequalities

In this section, we verify that $K_{22,t}$ is a pseudometric on $\mathcal{P}_r(\frac{1}{(4n)^2})$ for $t = O(r)$. In other words, the corresponding vectors form a valid SDP solution.

Theorem A.1. *For some $t = O(r)$, $K_{22,t}$ is a pseudometric on $\mathcal{P}_r(\frac{1}{(4n)^2})$.*

Proof. Let $\eta = \frac{1}{(4n)^2}$, and fix $u, v, w \in \mathcal{P}_r(\eta)$. To prove triangle inequality for $K_{s,t}$, it suffices to show that

$$1 + (\tfrac{1}{2} + \tfrac{1}{2}\overline{\alpha_t}(u, v))^s \geq (\tfrac{1}{2} + \tfrac{1}{2}\overline{\alpha_t}(u, w))^s + (\tfrac{1}{2} + \tfrac{1}{2}\overline{\alpha_t}(v, w))^s.$$

If both $\overline{\alpha_t}(u, w), \overline{\alpha_t}(v, w) \leq \frac{15}{16}$, then for $s = 22$, both terms are the right hand side are at most $\frac{1}{2}$, and the inequality is trivially satisfied. So we assume that $\overline{\alpha_t}(u, w) \geq \frac{15}{16}$ for the remainder of the proof.

By Corollary 4.3, to prove triangle inequality for $K_{22,t}$, it suffices to prove the same inequality for $K_{1,t}$ or $K_{2,t}$, i.e. one of the following inequalities.

$$\begin{aligned} 3 + \overline{\alpha_t}(u, v)[2 + \overline{\alpha_t}(u, v)] &\geq \overline{\alpha_t}(u, w)[2 + \overline{\alpha_t}(u, w)] + \overline{\alpha_t}(v, w)[2 + \overline{\alpha_t}(v, w)] \\ 1 + \overline{\alpha_t}(u, v) &\geq \overline{\alpha_t}(u, w) + \overline{\alpha_t}(v, w). \end{aligned}$$

Clearly both of these hold if $\lambda(u, w) = 1$ or if $\lambda(w, v) = 1$, so we assume this is not the case, and we are left to prove one of the following.

$$3 + \alpha_t(u, v)[2 + \alpha_t(u, v)] \geq \alpha_t(u, w)[2 + \alpha_t(u, w)] + \alpha_t(v, w)[2 + \alpha_t(v, w)] + 5\eta \tag{12}$$
$$1 + \alpha_t(u, v) \geq \alpha_t(u, w) + \alpha_t(v, w) + 2\eta, \tag{13}$$

recalling that $\overline{\alpha_t}(u, v) \leq \alpha_t(u, v) \leq (1 + \eta)\overline{\alpha_t}(u, v)$ for all $u, v \in \mathcal{P}_r(\eta)$. We remark that this loss in η will be acceptable beacuse when two points $u, v \in Q_n$ are distinct, they have $|\langle u, v\rangle| \leq 1 - \frac{4}{n}$, giving us $\approx \frac{1}{n}$ slack when the orbits of u, v, and w are distinct.

Case I (Strong matching): $\lambda(u, w), \lambda(v, w) \geq 1 - \frac{1}{2t}$.

Let $\lambda(v, w) = 1 - \delta, \lambda(u, w) = 1 - \varepsilon$, and observe that $\lambda(u, v) \geq 1 - (\delta + \varepsilon)$ by (9). Also, since $\lambda(u, w) \neq 1$ and $\lambda(w, v) \neq 1$, we have $\delta, \varepsilon \geq \frac{4}{n}$, and in particular $\eta \leq \varepsilon\delta$. We will verify (13). Write,

$$\alpha_t(v, w) \leq (1 - \delta)^{2t} + \left(\alpha_r(v, w) - (1 - \delta)^{2r}\right)^{t/r} \leq (1 - \delta)^{2t} + (\eta + 2r\delta)^{t/r}, \tag{14}$$

and similarly $\alpha_t(u, w) \leq (1 - \varepsilon)^{2t} + (\eta + 2r\varepsilon)^{t/r}$.

Using the preceding inequalities, to prove (13), it suffices to show that

$$1 + (1 - (\delta + \varepsilon))^{2t} - (1 - \delta)^{2t} - (1 - \varepsilon)^{2t} \geq (\eta + 2r\delta)^{t/r} + (\eta + 2r\varepsilon)^{t/r} + 5\eta. \tag{15}$$

But we have,

$$\begin{aligned}
1+(1-(\delta+\varepsilon))^{2t}-(1-\delta)^{2t}-(1-\varepsilon)^{2t} &= \sum_{i=2}^{2t}(-1)^i\binom{2t}{i}\left[\sum_{j=1}^{i-1}\binom{i}{j}\delta^j\varepsilon^{i-j}\right] \\
&\geq 2\binom{2t}{2}\delta\varepsilon-\binom{2t}{3}3\delta\varepsilon(\delta+\varepsilon) \\
&= t(2t-1)\delta\varepsilon\left([1-2(t-1)\delta]+[1-2(t-1)\varepsilon]\right) \\
&\geq 2t(2t-1)\delta\varepsilon\left(1-\frac{2(t-1)}{2t}\right) \\
&= (2t-1)\delta\varepsilon \\
&\geq ((2r+1)\delta)^{t/r}+((2r+1)\varepsilon)^{t/r}+5\varepsilon\delta,
\end{aligned}$$

where the final inequality holds for some $t=O(r)$ chosen large enough. This proves (15), recalling that $\eta\leq\varepsilon\delta$.

Case II (Weak matching): $\lambda(v,w)\leq 1-\frac{1}{2t}$.

Suppose that $\alpha_t(u,w)=1-\delta$. Our aim is to prove (12), which we write as

$$2\left(\alpha_t(v,w)-\alpha_t(u,v)\right)+\left(\alpha_t(v,w)-\alpha_t(u,v)\right)\left(\alpha_t(v,w)+\alpha_t(u,v)\right)\leq\delta(4-\delta)-2\eta. \tag{16}$$

Note that since $\overline{\alpha_t}(u,w)\geq\frac{15}{16}$, we have $\delta\leq\frac{1}{16}$. Furthermore, by Lemma 4.7, we have $\lambda(u,w)\geq 1-\frac{\delta+\eta}{2(t-r)}$. In particular, for $t=O(r)$ chosen large enough, we have $\lambda(u,w)\geq 1-\frac{1}{2t}$, which explains why cases I and II are exhaustive.

Now, if $\alpha_t(v,w)\geq 0.65$, then Lemma 4.7 implies $\lambda(v,w)\geq 1-\frac{0.35+\eta}{2(t-r)}\geq 1-\frac{0.45}{t}$ for $t\geq 2r$, which contradicts our assumption. We conclude that $\alpha_t(v,w)\leq 0.65$. In this case, we may assume that $\alpha_t(u,v)\leq 0.7$, since otherwise (13) is trivially satisfied, thus we have $\alpha_t(u,v),\alpha_t(v,w)\leq 0.7$.

The main idea in the "weak matching" case is to show that $\alpha_t(u,v)\gtrsim\alpha_t(v,w)$, but we cannot rely on a single "matched pair" (i.e. the triangle inqualities for λ) to do this. Instead, we argue that $\alpha_t(u,v)$ receives a large contribution on average.

To this end, write $\lambda(u,w)=1-\beta$, and let $\pi_0\in\Gamma$ be such that $|\langle\pi_0 u,w\rangle|=\lambda(u,w)$. Then,

$$\alpha_t(u,v)=\sum_{\pi\in\Gamma}|\langle\pi_0 u,\pi v\rangle|^{2t}\geq\sum_{\pi\in\Gamma}\left[\max(0,|\langle\pi_0 u,w\rangle|+|\langle w,\pi v\rangle|-1)\right]^{2t}\geq\sum_{\pi\in\Gamma}\left[\max(0,|\langle w,\pi v\rangle|-\beta)\right]^{2t}.$$

Let $I=\{\pi\in\Gamma:|\langle w,\pi v\rangle|\geq\beta\}$, and observe that

$$\sum_{\pi\notin I}|\langle w,\pi v\rangle|^{2t}\leq\beta^{2t-2r}\sum_{\pi\notin I}|\langle w,\pi v\rangle|^{2r}\leq\beta^{2(t-r)}\alpha_r(w,v)\leq\beta^{2(t-r)}(1+\eta).$$

Therefore,

$$
\begin{aligned}
\alpha_t(u,v) &\geq \sum_{\pi\in I}(|\langle w,\pi v\rangle| - \beta)^{2t} \\
&\geq \sum_{\pi\in I}|\langle w,\pi v\rangle|^{2t}\left(1-\frac{\beta}{|\langle w,\pi v\rangle|^{2t}}\right)^{2t} \\
&\geq \sum_{\pi\in I}|\langle w,\pi v\rangle|^{2t}\left(1-\frac{2\beta t}{|\langle w,\pi v\rangle|^{2t}}\right) \\
&\geq \left(\sum_{\pi\in\Gamma}|\langle w,\pi v\rangle|^{2t}\right) - (1+\eta)\left[\beta^{2(t-r)} - 2\beta t\right] \\
&\geq \alpha_t(w,v) - (1+\eta)\left[\delta^{2(t-r)} - (\delta+\eta)\frac{t}{t-r}\right].
\end{aligned}
$$

Plugging this into (16) and using $\alpha_t(u,v), \alpha_t(v,w) \leq 0.7$ yields,

$$3.4(1+\eta)\left(\delta^{2(t-r)} + (\delta+\eta)\frac{t}{t-r}\right) \leq \delta(4-\delta^2) - 2\eta.$$

Now, since $\lambda(u,w) \neq 1$, we have $\lambda(u,w) \leq 1 - \frac{4}{n}$, and using Lemma 4.7 gives $\delta \geq \frac{2t}{n}$; in particular, $\eta \leq \delta/16$. Combining this with $\delta \leq \frac{1}{16}$, it suffices to prove

$$3.7\left(2\delta^{2(t-r)} + \delta\frac{t}{t-r}\right) \leq 3.8\delta,$$

which certainly holds for some choice of $t = O(r)$. □

A.1 Integrality gaps

We now discuss the consequences of Theorem A.1 for integrality gaps.

Theorem A.2. *Let Γ be any group acting on $[n]$ with $\psi_\Gamma \leq 2^{0.1n}$. If $|\mathcal{P}_r(\frac{1}{(4n)^2})| \geq |Q_n|(1-n^{-2})$, then*

$$\mathsf{SDP}(Q_n/\Gamma) \leq O(r)\,\mathsf{SDP}(Q_n).$$

Proof. Let $\mathcal{P} = \mathcal{P}_r(\frac{1}{(4n)^2})$. Let $C \geq 1$ be such that $K_{22,Cr}$ is a pseudometric on $\mathcal{P}$, according to Theorem A.1. By Lemma 4.1, $K_{22,Cr}$ is negative-definite kernel, i.e. there exists a system of vectors $\{x_u\}_{u\in Q_n}$ such that $\|x_u - x_v\|^2 = K_{22,Cr}(u,v)$.

Fix some arbitrary $u_0 \in \mathcal{P}$. We define a new solution by

$$x'_u = \begin{cases} x_u & u\in\mathcal{P} \\ x_{u_0} & u\notin\mathcal{P}. \end{cases}$$

Certainly $\{x'_u\}_{u\in Q_n}$ is a Γ-invariant vector solution that satisfies the triangle inequalites. We are left to compute the value of this solution.

First, for $(u,v)\in E(Q_n)$ with $u,v\in\mathcal{P}$, by Theorem 4.8 and Lemma 4.5, we have

$$\|x_u - x_v\|^2 = K_{22,Cr}(u,v) \approx \rho_{22,Cr}(u,v) = O(r/n).$$

Hence,

$$\sum_{uv \in E(Q_n)} \|x'_u - x'_v\|^2 \lesssim |E(Q_n)|\frac{r}{n} + 4|E(Q_n \setminus \mathcal{P})| \lesssim |E(Q_n)|\frac{r}{n},$$

using $|Q_n \setminus \mathcal{P}| \leq |Q_n|/n^2$.

On the other hand, using Theorem 4.8 and Lemma 4.6,

$$\sum_{u,v \in Q_n} \|x'_u - x'_v\|^2 \geq \sum_{u,v \in \mathcal{P}} K_{22,Cr}(u,v) \gtrsim \sum_{u,v \in \mathcal{P}} \rho_{22,Cr}(u,v) \gtrsim |\mathcal{P}|^2 \gtrsim |Q_n|^2.$$

This verifies that $\mathsf{SDP}(Q_n/\Gamma) \leq O(r)\,\mathsf{SDP}(Q_n)$. □

Using this, we can recover the best-known integrality gap.

Corollary A.3. *If $\Gamma = \langle \sigma \rangle$ is the group generated by cyclic shifts, then* $\mathsf{SDP}(Q_n/\Gamma) \lesssim \mathsf{SDP}(Q_n)$.

Proof. An argument similar to that of (6) shows that for n large enough and some $r = O(1)$, $|\mathcal{P}_r(\frac{1}{(4n)^2})| \geq |Q_n|(1 - n^{-2})$. □

The problem with averaging over orbits. Of course, one might hope that using techniques more sophisticated than Theorem 2.1, it is possible to find nice groups Γ for which $\Phi(Q_n/\Gamma) \gtrsim f(n)\Phi(Q_n)$, where $f(n) \gg \log n$. In this case, one could hope to derive stronger integrality gaps. Indeed, Bourgain and Kalai [**?**] exhibit primitive permutation groups Γ which yield such bounds. Unfortunately, the following lemma poses a problem.

Lemma A.4. *For any group Γ acting on $[n]$, $\Phi(Q_n/\Gamma) \lesssim \Phi(Q_n)\log(n|\Gamma|)$.*

Proof sketch. Let $k \in \mathbb{N}$ and define $F_k : Q_n \to \{0,1\}$ by $F_k(u) = 1$ if there exist an $i \in [n]$ such that $u_i, u_{i+1}, \ldots, u_{i+k} < 0$. Let

$$S_k = \{u \in Q_n : F_k(v) = 1 \text{ for some } v \in \Gamma u\}. \tag{17}$$

It is clear that S_k is Γ-invariant. Now, there exists a $k \leq \log(n|\Gamma|)$ such that $|S_k|/|Q_n| \in [1/3, 2/3]$, since for a randomly chosen $u \in Q_n$, a fixed sequence will satisfy $u_i, u_{i+1}, \ldots, u_{i+k} < 0$ with probability 2^{-k}, and there are at most $n|\Gamma|$ such sequences under consideration in (17).

By a standard analysis, a randomly chosen $u \in S_k$ will, with high probability, have only $O(k)$ pivotal bits, implying that $|E(u, \bar{S}_k)|$ is typically $O(k) = O(\log(n|\Gamma|))$, which implies that $|E(S_k, \bar{S}_k)| \leq O(\log(n|\Gamma|))|S_k|$, and yields $\Phi(S_k) \leq O(\log(n|\Gamma|))|Q_n|^{-1} \approx O(\log(n|\Gamma|))\Phi(Q_n)$. □

The preceding lemma is problematic, because in order for $|\mathcal{P}_r(1/n^2)|$ to be almost everything, one has to take $r \gtrsim \frac{\log|\Gamma|}{\log n}$, This is because in a sum like

$$\alpha_t(u,v) = \sum_{\pi \in \Gamma} |\langle u, \pi v\rangle|^{2t}$$

the terms not corresponding to $\pi = \mathrm{id}$ can generally only be expected to be of order $(n^{-1/2})^{2t} = n^{-t}$, but there are possibly $|\Gamma|$ of these terms, implying that we need $t \approx \frac{\log|\Gamma|}{\log n}$ in order for these terms to have total magnitude $o(1)$. In the next section, we discuss how different vector solutions can be used with $r = O(1)$ for a specific example with $|\Gamma| \approx 2^{n^{\Omega(1)}}$.

B Larger orbits: Permutations of the rows

In this section, we discuss $m \times n$ sign matrices with $m = \mathrm{poly}(n)$, where Γ includes all permutations of the rows, meaning that our previous SDP solutions would not be adequate (as the orbits are now huge). Still, we give a (weak) SDP solution with $\mathsf{SDP}_{O(1)}(Q_{mn}/\Gamma) \approx \mathsf{SDP}(Q_n)$. Unfortunately, it is not difficult to see that $\Phi(Q_{mn}/\Gamma) \approx \Phi(Q_n)\log n$, meaning that we again achieve only an $\Omega(\log\log N)$ integrality gap. It is possible that a hierarchical version of this construction could give larger gaps.

B.1 The metric

For every $m, n \in \mathbb{N}$, let $X_{m,n} = \left(\frac{1}{\sqrt{n}}\{-1,1\}^n\right)^m \subseteq \mathbb{R}^{mn}$ be the space of sequences $(A_1, A_2, \ldots, A_m)$ with each $A_i \in \{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$. The symmetric group S_m acts in a natural way on $X_{m,n}$: For $\pi \in S_m$, we have $\pi(A) = \pi(A_1, \ldots, A_m) = (A_{\pi(1)}, \ldots, A_{\pi(m)})$. Let $\mathcal{X}_{m,n}$ be the set of orbits of $X_{m,n}$ under the S_m action. We define

$$\lambda_t(A,B) = \frac{1}{m} \max_{\pi:[m]\to[m]} \sum_{i=1}^{m} |\langle A_i, B_{\pi(i)}\rangle|^{2t},$$

where the maximum is over all bijections π.

Lemma B.1. *For any $A, B, C \in X_{m,n}$ and any $t \in \mathbb{N}$, we have*

$$\lambda_t(A,B) \geq \lambda_t(A,C) + \lambda_t(B,C) - 1.$$

Proof. Let $\pi, \pi' : [m] \to [m]$ be such that $\lambda_t(A,C) = \frac{1}{m}\sum_{i=1}^m |\langle A_i, C_{\pi(i)}\rangle|^{2t}$ and $\lambda_t(B,C) = \frac{1}{m}\sum_{i=1}^m |\langle B_i, C_{\pi'(i)}\rangle|^{2t}$. Then letting $\sigma = (\pi')^{-1} \circ \pi$, we have

$$\begin{aligned}
\lambda_t(A,B) &\geq \frac{1}{m}\sum_{i=1}^{m} |\langle A_i, B_{\sigma(i)}\rangle|^{2t} \\
&\geq -1 + \frac{1}{m}\sum_{i=1}^{m} |\langle A_i, C_{\pi(i)}\rangle|^{2t} + \frac{1}{m}\sum_{i=1}^{m} |\langle B_{\sigma(i)}, C_{\pi(i)}\rangle|^{2t} \qquad (18)\\
&= -1 + \lambda_t(A,C) + \lambda_t(B,C).
\end{aligned}$$

□

Next we define, for every $s, t \in \mathbb{N}$, the distance function $\rho_{s,t}(A,B) = 1 - (\frac{1}{2} + \frac{1}{2}\lambda_t(A,B))^s$.

Claim B.2. *For every $s, t \in \mathbb{N}$, $\rho_{s,t}$ is a metric on $\mathcal{X}_{m,n}$.*

Proof. First, it's clear that $\rho_{s,t}(A,B) = \rho_{s,t}(\pi A, B)$ for all $\pi \in S_m$ and $A, B \in X_{m,n}$. Also, $\rho_{s,t}(A,A) = 0$ because $\lambda_t(A,A) = 1$.

Now, consider $A, B, C \in X_{m,n}$. The triangle inequality $\rho_{s,t}(A,B) \leq \rho_{s,t}(A,C) + \rho_{s,t}(B,C)$ reduces to verifying

$$1 + (\tfrac{1}{2} + \tfrac{1}{2}\lambda_t(A,B))^s \geq (\tfrac{1}{2} + \tfrac{1}{2}\lambda_t(A,C))^s + (\tfrac{1}{2} + \tfrac{1}{2}\lambda_t(B,C))^s.$$

Write this as

$$1 + x^s \geq y^s + z^s. \tag{19}$$

Then $x, y, z \in [0,1]$ since $\lambda_t(A,B) \in [0,1]$ for all $A, B \in X_{m,n}$. Combining this with the fact that $1 + x \geq y + z$ from Lemma B.1, we conclude that (19) holds. □

Finally, we analyze the behavior of $\rho_{s,t}$ on "edges" of $X_{m,n}$ and on random pairs. If $A, A' \in X_{m,n}$, we write $A \sim A'$ if $\|A - A'\|_2^2 = \frac{4}{n}$ (i.e. the hamming distance between A and A' is one).

Lemma B.3 (Edges)**.** *If $A, A' \in X_{m,n}$ with $A \sim A'$, then $\rho_{s,t}(A, A') \leq \frac{2st}{mn}$.*

Proof. Observe that

$$\lambda_t(A, A') \geq \frac{1}{m} \sum_{i=1}^{m} |\langle A_i, A'_i \rangle|^{2t} = \frac{1}{m}\left(m - 1 + \left(1 - \frac{2}{n}\right)^{2t}\right) \geq 1 - \frac{4t}{mn}.$$

hence $\rho_{s,t}(A, A') = 1 - (1 - \frac{2t}{mn})^s \leq \frac{2st}{mn}$. □

Lemma B.4 (Random pairs)**.** *Suppose that $A, B \in X_{m,n}$ are chosen independently and uniformly at random. Then*

$$\Pr\left[\lambda_t(A,B) \geq Ln^{-t}\right] \leq 2me^{-\frac{1}{2}L^{1/t}}.$$

In particular, for any $s, t \in \mathbb{N}$, we have $\Pr[\rho_{s,t}(A,B) \geq \frac{1}{4}] \geq \Pr[\lambda_t(A,B) \leq \frac{1}{2}] \geq \frac{1}{2}$.

B.2 An equivalent negative-definite kernel

We now define, for any $t \in \mathbb{N}$, two kernels. For $A, B \in X_{m,n}$, let

$$\alpha_t(A,B) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{m} |\langle A_i, B_j \rangle|^{2t},$$

and

$$\overline{\alpha_t}(A,B) = \frac{\alpha_t(A,B)}{\sqrt{\alpha_t(A,A)\,\alpha_t(B,B)}}.$$

Lemma B.5. *For every $t \in \mathbb{N}$, α_t and $\overline{\alpha_t}$ are both positive semi-definite kernels on $\mathcal{X}_{m,n}$.*

Proof. Define maps $f, \overline{f} : X_{m,n} \to \mathbb{R}^{n^t}$ by $f(A) = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} A_i^{\otimes 2t}$ and $\overline{f}(A) = f(A)/\|f(A)\|_2$. Then $\langle f(A), f(B) \rangle = \alpha_t(A,B)$ and $\langle \overline{f}(A), \overline{f}(B) \rangle = \overline{\alpha_t}(A,B)$. Clearly f is invariant under the S_m action on $X_{m,n}$. □

For every $s, t \in \mathbb{N}$, define a negative-definite kernel on $\mathcal{X}_{m,n}$ by

$$K_{s,t}(A,B) = 1 - \left(\frac{1}{2} + \frac{\overline{\alpha_t}(A,B)}{2}\right)^s.$$

For $r \in \mathbb{N}$, let

$$\mathcal{N}_r(\eta) = \left\{A \in X_{m,n} : |\langle A_i, A_j \rangle|^{2r} \leq \frac{\eta}{m} \quad \forall i \neq j \in [m]\right\}$$

be the set of elements in $X_{m,n}$ with small self-correlation. In particular, $A \in \mathcal{N}_r(\eta)$ implies that $\alpha_r(A,A) \leq 1 + \eta$. Using Cauchy-Schwarz, we have $\alpha_r(A,B) \leq \sqrt{\alpha_r(A,A)\,\alpha_r(B,B)}$, hence $A, B \in \mathcal{N}_r(\eta)$ implies $\alpha_r(A,B) \leq 1 + \eta$ as well.

Lemma B.6 (Heavy matchings)**.** *Suppose that $t \geq 2r$, $\eta \leq \frac{1}{16}$, $\delta \in [0,1]$, and $A, B \in \mathcal{N}_r(\eta)$. Then $\alpha_t(A,B) \geq 1-\delta$ implies that*

$$\lambda_t(A,B) \geq 1 - (10\delta + 2\eta)$$

Proof. Define $\alpha_i = 1 - \sum_{j=1}^m |\langle A_i, B_j\rangle|^{2t}$ and $\beta_i = \max_{j\in[m]} |\langle A_i, B_j\rangle|$. Then,

$$1 - \alpha_i \leq \beta_i^{2t-2r} \sum_{j=1}^m |\langle A_i, B_j\rangle|^{2r} \leq \beta_i^{2(t-r)} \sqrt{\|A_i\|_2 \cdot \alpha_r(B,B)} \leq \beta_i^{2(t-r)}(1+\eta),$$

so we have

$$\beta_i^{2t} \geq \left(\frac{1-\alpha_i}{1+\eta}\right)^{\frac{t}{t-r}} \geq 1 - \frac{t}{t-r}(\alpha_i + \eta) \geq 1 - 2(\alpha_i + \eta). \tag{20}$$

Now suppose that

$$\alpha_t(A,B) = \frac{1}{m}\sum_{i=1}^m (1-\alpha_i) \geq 1 - \delta.$$

Let $S = \{i \in [m] : \alpha_i \leq \frac{1}{8}\}$. Clearly $|S| \geq (1-8\delta)m$ since $\sum_{i=1}^m \alpha_i \leq \delta m$. Define a mapping $\pi : S \to [m]$ by $\pi(i) = \operatorname{argmax}_{j\in[m]} |\langle A_i, B_j\rangle|^{2t}$.

We claim that π is injective. Observe that for $i \in S$, (20) implies that $\beta_i^{2t} \geq 1 - 2\left(\frac{1}{8} + \eta\right) \geq \frac{5}{8}$. So if $\pi(i) = \pi(j)$ for $i \neq j \in S$, then we have

$$|\langle A_i, A_j\rangle|^{2t} \geq |\langle A_i, B_{\pi(i)}\rangle|^{2t} + |\langle A_j, B_{\pi(i)}\rangle|^{2t} - 1 \geq \frac{1}{4},$$

which contradicts the fact that for $A \in \mathcal{N}_r(\eta)$, we have $|\langle A_i, A_j\rangle|^{2t} \leq |\langle A_i, A_j\rangle|^{2r} \leq \frac{\eta}{m} \leq \frac{1}{16}$.

Since π is injective, it follows that

$$\lambda_t(A,B) \geq \frac{1}{m}\sum_{i\in S} \beta_i^{2t} \geq \frac{1}{m}\sum_{i\in S}(1 - 2(\alpha_i + \eta)) \geq \frac{|S|}{m} - 2(\delta + \eta) \geq 1 - (10\delta + 2\eta).$$

□

Even though $K_{s,t}$ may not be a metric, we show that it is always close to $\rho_{s,t}$.

Theorem B.7 (Bi-lipschitz equivalence)**.** *There exists a universal constant $C \geq 1$ such that for any $t \geq 2r$, the distance functions $K_{s,t}$ and $\rho_{s,t}$ are C-bi-lipschitz equivalent when restricted to $\mathcal{N}_r(\frac{1}{20mn})$.*

Proof. If $A = \pi(B)$ for some $\pi \in S_m$, then clearly $\lambda_t(A,B) = \overline{\alpha_t}(A,B) = 1$, hence $\rho_{s,t}(A,B) = K_{s,t}(A,B) = 0$. Let $\eta = \frac{1}{20mn}$.

Consider $A, B \in \mathcal{N}_r(\eta)$ where A and B are in different equivalence classes of $\mathcal{X}_{m,n}$. Then clearly we have

$$\lambda_t(A,B) \leq \frac{1}{m}\left(m - 1 + \left(1 - \frac{2}{n}\right)^{2t}\right) \leq 1 - \frac{2}{mn}. \tag{21}$$

Now suppose that $\overline{\alpha_t}(A,B) = 1-\delta$ for some $\delta \in [0,1]$. In that case, $\alpha_t(A,B) \geq \overline{\alpha_t}(A,B) \geq 1-\delta$, so Lemma B.6 implies that $\lambda_t(A,B) \geq 1 - (10\delta + 2\eta)$. From (21), we conclude that $\delta \geq \frac{1}{6mn}$. This, in turn, implies that $\eta \leq \delta/3$, which gives $\lambda_t(A,B) \geq 1 - 11\delta$.

Finally, we observe that

$$\overline{\alpha_t}(A,B) \geq (1-\eta)\alpha_t(A,B) \geq (1-\delta/3)\alpha_t(A,B) \geq (1-\delta/3)\lambda_t(A,B),$$

hence $\lambda_t(A,B) \leq (1-\delta)(1+\delta/3) \leq 1-\frac{2\delta}{3}$. We conclude that $1-\lambda_t(A,B)$ and $1-\overline{\alpha_t}(A,B)$ are within an $O(1)$ factor of each other for all $A,B \in \mathcal{N}_r(\eta)$. This immediately implies that $K_{s,t}(A,B)$ and $\rho_{s,t}(A,B)$ are within an $O(1)$ multiplicative factor. □

The final result of this section concerns how large one needs to choose r (and hence t) so that $\mathcal{N}_r(\frac{1}{20mn})$ contains most of the points of $X_{m,n}$.

Lemma B.8. *Let $\eta = \frac{1}{20mn}$, and consider $A \in X_{m,n}$ chosen uniformly at random. For any $\tau = \tau(m,n)$, there exists a choice of $r \approx \frac{\log m}{\log n - \log\log\frac{m}{\tau}}$ for which*

$$\Pr\left[A \notin \mathcal{N}_r(\eta)\right] \leq \tau.$$

Proof. Let $A_i, A_j \in \{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$ be chosen independently at random, then

$$\Pr\left[A \notin \mathcal{N}_r(\eta)\right] \leq m^2 \Pr\left[|\langle A_i, A_j\rangle| \geq \left(\frac{1}{20m^2 n}\right)^{1/2r}\right] \leq 2m^2 \exp\left(\frac{-n}{2(20m^2 n)^{1/r}}\right).$$

Simplifying yields the desired conclusion. □

The point is that we can choose any $m = \mathrm{poly}(n)$ and $\tau = 2^{-n^{0.1}}$, and we still only need $r = O(1)$.

C PSD flows and triangle inequalities

In this section, we discuss the question of whether $\mathsf{SDP}_{O(1)}(G) \approx \mathsf{SDP}(G)$ for every graph G, i.e. whether the weak triangle inequalities can always be converted to strong triangle inequalities with only an $O(1)$ loss. This is most nicely stated in the setting of the SDP dual.

Let $G = (V,E)$ be a finite, undirected graph, and for every pair $u,v \in V$, let $\mathcal{P}_{uv}$ be the set of all paths between u and v in G. Let $\mathcal{P} = \bigcup_{u,v\in V} \mathcal{P}_{uv}$. A *flow in* G is a mapping $F : \mathcal{P} \to \mathbb{R}_{\geq 0}$. We define, for every vertex $(u,v) \in E$, the *congestion on* (u,v) as

$$C_F(u,v) = \sum_{p\in\mathcal{P}:(u,v)\in p} F(p).$$

For any $u,v \in V$, let $F[u,v] = \sum_{p\in\mathcal{P}_{uv}} F(p)$ be the amount of flow sent between u and v.

The standard "maximum concurrent flow" problem is simply

$$\mathsf{mcf}(G) = \text{maximize } \left\{D : \forall u,v, F[u,v] \geq D \text{ and } \forall (u,v) \in E, C_F(u,v) \leq 1.\right\}$$

If we define the symmetric matrix

$$A_{u,v} = F[u,v] - D + \mathbf{1}_{\{(u,v)\in E\}} - C_F(u,v),$$

then certainly every feasible flow of value D satisfies $A_{u,v} \geq 0$ for all $u, v \in V$. In fact, we can combine the two types of flow constraints (demand/congestion) together, and get the same thing:

Exercise: $\mathsf{mcf}(G) = \max\{D : A_{u,v} \geq 0\, \forall u, v\}$

Now, the dual of the Sparsest Cut SDP is precisely the same thing, but with a *global* constraint on A, instead of having a constraint per entry:

$$\mathsf{SDP}(G) = \max\{D : L(A) \succeq 0\}.$$

Here, $L(A)$ denotes the Laplacian of A, i.e.

$$L(A)_{i,j} = \begin{cases} \sum_{k \neq i} A_{i,k} & i = j \\ -A_{i,j} & \text{otherwise.} \end{cases}$$

and we write $L(A) \succeq 0$ to denote that $L(A)$ is positive semi-definite.

Now, if we write, for some $\kappa \geq 1$,

$$A^{(\kappa)}_{u,v} = F[u, v] - D + \kappa \cdot \mathbf{1}_{\{(u,v) \in E\}} - C_F(u, v),$$

then clearly

$$\max\{D : A^{(\kappa)}_{u,v} \geq 0\, \forall u, v\} \geq \max\{D : A_{u,v} \geq 0\, \forall u, v\}$$

because we have bumped up the edge capacities. On the other hand, given an $A^{(\kappa)}$-feasible flow of value D, we can always get an actual feasible flow with value D/κ by simply scaling down the flow by factor $1/\kappa$, i.e.

$$\max\{D : A^{(\kappa)}_{u,v} \geq 0\, \forall u, v\} = \kappa \cdot \max\{D : A_{u,v} \geq 0\, \forall u, v\}.$$

Question 1. *Is the same kind of thing true for "PSD-flows"? In other words, are*

$$\max\{D : L(A) \succeq 0\} \qquad \textit{and} \qquad \max\{D : L(A^{(\kappa)}) \succeq 0\}$$

related by a factor depending only on κ?

If this question has a positive answer, then it makes integrality gaps for the Sparsest Cut SDP much easier to understand, because SDP duality shows that $\mathsf{SDP}(G) = \max\{D : L(A) \succeq 0\}$ while $\mathsf{SDP}_\kappa(G) = \max\{D : L(A^{(\kappa)}) \succeq 0\}$.

The answer to this question is affirmative if we can decouple the $L(A) \succeq 0$ constraint into two constraints, i.e. let $X_{u,v} = F[u, v] - D$ for $u \neq v$ and let $Y_{u,v} = \mathbf{1}_{(u,v) \in E} - C_F(u, v)$.

Question 2. *Can we relate (e.g. within constant factors)* $\max\{D : L(A) \succeq 0\}$ *to* $\max\{D : L(X) \succeq 0$ *and* $L(Y) \succeq 0\}$ *as we can for normal flows? It is easy to check that this would give an affirmative answer to Question 1.*