Subexponential Algorithms for d-to-1 Two-Prover Games and for Certifying Almost Perfect Expansion

David Steurer*

November 11, 2010

Abstract

A question raised by the recent subexponential algorithm for Unique Games (Arora, Barak, Steurer, FOCS 2010) is what other "hard-looking" problems admit good approximation algorithms with subexponential complexity.

In this work, we give such an algorithm for d-to-1 two-prover games, a broad class of constraint satisfaction problems. Our algorithm has several consequences for Khot's *d-to-1 Conjectures*. We also give a related subexponential algorithm for certifying that small sets in a graph have almost perfect expansion. Our algorithms follow the basic approach of the algorithms in (Arora, Barak, Steurer, FOCS 2010), but differ in the implementation of the individual steps.

Key ingredients of our algorithms are a local version of Cheeger's inequality that works in the regime of almost perfect expansion, and a graph decomposition algorithm that finds for every graph, a subgraph with at least an ε fraction of the edges such that every component has at most $n^{O(\log(1/\lambda)/\log(1/\varepsilon))^{1/2}}$ eigenvalues larger than λ.

*Microsoft Research New England, Cambridge, MA.

Contents

1 Introduction

One of the main question raised by the recent subexponential algorithm for UNIQUE GAMES [ABS10] is what other "hard-looking" problems admit improved approximations in subexponential time. For example, is there an algorithm for MAX CUT that beats the Goemans–Williamson approximation by $\varepsilon > 0$ and runs in time $\exp(n^{f(\varepsilon)})$ for a function $f(\varepsilon)$ that tends to 0 as $\varepsilon \to 0$? Similarly, is there an algorithm for SPARSEST CUT with approximation ratio C and running time $\exp(n^{g(C)}$ for a function $g(C)$ that tends to 0 as $C \to \infty$?

In this work, we give subexponential approximation algorithms for two other "hard-looking" problems: (1) the problem of approximating the optimal value of d-to-1 two-prover games, and (2) the problem of certifying that all small vertex sets in a graph have almost perfect expansion.

Khot's Conjectures. In his seminal paper [Kho02], Khot proposed two conjectures about the computational complexity of two-prover games (a class of constraint satisfaction problems). The first one, the *Unique Games Conjecture,* turns out to have many very strong implications for the hardness of approximation. A confirmation of the conjecture would essentially settle the approximability of many basic optimization problems, like MAX CUT [KKMO07, MOO05], VERTEX COVER [KR08], and, in fact, every constraint satisfaction problem [Rag08]. The Unique Games Conjecture asserts that for a certain constraint satisfaction problem, called UNIQUE GAMES, it is NP-hard to distinguish between the case that most of the constraints can be satisfied, at least a $1 - \varepsilon$ fraction, and the case that almost none of the constraints can be satisfied, at most an ε fraction. (Here, $\varepsilon > 0$ can be arbitrarily close to 0. See Section 2.2 for a formal definition.)

Khot's second conjecture, the *d-to-1 Conjecture* also has very interesting (but fewer known) implications: the first strong approximation hardness for various graph coloring problems [DMR09, DS10, GS11] and optimal hardness for MAX 3-CSPs with perfect completeness [OW09, Tan09]. The d-to-1 Conjecture asserts that for d-TO-1 GAMES, a more general constraint satisfaction problem than UNIQUE GAMES, it is NP-hard to distinguish between the case that all constraints can be satisfied and the case that almost none of the constraints can be satisfied, at most an ε fraction. (Again, $\varepsilon > 0$ can be arbitrarily close to 0. See Section 2.2 for a formal definition.)

Resolving Khot's conjectures is one of the important open problems in approximation complexity. In contrast to many other conjectures in complexity, there is no consensus among researchers whether the Unique Games Conjecture or the d-to-1 Conjecture are true or not. One of the few concrete evidences for the truth of the conjectures are integrality gaps (which one can view as lower bounds in restricted computational models defined by hierarchies of mathematical programming relaxations). For UNIQUE GAMES, relatively strong integrality gaps are known [KV05, RS09]. For d-TO-1 GAMES, basic integrality gaps (with perfect completeness) were recently shown [GKO+10].

Arora, Barak, and the author [ABS10] showed that UNIQUE GAMES has a subexponential algorithm. This algorithm achieves in time $\exp(n^{O(\varepsilon^{1/3})})$ an approximation that the Unique Games Conjecture asserts to be NP-hard to achieve, that is, it distinguish between the case that a $1 - \varepsilon$ fraction of the constraints can be satisfied and the case that at most an ε fraction of the constraints can be satisfied. This algorithm demonstrates that hardness results based on the Unique Games Conjecture cannot rule out subexponential algorithms. (Certain approximation hardness results based on LABEL COVER (e.g., [Hås01, MR08]) do rule out subexponential algorithms assuming the Exponential Time Hypothesis [IPZ01], namely

3-Sat requires time $2^{\Omega(n)}$.) Another consequence of this algorithm is that if the Unique Games Conjecture is true, then Unique Games is a problem with *intermediate complexity* (strictly between polynomial and strongly exponential complexity).

Building on this work, we show that also the more general d-to-1 Games problem has a subexponential algorithm, achieving an approximation that the d-to-1 Conjecture asserts to be NP-hard to achieve.

Theorem (Informal, see Theorem 2.5)**.** *There exists an algorithm that given a satisfiable d-to-1 game, finds an assignment that satisfies an ε fraction of the constraints in time*

$$\exp\left(n^{O(1/\log(1/\varepsilon))^{1/2}}\right).$$

(Here, the $O(\cdot)$ notation hides factors depending polynomially on d and the alphabet size of the game.)

Like the algorithm for Unique Games, this algorithm demonstrates that hardness results based on the d-to-1 Conjecture (and its variants) do not rule out subexponential algorithms, e.g., opening the possibility of an algorithm that colors a 3-colorable graph with C colors in time $\exp(n^{f(C)})$, where f tends to 0 as $C \to \infty$.

Our algorithm also shows that if the d-to-1 Conjecture (or the Unique Games Conjecture) is true, then d-to-1 Games has intermediate complexity (similar to Unique Games). Furthermore, any reduction from 3-Sat or Label Cover proving the d-to-1 Conjecture must have large polynomial blow-up, at least $n^{\Omega(\log(1/\varepsilon))^{1/2}}$ where ε is the desired soundness (assuming that 3-Sat does not have subexponential algorithms).

The best known polynomial-time algorithm for d-to-1 Games [CMM06] can find an assignment for satisfiable d-to-1 games that satisfies roughly a $1/k^{1-O(1/d^{1/2})}$ fraction of constraints, where k is the alphabet size of the game. For the case, we are interested in the alphabet size of the game could be logarithmic in the instance size. (For significantly larger alphabet size, the best algorithm for d-to-1 Games satisfies a $2^{-O(\log n \cdot \log d)^{1/2}}$ fraction of constraints [Tre05].)

Expansion of Small Sets in Graphs. The (edge) expansion of small sets in graphs is closely connected to the Unique Games problem. Raghavendra and the author [RS10] give a reduction from Small-Set Expansion, the problem of approximating the expansion of small sets in graphs, to Unique Games. This reduction shows that Unique Games is a harder problem than Small-Set Expansion. In particular, the Unique Games Conjecture is true if the following hypothesis, the *Small-Set Expansion Hypothesis*, is true: it is NP-hard to distinguish between the case that there exists a small non-expanding vertex set, with expansion at most ε, and the case that all small vertex sets have almost perfect expansion, expansion at least $1 - \varepsilon$. (Here, $\varepsilon > 0$ can be arbitrarily small. See [RS10] for a formal statement.)

The subexponential algorithm for Unique Games [ABS10] gives further evidence for the close connection between Small-Set Expansion and Unique Games: At the heart of this algorithm lies a subexponential algorithm for approximating Small-Set Expansion.

In this work, we demonstrate a similar connection between d-to-1 Games and Small-Set Expansion in the regime, where the minimum expansion of small sets is close to 1. More concretely, our subexponential algorithm for d-to-1 Games builds on the following subexponential algorithm for certifying that small sets in a graph have almost perfect expansion.

Theorem (Informal, see Theorem 2.2)**.** *There exists an algorithm that given a graph containing a vertex set with volume at most δ and expansion at most $1 - \varepsilon$, finds in time* $\exp(n^{\beta})$ *a vertex set with volume at most δ and expansion at most* $1 - \varepsilon^{O(1/\beta)}$.

It is interesting to note that certifying almost perfect expansion for random graphs is easy. For random graphs, the eigenvalue gap is an efficiently computable certificate that small sets have almost perfect expansion. Concretely, in a graph with (normalized) eigenvalue gap $1 - \varepsilon$, every set with volume at most δ has expansion at least $1 - \delta - \varepsilon$. (A random regular graph with degree $O(1/\varepsilon^2)$ will have eigenvalue gap at least $1 - \varepsilon$ with high probability.) However, in general, graphs in which small sets have close to perfect expansion do not necessarily have eigenvalue gap close to 1. In fact, there are graphs with eigenvalue gap ε in which sets of volume $\delta < 1/2$ have expansion at least $1 - \delta^{O(\varepsilon)}$. (Examples for this are hypercontractive graphs, like Boolean noise graphs.)

The problem of certifying that small sets in a graph have almost perfect expansion is related to the Densest k-Subgraph problem. In this problem, one is given graph and the goal is to find a vertex set of size k with the maximum number of edges staying inside. Achieving approximation ratio C for Densest k-Subgraph is roughly equivalent (at least for regular graphs) to the problem of distinguishing between the case there exists a set with volume k/n and expansion at most $1 - \varepsilon$ and the case that all sets of volume k/n have expansion at least $1 - \varepsilon/C$. The best known polynomial-time algorithm for Densest k-Subgraph has approximation ratio $C = O(n^{1/4})$ [BCC+10]. This approximation guarantee is incomparable to ours. Assuming we choose $\beta = \Omega(1)$, the approximation guarantee of our subexponential algorithm is better when ε is not too small, namely $\varepsilon > 1/n^{o(1)}$. For smaller ε, namely $\varepsilon < 1/n^{\Omega(1)}$, the approximation guarantee of [BCC+10] remains the best known.

1.1 Techniques

Both of our algorithms follow the basic approach of the algorithms in [ABS10]. In this section, we describe this basic approach and outline some of the differences between the algorithms here and in [ABS10]. (Unlike the rest of the paper, the discussion here assumes some familiarity with graph expansion and two-prover games.)

Subexponential Algorithms for Small-Set Expansion. One of the algorithms of [ABS10] has the following approximation guarantee for Small-Set Expansion: Given a graph containing a vertex set S with volume δ and expansion at most ε, the algorithm finds in time $\exp(n^{O(\varepsilon)}/\delta)$ a vertex set with volume at most δ and expansion at most 0.9.

This algorithm is based on the following two facts: (1) Consider the eigenvectors of the graph (or better, its stochastic adjacency matrix) with eigenvalue close to 1, say larger than $1 - O(\varepsilon)$. If S is a vertex set with expansion at most ε, then its indicator vector is close to the subspace spanned by these eigenvectors. Hence, if there are only few such eigenvectors, we can simply find a set close to S by enumerating this subspace (suitably discretized). This enumeration takes time exponential in the dimension of the subspace. (2) On the other hand, [ABS10] show that if there are many eigenvalues close to 1 (more than $n^{O(\varepsilon)}/\delta$), then not all local random walks can mix quickly. Using an appropriate local variant of Cheeger's inequality, one can show that this can only happen if one of the level sets of a local random walk does not expand.

Suppose we are interested in certifying that small sets in a graph have almost perfect expansion. In other words, if the graph contains a small set S with expansion at most $1-\varepsilon$, we want to be able to find a small set with expansion at most $1-\varepsilon'$ for some $\varepsilon' > 0$ (depending on ε). In this situation, one wants to look at all eigenvectors with eigenvalue larger than $\Omega(\varepsilon)$. Similar to before, one can show that the indicator vector of S is $\Omega(\varepsilon)$-correlated with the subspace spanned by these eigenvectors. Given this information, one could find a vertex set S' which agrees with S on an $\Omega(\varepsilon)$ fraction of its vertices. However, this fact alone is not enough to imply that S' has expansion bounded away from 1. A fairly subtle argument (see Section 3.1) shows that even if S is only $\Omega(\varepsilon)$-correlated with the subspace, it is still possible to decode a set with expansion bounded away from 1, namely at most $1-\varepsilon^{O(1)}$. Hence, if there are few eigenvectors with eigenvalue $\Omega(\varepsilon)$, we can solve the problem again by enumerating a suitable discretization of the corresponding subspace.

On the other hand, if there are many significant eigenvalues, say at least n^{β} eigenvalues larger than $\Omega(\varepsilon)$, the arguments of [ABS10] show the following: there exists a vertex i and number $t \in \mathbb{N}$ such that the t-step random walk from i is concentrated on only a small fraction of the vertices. In addition, the collision probability of this random walk does not decrease by much if we take on more random step, concretely the drop in collision probability is at most $C = 1/\varepsilon^{O(1/\beta)}$. We would like to argue that this implies that there exists a vertex set around vertex i with expansion bounded away from 1. Unfortunately, the usual statement of Cheeger's inequality does not apply in this situation, because C is much larger than 1. (In [ABS10], one can achieve that the collision probability stays essentially the *same* when one takes one more random step, which means that C is very close to 1. In this case, the usual form of Cheeger's inequality applies and yields a small vertex set with low expansion, namely close to 0.) We show a tighter version of Cheeger's inequality, which applies also when C is much larger than 1. In our situation, this version of Cheeger's inequality yields a small vertex set with expansion at most $1 - O(1/C^2)$.

Subexponential Algorithms for 2-Prover Games. Like the algorithm for UNIQUE GAMES [ABS10], our algorithm for d-TO-1 GAMES first decomposes the constraint graphs into components with nice eigenvalue distribution and then uses subspace enumeration to find good assignments for the components. Unlike the decomposition for UNIQUE GAMES, our decomposition for d-TO-1 GAMES has to control eigenvalues close to 0 (eigenvalues larger than $1/d^{O(1)}$). The price that we pay for the additional control on eigenvalues close to 0 is that we might have to remove *all* but a small fraction of the edges of the graph. b The subspace enumeration for d-TO-1 GAMES is somewhat more involved than for UNIQUE GAMES. Essentially the algorithm uses some form of list decoding to find good assignments for the components. In this way, one can get an assignment that agrees with an optimal assignment on an $1/d^{O(1)}$ fraction of the vertices. Let's denote this set by T. (However, we have know way of computing this set.) At this point, we know that all constraints that stay within this vertex set T are satisfied by the assignment we found. Unfortunately, it could be the case that the edges that stay inside of T make up much fewer than a $1/d^{O(1)}$ fraction of the constraints of the game (i.e., the set could have close-to-perfect expansion in the constraint graph). To fix this problem, we "unlabel" a constant fraction of randomly chosen vertices and compute new labels for these vertices using a greedy algorithm. One can show that the fraction of constraints satisfied by this new assignment is up to a constant factor at least the fraction of constraints that are *incident* to the vertex set T. (However, it is not guaranteed that it actually satisfies a constant fraction *of* the constraint incident to the set T.) Since T has volume $1/d^{O(1)}$, the fraction of constraints incident to it is at least $1/d^{O(1)}$. It follows that our assignment satisfies a $1/d^{O(1)}$ fraction of constraints.

1.2 Organization

In the next section Section 1.3 (Notation), we introduce some convenient notation for graphs, expansion, and eigenvalues (which is mostly standard). In Section 2 (Main Results), we give detailed proofs of our main results assuming several key theorems. We prove these key theorems in the remaining technical sections Section 3 (Subspace Enumeration), Section 4 (Threshold Rank versus Expansion close to 1), and Section 5 (Graph Decomposition).

1.3 Notation

Graphs and Expansion. Let G be a (possibly weighted) graph with vertex set V. We write $ij \sim G$ to denote a random edge ij of the graph. (If the graph is weighted, we sample an edge with probability proportional to its weight.) We assume that this edge distribution is symmetric, so that a vertex pair ij has the same probability as the vertex pair ji. Abusing notation, we write $i \sim G$ to denote a random vertex of G, distributed like the first endpoint of a random edge of G.

For vertex subsets $S, T \subseteq V$, we let $G(S,T) := \mathbb{P}_{ij\sim G}\{i \in S, j \in T\}$ be the fraction of edges going from S to T. For a vertex i, its *(fractional) degree* $d_i = G(\{i\}, V)$ is the fraction of edges leaving the vertex. The *volume* of a vertex set S is defined as $\mu(S) := \sum_{i\in S} d_i$. Alternatively, $\mu(S) = \mathbb{P}_{i\sim G}\{i \in S\}$. The *expansion* of a vertex set S is given by $\Phi(S) := G(S, V \setminus S)/\mu(S)$. Note that $\Phi(S) = \mathbb{P}_{ij\sim G}\{j \notin S \mid i \in S\}$.

Linear Algebra for Non-Regular Graphs. For the construction of the graph decomposition used in the proof of Theorem 2.5 (Subexponential Algorithm for d-to-1 Games), it is most convenient to allow non-regular graphs. (It's no longer easily possible to avoid dealing with non-regular graphs as it is in the case of Unique Games [ABS10].) In the following, we define appropriate linear algebra notions (matrices, inner products, eigenvalues) for non-regular graphs (which requires a bit more care than for regular graphs).

We let $L_2(V)$ be the space of functions $f \colon V \to \mathbb{R}$ endowed with the inner product $\langle f, g\rangle := \mathbb{E}_{i\sim G} f_i g_i$ (For ease of notation, we write f_i to denote the value of the function f at a vertex i.) For a function $f \in L_2(V)$, we define two norms, $\|f\| := \langle f, f\rangle^{1/2}$ and $\|f\|_1 := \mathbb{E}_{i\sim G}|f_i|$. The graph G naturally corresponds to a *Markov operator* K_G on $L_2(V)$,

$$K_G f(i) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{j\sim G|i} f(j) .$$

Here, $j \sim G \mid i$ denotes a random neighbor of the vertex i in G. (Formally, we sample a random edge of G and condition on the first endpoint being i.) It is straight-forward (but somewhat tedious) to check the following facts about the operator K_G. (The interested reader can find their simple proofs in Section A.1.)

Fact 1.1 (Facts about Markov Operators)**.**

1. *The operator K_G is self-adjoint, so that $\langle f, K_G g\rangle = \langle K_G f, g\rangle$ for all $f, g \in L_2(V)$.*
2. *The operator K_G is contractive in norm $\|\cdot\|$, so that $\|K_G f\| \leqslant \|f\|$ for all $f \in L_2(V)$.*
3. *The operator K_G is contractive in norm $\|\cdot\|_1$, so that $\|K_G f\|_1 \leqslant \|f\|_1$ for all $f \in L_2(V)$.*

Since K_G is self-adjoint, its eigenvalues are real numbers and there exists an orthonormal basis of $L_2(V)$ formed by its eigenfunctions. Since K_G is contractive with respect to $\|\cdot\|$, the eigenvalues of K_G are between -1 and 1. In fact, the all-ones function $f = \mathbb{1}_V$ is an eigenfunction of K_G with eigenvalue 1. When we refer to eigenvalues and eigenfunctions of a graph, we mean the eigenvalues and eigenfunctions of its Markov operator.

Following [ABS10], we define the *threshold rank* and the *soft-threshold rank* of a graph. For $\tau > 0$, the *threshold rank* of G, denoted $\mathrm{rank}_\tau(G)$, is the number of eigenvalues of K_G strictly larger than τ. In some situations, the related notion of *soft-threshold rank* is more convenient. For $\tau > 0$, the *soft-threshold rank* of G, denoted $\mathrm{rank}^*_\tau(G)$, is the infimum of $\sum_i \lambda_i^{2t}/\tau^{2t}$ over all $t \in \mathbb{N}$, where $\lambda_1, \ldots, \lambda_{|V|}$ are the eigenvalues of the operator K_G. The soft-threshold rank is always an upper bound on the threshold rank, so that $\mathrm{rank}_\tau(G) \leqslant \mathrm{rank}^*_\tau(G)$ for all thresholds $\tau > 0$.

2 Main Results

A key ingredient for several parts of our algorithms is the following local variant of Cheeger's bound that works in regime when the expansion is close to 1. (We remark that finding vertex sets with expansion bounded away from 1 is only non-trivial for sufficiently small sets. For example, a random vertex set with volume $1/2$ has typically expansion bounded away from 1, namely $1/2$. On the other hand, a random vertex set of volume δ is expected to have expansion roughly $1 - \delta$.) We prove the following theorem in Section 6.

Theorem 2.1 (Local Cheeger Bound for Expansion close to 1)**.** *Let G be a regular graph with vertex set V and let $f \in L_2(V)$ be a real-valued function on V such that $\|K_G f\|^2 \geqslant \varepsilon \|f\|^2$ and $\|f\|_1^2 \leqslant \delta \|f\|^2$. Then, given the function f, one can compute in polynomial time a vertex set with volume at most δ and expansion at most $1 - \Omega(\varepsilon^2)$.*

The usual Cheeger bound (more precisely, its local variant used in [ABS10]) asserts that if $\|K_G f\|^2 \geqslant (1-\eta)\|f\|^2$ and $\|f\|_1^2 \leqslant \delta\|f\|^2$, then one can efficiently find a vertex set (among the level sets of f^2) with small volume, say at most 2δ, and low expansion, namely at most $O(\sqrt{\eta})$. In addition, one has to assume that G is "lazy", that is, at least half of the degree of every vertex is due to self-loops (formally, $\mathbb{P}_{j\sim G|i}\{j = i\} \geqslant 1/2$ for every vertex i). It is important that Theorem 2.1 can avoid the assumption that G is lazy (the theorem is trivally true for lazy graphs, because in lazy graphs every nonnegative function satisfies $\|K_G f\|^2 \geqslant 1/4\|f\|^2$ and every vertex set has expansion at most $1/2$.) The main feature of Theorem 2.1 is that it applies whenever the ratio $\|K_G f\|^2/\|f\|^2$ is bounded away from 0 (whereas the usual version of Cheeger's inequality applies only if this ratio is sufficiently close to 1).

2.1 Algorithm for Small-Set Expansion close to 1

In this section, we explain the components of the following subexponential algorithm for certifying that all small sets in a graph have almost perfect expansion.

Theorem 2.2 (Subexponential Certificates for Almost Perfect Expansion)**.** *There exists an algorithm that given a graph containing a vertex set with volume at most δ and expansion at most $1 - \varepsilon$, finds in time $\exp(n^\beta/\delta)$ a vertex set with volume at most δ and expansion at most $1 - \varepsilon^{O(1/\beta)}$. (Here, ones assumes ε bounded away from 1, say $\varepsilon < 0.9$, and β not too small, $\beta \gg \log\log n/\log n$).*

The algorithm consists of two parts. If the number of eigenvalues larger than $\varepsilon/2$ is smaller than n^β/δ, we can use the following approximation algorithm based on subspace enumeration in order to find a set of volume δ and expansion at most $1 - \varepsilon^{O(1)}$. We prove of the theorem in Section 3.1.

Theorem 2.3 (Subspace Enumeration for Expansion close to 1)**.** *There exists an algorithm that given a graph containing a vertex with volume at most δ and expansion at most $1 - \varepsilon$, finds a vertex with volume at most δ and expansion at most $1 - \varepsilon^{O(1)}$. The running time of the algorithm is exponential in $\operatorname{rank}_{\varepsilon/2}(G)$.*

On the other hand, if the number of eigenvalues larger than $\varepsilon/2$ is larger than n^{β}/δ, we can use the following algorithm for $\lambda = \varepsilon/2$ (which based on local random walks) to find a vertex set with volume at most δ and expansion at most $1 - \varepsilon^{O(1/\beta)}$. We prove this theorem in Section 4.

Theorem 2.4 (Threshold Rank vs Expansion close to 1)**.** *Let G be a graph with n vertices such that $\operatorname{rank}_{\lambda}(G) > n^{\beta}/\delta$. Then, G contains a vertex set with volume at most δ and expansion at most $1 - \lambda^{O(1/\beta)}$. Furthermore, there exists a polynomial-time algorithm that given G and δ, finds such a vertex set. (Here, we assume that the λ is bounded away from 1, say $\lambda < 0.9$, and β is not too small, $\beta \gg 1/\log n$.)*

Combining the above theorems in the way we described gives the proof of Theorem 2.2.

Proof of Theorem 2.2. Let G be the given graph (which is promised to contain a vertex set of volume at most δ and expansion at most $1 - \varepsilon$.) Our goal is to find a set with volume at most δ and expansion at most $1 - \varepsilon^{O(1/\beta)}$ in time $\exp(n^{\beta}/\delta)$. To this end, we distinguish two cases. The first case is that G has more than n^{β}/δ eigenvalues larger than $\lambda := \varepsilon/2$. In this case, Theorem 2.4 tells us that we can find in polynomial time a vertex set with volume at most δ and expansion at most $1 - \varepsilon^{O(1/\beta)}$. Otherwise, if G has at most n^{β}/δ eigenvalues larger than λ, then the algorithm in Theorem 2.3 allows us to find in time $\exp(n^{\beta}/\delta)$ a vertex set with volume at most δ and expansion at most $1 - \varepsilon^{O(1)}$. ◻

2.2 Algorithm for d-to-1 Games

In this section, we give a subexponential algorithm for d-to-1 Games with the following approximation guarantee. (We parametrized the approximation guarantee slightly differently than for the informal theorem statement in introduction. The two approximation guarantees are equivalent.)

Theorem 2.5 (Subexponential Algorithm for d-to-1 Games)**.** *There exists an algorithm that given a satisfiable d-to-1 game, finds in time $\exp(k^2 d n^{\beta})$ an assignment that satisfies a $(1/d)^{O(1/\beta^2)}$ fraction of the constraints. (Here, one assumes $d \geqslant 2$ and β not too small, $\beta \gg \log\log n/\log n$.)*

(We remark that the algorithm also works for almost satisfiable d-to-1 games (as long as optimal solutions violate less than a $(1/d)^{O(1/\beta^2)}$ fraction of the constraints.)

Before proving the theorem, we first give a formal definition of d-to-1 games: Let $k \in \mathbb{N}$ and $[k] = \{1, \ldots, k\}$. We say a binary relation $P \subseteq [k] \times [k]$ is *d-to-1* if P contains for every label $a \in [k]$, at most d pairs of the form $(a, b) \in P$ and at most d pairs of the form $(b, a) \in P$. (We are abusing the term d-to-1 here. The more common definition would require that for every $a \in [k]$, there is at most one b with $(b, a) \in P$. For us, the more symmetric and more general definition is convenient.) We also add a technical condition[1] (which is probably not necessary), namely that P contains for every $a \in [k]$, *at least* one pair of the form $(a, b) \in P$ and at least one pair of the form $(b, a) \in P$.

A *d-to-*1 *game* Γ with vertex set V and alphabet $[k] = \{1, \ldots, k\}$ is specified by a list of d-to-1 constraints, represented as triples (i, j, P) where $i, j \in V$ and P is a d-to-1 binary relation over $[k]$. An assignment $x \in [k]^V$ satisfies a constraint (i, j, P) if $(x_i, x_j) \in P$. The *value* of an assignment x for a game Γ is the fraction of constraints of Γ satisfied by x. In the d-to-1 Games problem, we are given a d-to-1 game Γ and the goal is to find an assignment x with maximum value. We let $\mathrm{opt}(\Gamma)$ denote the fraction of constraints of Γ satisfied by an optimal assignment. A game Γ is *satisfiable* if it has optimal value $\mathrm{opt}(\Gamma) = 1$.

We will allow the constraints of a game Γ to be weighted (with non-negative coefficients). In this case, it is convenient to represent the game as a *distribution* over constraints. We write $(i, j, P) \sim \Gamma$ to sample a random constraint from a game Γ. We will assume that this distribution is *symmetric*, in the sense that the constraint (i, j, P) has the same probability as the constraint (j, i, P^{-1}), where $P^{-1} = \{(b, a) \mid (a, b) \in P\}$ is the inverse relation of P.

Khot [Kho02] conjectures the following for every $d \in \mathbb{N}$ with $d \geqslant 2$:

d-to-1 Conjecture: For every constant $\varepsilon > 0$, there exists an alphabet size k such that given a d-to-1 game Γ with alphabet size k, it is NP-hard to distinguish between $\mathrm{opt}(\Gamma) = 1$ and $\mathrm{opt}(\Gamma) \leqslant \varepsilon$.

The Unique Games Conjecture is an analog of the d-to-1 Conjecture for $d = 1$. (However, one gives up perfect completeness, i.e., $\mathrm{opt}(\Gamma) = 1$, because such 1-to-1 games are trivial). Due to the lack of perfect completeness, the Unique Games Conjecture is incomparable to any d-to-1 Conjecture as far as we know. One usually refers to 1-to-1 games as *unique games*.

Unique Games Conjecture: For every constant $\varepsilon > 0$, there exists an alphabet size k such that given a unique (1-to-1) game Γ with alphabet size k, it is NP-hard to distinguish between $\mathrm{opt}(\Gamma) \geqslant 1 - \varepsilon$ and $\mathrm{opt}(\Gamma) \leqslant \varepsilon$.

For our algorithm, the *constraint graph* $G = G(\Gamma)$ of a d-to-1 game will be important. This graph specifies which vertices appear in a constraint together. Formally, the edge distribution of G is the marginal distribution of the pair (i, j) for a random constraint $(i, j, P) \sim \Gamma$.

The following two theorems are needed for the proof of Theorem 2.5. The first theorem allows us to partition any graph into components with small soft-threshold rank, smaller than n^β for threshold λ, while preserving at least a $\lambda^{O(1/\beta^2)}$ fraction of the constraints.

Theorem 2.6 (Graph Decomposition)**.** *Let G be a graph with n vertices. (If G is weighted, assume that edges have polynomial weight.) Then for every $\beta, \lambda > 0$, there exists a subgraph G_0 such that every connected component A of G_0 satisfies $\mathrm{rank}^*_\lambda(A) \leqslant n^\beta$ and G_0 contains at least a $\lambda^{O(1/\beta^2)}$ fraction of the edges of G. Furthermore, there exists a polynomial time algorithm that given the graph G and parameters β and λ, finds such a subgraph G_0. (Here, one assumes λ bounded away from 1, say $\lambda < 0.9$, and β not too small, so that $\beta \gg 1/\log n$.)*

[1] The only reason we add this condition is because of Lemma 3.4, which seems to be false if some labels never participate in satisfying configurations for constraints. We remark that the proof of Lemma 3.4 would go through also for d-to-1 two-prover games in the sense of Khot [Kho02], i.e., where we have a bipartite constraint graph with different alphabets for each side. Even in general, it is probably possible to fix Lemma 3.4 by pruning labels that participate in exceptionally few constraints as satisfying configurations.

After we decomposed the constraint graph using the previous theorem, we apply the following algorithm to the individual components of the decomposition. (Since the original instance was promised to be satisfiable, also all components of the decomposition are satisfiable d-to-1 games.)

Theorem 2.7 (Subspace Enumeration for d-to-1 Games)**.** *There exists an algorithm that given a satisfiable d-to-1 game Γ, finds an assignment that satisfies a $1/d^{O(1)}$ fraction of the constraints of Γ. The running time of the algorithm is at most exponential in $k^2 d \cdot \mathrm{rank}^*_{1/d^{O(1)}}(G)$, where $G = G(\Gamma)$ is the constraint graph.*

Combining the above theorems in the way we described proves Theorem 2.5.

Proof of Theorem 2.5. Let Γ be a given satisfiable d-to-1 game with alphabet size k. Let $G = G(\Gamma)$ be the constraint graph of the game Γ. Our goal is to find in time $\exp(dk^2 n^\beta)$ an assignment for Γ that satisfies an $1/d^{O(1/\beta^2)}$ fraction of constraints. Let $\lambda = 1/d^{O(1)}$ and let G_0 be the subgraph of G obtained from Theorem 2.6. Every connected component of G_0 corresponds to subgame of Γ.

Since these subgames are satisfiable and their constraint graphs have soft-threshold rank at most n^β, Theorem 2.7 allows us to find in time $\exp(dk^2 n^\beta)$ assignments for these subgames that satisfy a $1/d^{O(1)}$ fraction of constraints. Since a G_0 contains a $1/d^{O(1/\beta^2)}$ fraction of the constraints of Γ, the combination of these assignments also satisfy a $1/d^{O(1/\beta^2)}$ fraction of constraints of Γ. □

3 Subspace Enumeration

In this section, we develop subspace enumeration algorithms for expansion close to 1 and for d-to-1 games. Both algorithms follow the same scheme. First, one shows that if there exists a good solution than it has to be correlated with a certain subspace (that one can enumerate). Second, one shows how to decode an approximate solution from the projection of a good solution into this subspace.

For us, the main difficulty here is that the projection only guaranteed to be *slightly* correlated with a good solution. (In contrast, for the problems considered in [ABS10], the projection was *close* to a good solution, which made the decoding easy.) As a consequence, our decoding procedures are more involved compared to the previous decoding procedures.

For a graph G with vertex set V, a function $f \in L_2(V)$, and $\lambda > 0$, we let $f^{>\lambda}$ denote the projection of f into the subspace of $L_2(V)$ spanned by the eigenfunctions of G with eigenvalue larger than λ (in absolute[2] value).

The following simple lemma is the common basis of our subspace enumeration algorithms. This lemma appears in very similar (slightly weaker) form in previous works that use subspace enumeration [KT07, Kol10, ABS10]. (We cannot directly reference one of their lemmas only because we are interested in expansion close to 1. But we still omit the proof.)

Lemma 3.1 (Subspace Enumeration)**.** *Let G be a graph with vertex V and let $f = \mathbb{1}_S \in L_2(V)$ be the indicator function of a vertex set $S \subseteq V$ with expansion at most $1 - \varepsilon$. Then,*

$$\langle f, f^{>\lambda} \rangle \geqslant \tfrac{1}{(1-\lambda)}(\varepsilon - \lambda)\|f\|^2$$

[2]It would be enough to consider only the *positive* eigenvalues larger than λ. We choose here to include the negative ones as well in order to be consistent with the definition of threshold rank.

Furthermore, there exists an algorithm that given G computes in time $\operatorname{poly}(n) \cdot O(1/\eta)^{\operatorname{rank}_\lambda(G)}$ *a list of functions containing a function* $\tilde{f}$ *that is* η*-close to* $f^{>\lambda}$ *(that is,* $\|f^{>\lambda} - \tilde{f}\| \leqslant \eta \|f^{>\lambda}\|$*).*

3.1 Subspace Enumeration for Expansion close to 1

In this section, we prove the following theorem.

Theorem (Restatement of Theorem 2.3)**.** *There exists an algorithm that given a graph containing a vertex with volume at most* δ *and expansion at most* $1 - \varepsilon$*, finds a vertex with volume at most* δ *and expansion at most* $1 - \varepsilon^{O(1)}$*. The running time of the algorithm is exponential in* $\operatorname{rank}_{\varepsilon/2}(G)$*.*

Suppose S is the promised vertex set with volume at most δ and expansion at most $1 - \varepsilon$. By Lemma 3.1, we can find in time exponential in $\operatorname{rank}_\lambda(G)$ a function that is close to the projection $f^{>\lambda}$ of the indicator function $f = \mathbb{1}_S$. For simplicity, assume find a function that is exactly equal to the projection $f^{>\lambda}$. (Standard continuity arguments show that a function close enough to the projection will also work.) Notice that if we choose $\lambda = \varepsilon/2$, then Lemma 3.1 tells us $\langle f, f^{>\lambda}\rangle \geqslant \Omega(\varepsilon)\|f\|^2$. Hence, if we combine the following lemma with Theorem 2.1 (Local Cheeger for Expansion close to 1), then we can find a set with volume at most δ and expansion at most $1 - \varepsilon^{O(1)}$ as required for Theorem 2.3.

Lemma 3.2. *Let* $f = \mathbb{1}_S$ *be the indicator of a vertex set* S *with volume* δ*. Suppose* f *is* ε*-correlated with its projection* $g = f^{>\lambda}$*, so that* $\langle f, g\rangle \geqslant \varepsilon\|f\|^2$*. Then, there exists a level set* S' *of the function* g^2 *with volume* $\mu(S') = O(\delta/\varepsilon^2)$ *such that the restriction of* g *to this level set satisfies* $\|Kg_{|S'}\|^2 \geqslant \Omega(\lambda^2\varepsilon^4)\|g_{|S'}\|^2$*.*

Proof. Consider the level set $S' = \{i \in V \mid g_i^2 \geqslant t\}$ of g^2 and define $g' := g_{|S'}$ to be the restriction of g to its level set S'. (We choose the threshold $t = \Omega(\varepsilon^2)$ at the end of the proof.) We can upper bound the volume of the level set S' by

$$\mu(S') \leqslant \mathop{\mathbb{P}}_{i \in V}\left\{g_i^2 > t\right\} \leqslant \|g\|^2/t \leqslant \|f\|^2/t = \delta/t\,.$$

In order to show that g' satisfies $\|Kg'\|^2 \geqslant \Omega(\lambda\varepsilon^4)\|g'\|^2$, we will argue that the restriction $g_{|S}$ and $g' = g_{|S'}$ are correlated (which implies that also g and g' are correlated).

Claim 3.2.1. *The correlation of* g *and* g' *is at least* $\langle g, g'\rangle \geqslant \langle g_{|S}, g_{|S'}\rangle \geqslant \alpha_{\varepsilon,t}\|f\|^2$ *for* $\alpha_{\varepsilon,t} = (\varepsilon/2 - \sqrt{t})^2$*.*

First, we lower bound the norm of the restriction $g_{|S}$, using the fact that g is correlated with the function f, whose support is S,

$$\|g_{|S}\| = \|f_{|S} - (f-g)_{|S}\| \geqslant \|f_{|S}\| - \|f - g\| = \|f\| - \|f - g\| \geqslant (1 - \sqrt{1-\varepsilon})\|f\| \geqslant (\varepsilon/2)\|f\|\,.$$

On the other hand, we can upper bound the norm of the restriction $g_{|S\setminus S'}$ (using the fact that $g^2 \leqslant t$ outside of the set S'),

$$\|g_{S\setminus S'}\|^2 \leqslant t\,\mu(S \setminus S') \leqslant t\|f\|^2\,.$$

Combining these bounds, we can lower bound the correlation of $g_{|S}$ and $g_{|S'}$,

$$\langle g_{|S}, g_{|S'}\rangle = \|g_{|S\cap S'}\|^2 = \|g_{|S} - g_{|S\setminus S'}\|^2 \geqslant (\|g_{|S}\| - \|g_{|S\setminus S'}\|)^2 \geqslant \left(\varepsilon/2 - \sqrt{t}\right)^2\|f\|^2 = \alpha_{\varepsilon,t}\|f\|^2\,,$$

which proves our claim.

The above claim shows that the function g' is $\alpha_{\varepsilon,t}$-correlated with a function g that lives in the span of the eigenfunctions with eigenvalue larger than λ. Hence, we can lower bound the norm of the projection of g' into this space by $\|(g')^{>\lambda}\|^2 \geqslant \alpha_{\varepsilon,t}^2\|g'\|^2$. Thus, $\|Kg'\|^2 = \|K(g')^{>\lambda}\|^2 + \|K(g')^{\leqslant\lambda}\|^2 \geqslant \lambda^2\alpha_{\varepsilon,t}^2\|g'\|^2$

If we choose $t = \varepsilon^2/10$, then $\alpha_{\varepsilon,t} = \Omega(\varepsilon^2)$ and the volume of S' is bounded by $\mu(S') \leqslant \|f\|^2/t = O(\delta/\varepsilon^2)$. Furthermore, the function g' satisfies $\|Kg'\|^2 \geqslant \Omega(\lambda^2\varepsilon^4)\|g'\|^2$. □

3.2 Subspace Enumeration for d-to-1 Games

Let Γ be a d-to-1 game with vertex set $V = \{1, \ldots, n\}$ and alphabet $[k] = \{1, \ldots, k\}$. Let $G = G(\Gamma)$ be the constraint graph of the game. In this section, we prove the following result.

Theorem (Restatement of Theorem 2.7)**.** *There exists an algorithm that given a satisfiable d-to-1 game Γ, finds an assignment that satisfies a $1/d^{O(1)}$ fraction of the constraints of Γ. The running time of the algorithm is at most exponential in $k^2 d \cdot \operatorname{rank}^*_{1/d^{O(1)}}(G)$, where $G = G(\Gamma)$ is the constraint graph.*

As for unique games, one can define a *label-extended graph* $\hat{G} = \hat{G}(\Gamma)$ with vertex set $\hat{V} = V \times [k]$. The edge distribution of $\hat{G}$ is given by the following sampling procedure:

1. Sample a random constraint $(i, j, P) \sim \Gamma$.
2. Sample a random label pair $(a, b) \in P$.
3. Output an edge between the vertices $(i, a) \in \hat{V}$ and $(j, b) \in \hat{V}$.

The following simple lemma shows that a good assignment for the d-to-1 game Γ corresponds to a vertex set in $\hat{G}$ with volume $1/k$ and expansion roughly $1 - 1/d^{O(1)}$. (We omit the proof.)

Lemma 3.3. *Let $x \in [k]^V$ be an assignment for Γ with value at least $1 - \varepsilon$. Then, the vertex set $S = \{(i, a) \mid x_i = a\}$ has volume $1/k$ and expansion $1 - 1/d^{O(1)} + \varepsilon$.*

The following (somewhat tedious) lemma shows that the soft-threshold rank of the label extended graph $\hat{G}$ can be bounded from above in terms of the soft-threshold rank of the constraint graph G and the alphabet size k. We defer the proof to Section A.2.

Lemma 3.4. *For every $\tau > 0$, we have $\operatorname{rank}^*_\tau(\hat{G}) \leqslant k^2 d \cdot \operatorname{rank}^*_\tau(G)$.*

In order to prove Theorem 2.7, let x be an optimal assignment for Γ and let S be the corresponding vertex set of $\hat{G}$. By Lemma 3.3, the set S has volume $1/k$ and expansion at most $1 - 1/d^{O(1)}$. Hence, using Lemma 3.1, we can find in time exponential in $\operatorname{rank}_\lambda(\hat{G}) \leqslant k^2 d \operatorname{rank}^*_\lambda(G)$ for $\lambda = 1/d^{O(1)}$ a function $g \in L_2(V)$ that is close to the projection $f^{>\lambda}$ of the indicator $f = \mathbb{1}_S$. In particular, we arrange that the function g satisfies $\langle g, \mathbb{1}_S\rangle \geqslant d^{-O(1)}\|\mathbb{1}_S\|^2$ and $\|g\|^2 \leqslant \|\mathbb{1}_S\|^2$. Then, using the following algorithm, we can find an approximate assignment for the game Γ. This concludes the proof of Theorem 2.7.

Lemma 3.5. *Let x be an assignment for Γ with value 1 and let S be the vertex set of $\hat{G}$ corresponding to this assignment. Then, given a function $g \in L_2(\hat{V})$ with $\langle g, \mathbb{1}_S\rangle \geqslant \varepsilon\|\mathbb{1}_S\|^2$ and $\|g\|^2 \leqslant \|\mathbb{1}_S\|^2$, we can compute in polynomial-time an assignment x' that satisfies at least an $\varepsilon^{O(1)}$ fraction of the constraints of Γ.*

Proof. For $t = \varepsilon^2/100$, consider the level set $S' := \{(i,a) \mid g_{i,a}^2 > t\}$ of g^2. For every vertex $i \in V$, define a label set $S'_i = \{a \mid (i,a) \in S'\}$. Our goal is to show that based on the sets S'_i, one can decode an assignment (in a randomized way) that satisfies an $\varepsilon^{O(1)}$ fraction of constraints (in expectation).

Claim 3.5.1. *The label set S'_i cannot be too large for many vertices, so that*

$$\mathop{\mathbb{P}}_{i\sim G}\left\{|S'_i| > 1/\varepsilon^5\right\} \leqslant O(\varepsilon^3)\,.$$

We can upper bound the volume of S',

$$\mu(S') = \mathop{\mathbb{P}}_{(i,a)\sim\hat{G}}\left\{g_{i,a}^2 > t\right\} \leqslant \|g\|^2/t \leqslant O(1/\varepsilon^2)\cdot\tfrac{1}{k}\,.$$

Therefore,

$$\mathop{\mathbb{P}}_{i\sim G}\left\{|S'_i| > 1/\varepsilon^5\right\} \leqslant \varepsilon^5 \mathop{\mathbb{E}}_{i\sim G}|S'_i| = \varepsilon^5 k\mu(S') \leqslant O(\varepsilon^3)\,.$$

Claim 3.5.2. *The label set S'_i contains the "correct label", i.e., the label assigned by x, for about an ε^2 fraction of the vertices, so that*

$$\mathop{\mathbb{P}}_{i\sim G}\left\{x_i \in S'_i\right\} = \Omega(\varepsilon^2)\,.$$

Using Cauchy–Schwarz, we can upper bound the correlation of g and $\mathbb{1}_S$ in terms of the volume $\mu(S \cap S')$.

$$\varepsilon/k \leqslant \langle f, \mathbb{1}_S\rangle \leqslant \langle f, \mathbb{1}_{S\cap S'}\rangle + \sqrt{t}\cdot\mu(S) \leqslant \|f\|\cdot\mu(S\cap S')^{1/2} + \sqrt{t}\cdot\mu(S).$$

It follows that $\mu(S \cap S') \geqslant \Omega(\varepsilon^2)\cdot 1/k$. Therefore,

$$\mathop{\mathbb{P}}_{i\sim G}\left\{x_i \in S'_i\right\} = \mathop{\mathbb{E}}_{i\sim G}|S'_i \cap S| = k\cdot\mu(S\cap S') \geqslant \Omega(\varepsilon^2)\,.$$

Consider the following distribution over assignments $y \in [k]^V$. To every vertex, we assign a random label $y_i \in S'_i$. (If S'_i is empty, we assign a random label from $[k]$.) It is straight-forward to verify that the above claims imply that x and y agree on at least an $\varepsilon^{O(1)}$ fraction of constraints in expectation.

Let T be the set of vertices that x and y agree on. (Note that we cannot compute the set T.) At this point, we know that all constraints with both endpoints in T are satisfied. Unfortunately, this fraction of constraints could be much less than the volume of T. (In fact, it could be the case that all edges incident to T leave the set.)

However, there is an easy way to fix the assignment y such that indeed an $\varepsilon^{O(1)}$ fraction of constraints will be satisfied. After sampling the assignment y, we compute an assignment x' by the following algorithm:

1. Sample a random subset of vertices A of volume $1/2$.
2. For every vertex $i \notin A$, keep the label of i and assign $x'_i = y_i$.
3. For every vertex $i \in A$, compute a new label $x'_i \in [k]$ such that the assignment x' satisfies as many constraints of the form (i, j, P) with $j \notin A$ as possible. Formally,

$$x'_i = \operatorname{argmax}_{a\in[k]} \left|\{(i,j,P) \mid j \notin A,\ (a, y_j) \in P\}\right|$$

We claim that the weight of constraints satisfied by the assignment x' is at least the number of constraints between the vertex set A and the vertex set $T \setminus A$. This claim implies the lemma, because the expected volume of T is $\varepsilon^{O(1)}$ and every constraint incident to T goes between A and $T \setminus A$ with constant probability (since A is a random subset of volume $1/2$).

To verify the claim, note that for every vertex $i \in A$ there exists a label (namely x_i) that satisfies all constraints of the form (i, j, P) with $j \in T \setminus A$ (using the fact that y and x' agree with x on T). Hence, when the algorithm chooses in step 3 the label x'_i for vertex i, we are guaranteed to satisfy at least as many constraints as there are between i and $T \setminus A$. In total, the number of satisfied constraints is at least the number of constraints between A and $T \setminus A$, as desired. □

4 Threshold Rank versus Expansion close to 1

The following theorem is a restatement of Theorem 2.4 with the only difference being that we assume a lower bound on the soft-threshold rank instead of the threshold rank. However, since the soft-threshold rank is always larger than the threshold-rank, this theorem directly implies Theorem 2.4.

Theorem 4.1. *Let G be a graph with n vertices such that $\operatorname{rank}^*_\lambda(G) > n^\beta/\delta$. Then, G contains a vertex set with volume at most δ and expansion at most $1 - \lambda^{O(1/\beta)}$. Furthermore, there exists a polynomial-time algorithm that given G and δ, finds such a vertex set. (Here, we assume that the λ is bounded away from 1, say $\lambda < 0.9$, and β is not too small, $\beta \gg 1/\log n$.)*

Proof. Let $V = \{1, \ldots, n\}$ be the vertex set of G. Consider the orthonormal basis $f^{(1)}, \ldots, f^{(n)}$ of $L_2(V)$ given by the scaled indicators $f^{(i)} = \sqrt{1/d_i}\,\mathbb{1}_{\{i\}}$. Note that $\|f^{(i)}\|_1^2 = d_i$ and thus, $\sum_i \|f^{(i)}\|_1^2 = 1$. Let $K = K_G$ be the Markov operator of G and consider the functions $f^{(i,t)} = K^t f^{(i)}$ for $t \in \mathbb{N}$.

Claim. *For $R = (\beta \log n)/(2\log(1/\lambda))$, there exists a vertex i_0 with*

$$\|f^{(i_0,R)}\|^2 \geqslant (\|f^{(i_0)}\|_1^2 + 1/n)/2\delta\,.$$

The assumed lower bound on the soft-threshold rank of G implies that all $t \in \mathbb{N}$ (in particular, $t = R$),

$$\sum_i \|f^{(i,t)}\|^2 = \sum_i \langle f^{(i)}, K^{2t} f^{(i)} \rangle = \operatorname{Tr} K^{2t} > \lambda^{2t} n^\beta/\delta\,.$$

For $t = R$, the right-hand side evaluates to $1/\delta$. Hence, by averaging, there exists a vertex i_0 such that $\|f^{(i_0,R)}\|^2 \geqslant (1/n + \|f^{(i_0)}\|_1^2)/2\delta$. (Here, we are using that $\sum_i \|f^{(i)}\|_1^2 = 1$.)

Claim. *There exists $t \in \{0, \ldots, R-1\}$ such that $\|K f^{(i_0,t)}\|^2 \geqslant \lambda^{2/\beta} \|f^{(i_0,t)}\|^2$.*

By the previous claim, $\|f^{(i_0,R)}\|^2 \geqslant (1/n)\|f^{(i_0,1)}\|^2$ (assuming $\delta \leqslant 1/2$). Hence, by averaging, there exists $t \in \{0, \ldots, R-1\}$ such that $\|f^{(i_0,t+1)}\|^2/\|f^{(i_0,t)}\|^2 \geqslant (1/n)^{1/R} = \lambda^{2/\beta}$. The claim follows, because $f^{(i_0,t+1)} = K f^{(i_0,t)}$.

The proof of the theorem follows by applying Theorem 2.1 (Local Cheeger Bound) to the function $f = f^{(i_0,t)}$. On the one hand, by the second claim, this function satisfies $\|Kf\|^2 \geqslant \lambda^{2/\beta}\|f\|^2$. On the other hand, by the first claim, the function satisfies $\|f\|^2 \geqslant \|f^{(i_0,R)}\|^2 \geqslant \|f^{(i_0)}\|_1^2/2\delta \geqslant \|f\|_1^2/2\delta$. (Here, we are using that the operator K is contractive with respect to the norms $\|\cdot\|$ and $\|\cdot\|_1$.) Hence, Theorem 2.1 allows us to find a vertex set with volume 2δ and expansion at most $1 - O(\lambda^{4/\beta})$. (We omitted this factor 2 for the volume in theorem statement, because it can be absorbed into the $O(\cdot)$-notation.) □

5 Graph Decomposition for Eigenvalues close to 0

Theorem (Restatement of Theorem 2.6)**.** *Let G be a graph with n vertices. (If G is weighted, assume that edges have polynomial weight.) Then for every $\beta, \lambda > 0$, there exists a subgraph G_0 such that every connected component A of G_0 satisfies $\operatorname{rank}^*_\lambda(A) \leqslant n^\beta$ and G_0 contains at least a $\lambda^{O(1/\beta^2)}$ fraction of the edges of G. Furthermore, there exists a polynomial time algorithm that given the graph G and parameters β and λ, finds such a subgraph G_0. (Here, one assumes λ bounded away from 1, say $\lambda < 0.9$, and β not too small, so that $\beta \gg 1/\log n$.)*

Proof. The decomposition procedure is very similar to the procedure for the graph decomposition in [ABS10]:

> As long as there exists a connected component A that violates the upper bound on the soft-threshold rank, so that $\operatorname{rank}^*_\lambda(A) > n^\beta$, use Theorem 4.1 to find a set vertex $S \subseteq V$ with $\mu(S) \leqslant n^{-\Omega(\beta)}\mu(A)$ and $\Phi(S) \leqslant 1 - \lambda^{O(1/\beta)}$, and then subdivide the component A into two parts S and $A \setminus S$ by removing the edges between them.

Note that the graph changes in the course of the decomposition procedure. Therefore, also the notion of volume and expansion change. To lower bound the fraction of edges that remain when the procedure terminates, we introduce *fictitious edge weights*. (We don't use these weights for the decomposition, we only use them to analyze the procedure.) For simplicity, we assume that all edges in the graph have the same actual weight. Initially, we let the fictitious weights be equal to the actual edge weights. Whenever we subdivide a connected component A using a subset $S \subseteq A$, we distribute the fictitious weight of the removed edges equally among the edges that stay in S. (Notice that this charging scheme maintains the invariant that all edges in the same connected component have the same fictitious weight.) Since the expansion of S is at most $1 - \lambda^{O(1/\beta)}$, at least an $\lambda^{O(1/\beta)}$ fraction of its edges stay inside. Hence, the fictitious weight of the edges in S increases by at most a factor of $1/\lambda^{O(1/\beta)}$. Note that the total fictitious weight in the graph remains the same during the decomposition procedure. Hence, if we can upper bound the maximum fictitious weight at the end of the procedure, we can lower bound the number of edges that remain in the graph. (The fictitious weight of an edge indicates how often we are "overcounting" the edge.)

To bound the maximum fictitious weight of an edge, consider any vertex $i \in V$. How often can we subdivide the component of this vertex? We claim that its component can be subdivided at most $O(1/\beta)$ times. The reason is that with every subdivision the volume of the component of vertex i shrinks by a factor of at least $n^{\Omega(\beta)}$. Hence, after $O(1/\beta)$ subdivisions, the fraction of edge of G that are contained in the component is less than $1/n^{O(1)}$, which means that the component consists just of the vertex i. (Recall that we assumed that edges have polynomial weight.) It follows that we subdivide a component of a vertex at most $O(1/\beta)$ times and hence the maximum fictitious edge weight at the end of the decomposition procedure is bounded by $1/\lambda^{O(1/\beta^2)}$, which means that at least a $\lambda^{O(1/\beta^2)}$ fraction of the original edges remain in the graph. ◻

6 Local Cheeger Bound for Expansion close to 1

In this section, we prove the following local version of Cheeger's bound that works for expansion close 1. An important difference to other versions of Cheeger's bound is that we consider the quadratic form $\langle f, K^2 f\rangle = \|Kf\|^2$ instead of the form $\langle f, Kf\rangle$. In [ABS10], this issue was resolved by considering lazy graphs. Since we are interested in expansion close to 1, passing to lazy graphs is prohibitive.

Theorem (Restatement of Theorem 2.1)**.** *Let G be a regular graph with vertex set V and let $f \in L_2(V)$ be a real-valued function on V such that $\|K_G f\|^2 \geqslant \varepsilon \|f\|^2$ and $\|f\|_1^2 \leqslant \delta \|f\|^2$. Then, given the function f, one can compute in polynomial time a vertex set with volume at most δ and expansion at most $1 - \Omega(\varepsilon^2)$.*

The proof of this theorem relies on the following lemma. The proof of this lemma is essentially the same as the proof of the local variant of Cheeger's bound used in [ABS10] (which appeared in several previous works, for example, [DI98, GMT06, RST10].) The only difference is that one step has to be done more carefully[3] (when applying Cauchy–Schwarz). We give a self-contained proof at the end of this section.

Lemma 6.1 (Local Cheeger Bound)**.** *For every function $f \in L_2(V)$, there exists a level set $S \subseteq V$ of the function f^2 with volume $\mu(S) \leqslant \delta$ and expansion*

$$\Phi(S) \leqslant \frac{\sqrt{1 - \langle f, K_G f\rangle^2 / \|f\|^4}}{1 - \|f\|_1^2 / \delta \|f\|^2}\,.$$

Assuming the above lemma, we can prove Theorem 2.1.

Proof of Theorem 2.1. Without loss of generality, we may assume that f is nonnegative. Consider the function $g = Kf + f$. This function satisfies $\|g\|_1^2 \leqslant (\|Kf\|_1 + \|f\|)^2 \leqslant 4\|f\|_1^2$. Similarly, $\|g\|^2 \leqslant 4\|f\|^2$. On the other hand, $g \geqslant f$ and thus $\|g\|^2 \geqslant \|f\|^2$. (Here, we used that both f and g are nonnegative.) Furthermore, $\langle g, Kg\rangle = \langle f, Kf\rangle + 2\|Kf\|^2 + \langle f, K^3 f\rangle \geqslant 2\|Kf\|^2$ (again using nonnegativity). It follows that $\|g\|_1^2 \leqslant O(\delta)\|g\|^2$ and $\langle g, Kg\rangle \geqslant \Omega(\varepsilon)\|g\|^2$. Hence, by Lemma 6.1 one of the level sets of g^2 satisfies the requirements of the theorem. □

6.1 Proof of Lemma 6.1

Lemma (Restatement of Lemma 6.1)**.** *For every function $f \in L_2(V)$, there exists a level set $S \subseteq V$ of the function f^2 with volume $\mu(S) \leqslant \delta$ and expansion*

$$\Phi(S) \leqslant \frac{\sqrt{1 - \langle f, K_G f\rangle^2 / \|f\|^4}}{1 - \|f\|_1^2 / \delta \|f\|^2}\,.$$

Proof. Let $f \in L_2(V)$. Suppose $f^2 \leqslant 1$. Consider the following distribution over vertex subsets $S \subseteq V$:

1. Sample $t \in [0, 1]$ uniformly at random.
2. Output the set $S = \{i \in V \mid f_i^2 > t\}$.

Note that every set S in the support of this distribution is a level set of the function f^2. In the following lemmas, we establish simple properties of this distribution.

Claim 6.1.1. *The expected volume of S satisfies $\mathbb{E}_S\, \mu(S) = \|f\|^2$.*

[3] A similar improvement for the classical (*non-local*) version of Cheeger's inequality was noted in [RS07, Appendix B].

We calculate the expected volume as follows

$$\mathop{\mathbb{E}}_{S} \mu(S) = \mathop{\mathbb{E}}_{i\sim\mu} \mathop{\mathbb{P}}_{t\in[0,1]} \left\{ f_i^2 > t \right\} = \|f\|^2 . \qquad \square$$

Claim 6.1.2. *The second moment of* $\mu(S)$ *is at most* $\mathbb{E}_S\, \mu(S)^2 \leqslant \|f\|_1^2$.

We bound the expectation of $\mu(S)^2$ as follows

$$\mathop{\mathbb{E}}_{S} \mu(S)^2 = \mathop{\mathbb{E}}_{i,j\sim\mu} \mathop{\mathbb{P}}_{t} \left\{ \min\{f_i^2, f_j^2\} > t \right\} = \mathop{\mathbb{E}}_{i,j\sim\mu} \min\{f_i^2, f_j^2\} \leqslant \mathop{\mathbb{E}}_{i,j\sim\mu} f_i f_j = \|f\|_1^2 .$$

Claim 6.1.3. *Sets with volume larger than* δ *contribute to the expected volume at most* $\mathbb{E}_S\, \mu(S)\mathbb{1}_{\mu(S)>\delta} \leqslant \mathbb{E}_S\, \mu(S)^2/\delta$.

Immediate because $\mu(S)\mathbb{1}_{\mu(S)>\delta} \leqslant \mu(S)^2/\delta$ holds pointwise.

Claim 6.1.4. *The expected boundary of* S *is bounded by*

$$\mathop{\mathbb{E}}_{S} G(S, V \setminus S) \leqslant \|f\|^2 \sqrt{1 - \langle f, K_G f\rangle^2/\|f\|^4}.$$

We calculate the expected boundary of S and apply Cauchy–Schwarz,

$$\begin{aligned}
\mathop{\mathbb{E}}_{S} G(S, V \setminus S) &= \mathop{\mathbb{E}}_{ij\sim G} \mathop{\mathbb{P}}_{S} \{ i \in S \wedge j \notin S \} = \mathop{\mathbb{E}}_{ij\sim G} \mathop{\mathbb{P}}_{t} \left\{ f_i^2 > t \geqslant f_j^2 \right\} \\
&= \mathop{\mathbb{E}}_{ij\sim G} \max\left\{ f_i^2 - f_j^2, 0 \right\} = \tfrac{1}{2} \mathop{\mathbb{E}}_{ij\sim G} \left| f_i^2 - f_j^2 \right| = \tfrac{1}{2} \mathop{\mathbb{E}}_{ij\sim G} \left| f_i - f_j \right| \cdot \left| f_i + f_j \right| \\
&\leqslant \Big(\mathop{\mathbb{E}}_{ij\sim G} \tfrac{1}{2}(f_i - f_j)^2 \cdot \mathop{\mathbb{E}}_{ij\sim G} \tfrac{1}{2}(f_i + f_j)^2 \Big)^{1/2} \qquad \text{(using Cauchy–Schwarz)} \\
&= \langle f, (I - K_G) f\rangle^{1/2} \langle f, (I + K_G) f\rangle^{1/2} = \sqrt{\|f\|^4 - \langle f, K_G f\rangle^2} .
\end{aligned}$$

We combine the previous subclaims to complete the proof of Lemma 6.1. Let S^* be the level set of f^2 with volume at most δ and minimum expansion. Then,

$$\begin{aligned}
\Phi(S^*) &\leqslant \frac{\mathbb{E}_S\, G(S, V \setminus S)\mathbb{1}_{\mu(S)\leqslant\delta}}{\mathbb{E}_S\, \mu(S)\mathbb{1}_{\mu(S)\leqslant\delta}} \\
&\leqslant \frac{\mathbb{E}_S\, G(S, V \setminus S)}{\mathbb{E}_S\, \mu(S) - \mathbb{E}_S\, \mu(S)^2/\delta} \qquad \text{(using 6.1.3)} \\
&\leqslant \frac{\|f\|^2 \sqrt{1 - \langle f, K_G f\rangle^2/\|f\|^4}}{\|f\|^2 - \|f\|_1^2/\delta} \qquad \text{(using 6.1.1, 6.1.2, and 6.1.4)}.
\end{aligned}$$

Therefore, the set S^* satisfies the conclusion of the local Cheeger bound (Lemma 6.1).

7 Conclusion

An interesting question raised by this work is how closely graph expansion and d-to-1 games are related. A concrete question is whether Khot's d-to-1 Conjecture is implied by a conjecture about approximating graph expansion (in the same way as the Unique Games Conjecture is implied by a conjecture about the approximability of graph expansion [RS10].)

Acknowledgments

The author thanks Sanjeev Arora, Boaz Barak, and Prasad Raghavendra for valuable discussions.

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer, *Subexponential algorithms for unique games and related problems*, FOCS, 2010, pp. 563–572. 1, 2, 3, 4, 5, 6, 9, 14, 15

[AKK+08] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K. Vishnoi, *Unique games on expanding constraint graphs are easy*, STOC, 2008, pp. 21–28. 18

[BCC+10] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan, *Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph*, STOC, 2010, pp. 201–210. 3

[CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev, *Near-optimal algorithms for unique games*, STOC, 2006, pp. 205–214. 2

[DI98] Tassos Dimitriou and Russell Impagliazzo, *Go with the winners for graph bisection*, SODA, 1998, pp. 510–520. 15

[DMR09] Irit Dinur, Elchanan Mossel, and Oded Regev, *Conditional hardness for approximate coloring*, SIAM J. Comput. **39** (2009), no. 3, 843–873. 1

[DS10] Irit Dinur and Igor Shinkar, *On the conditional hardness of coloring a 4-colorable graph with super-constant number of colors*, APPROX-RANDOM, 2010, pp. 138–151. 1

[GKO+10] Venkatesan Guruswami, Subhash Khot, Ryan O'Donnell, Preyas Popat, Madhur Tulsiani, and Yi Wu, *SDP gaps for 2-to-1 and other label-cover variants*, ICALP (1), 2010, pp. 617–628. 1

[GMT06] Sharad Goel, Ravi Montenegro, and Prasad Tetali, *Mixing time bounds via the spectral profile*, Electron. J. Probab. **11** (2006), no. 1, 1–26 (electronic). MR MR2199053 (2007e:60075) 15

[GS11] Venkatesan Guruswami and Ali Kemal Sinop, *The complexity of finding independent sets in bounded degree (hyper)graphs of low chromatic number*, SODA, 2011, To appear. 1

[Hås01] Johan Håstad, *Some optimal inapproximability results*, J. ACM **48** (2001), no. 4, 798–859. 1

[IPZ01] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane, *Which problems have strongly exponential complexity?*, J. Comput. Syst. Sci. **63** (2001), no. 4, 512–530. 1

[Kho02] Subhash Khot, *On the power of unique 2-prover 1-round games*, STOC, 2002, pp. 767–775. 1, 8

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell, *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?*, SIAM J. Comput. **37** (2007), no. 1, 319–357. 1

[Kol10] Alexandra Kolla, *Spectral algorithms for unique games*, IEEE Conference on Computational Complexity, 2010, pp. 122–130. 9

[KR08] Subhash Khot and Oded Regev, *Vertex cover might be hard to approximate to within* $2-\varepsilon$, J. Comput. Syst. Sci. **74** (2008), no. 3, 335–349. 1

[KT07] Alexandra Kolla and Madhur Tulsiani, *Playing random and expanding unique games*, Unpublished manuscript available from the authors' webpages, to appear in journal version of [AKK+08], 2007. 9

[KV05] Subhash Khot and Nisheeth K. Vishnoi, *The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into* ℓ_1, FOCS, 2005, pp. 53–62. 1

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences invariance and optimality*, FOCS, 2005, pp. 21–30. 1

[MR08] Dana Moshkovitz and Ran Raz, *Two query PCP with sub-constant error*, FOCS, 2008, pp. 314–323. 1

[OW09] Ryan O'Donnell and Yi Wu, *Conditional hardness for satisfiable 3-CSPs*, STOC, 2009, pp. 493–502. 1

[Rag08] Prasad Raghavendra, *Optimal algorithms and inapproximability results for every CSP?*, STOC, 2008, pp. 245–254. 1

[RS07] Jaikumar Radhakrishnan and Madhu Sudan, *On Dinur's proof of the PCP theorem*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 1, 19–61 (electronic). MR 2265009 (2008f:68036) 15

[RS09] Prasad Raghavendra and David Steurer, *Integrality gaps for strong SDP relaxations of unique games*, FOCS, 2009, pp. 575–585. 1

[RS10] ———, *Graph expansion and the unique games conjecture*, STOC, 2010, pp. 755–764. 2, 16

[RST10] Prasad Raghavendra, David Steurer, and Prasad Tetali, *Approximations for the isoperimetric and spectral profile of graphs and related parameters*, STOC, 2010, pp. 631–640. 15

[Tan09] Linqing Tang, *Conditional hardness of approximating satisfiable Max 3CSP-q*, ISAAC, 2009, pp. 923–932. 1

[Tre05] Luca Trevisan, *Approximation algorithms for unique games*, FOCS, 2005, pp. 197–205. 2

A Deferred Proofs

A.1 Facts about Markov Operators

Fact (Restatement of Fact 1.1)**.** *1. The operator K_G is self-adjoint, so that $\langle f, K_G g\rangle = \langle K_G f, g\rangle$ for all $f, g \in L_2(V)$.*

2. *The operator K_G is contractive in norm $\|\cdot\|$, so that $\|K_G f\| \leqslant \|f\|$ for all $f \in L_2(V)$.*

3. *The operator K_G is contractive in norm $\|\cdot\|_1$, so that $\|K_G f\|_1 \leqslant \|f\|_1$ for all $f \in L_2(V)$.*

Proof. Item 1: From the definition of the Markov operator K_G and the inner product on $L_2(V)$,

$$\begin{aligned}\langle f, K_G g\rangle &= \mathop{\mathbb{E}}_{i\sim G} f_i (K_G g)_i \\ &= \mathop{\mathbb{E}}_{i\sim G} \mathop{\mathbb{E}}_{j\sim G|i} f_i g_j \\ &= \mathop{\mathbb{E}}_{j\sim G} \mathop{\mathbb{E}}_{i\sim G|j} f_i g_j \qquad \text{(since } G \text{ is undirected)} \\ &= \langle K_G f, g\rangle\end{aligned}$$

Item 2: From the definition of K_G and the norm $\|\cdot\|$ on $L_2(V)$,

$$\begin{aligned}\|K_G f\|^2 &= \mathop{\mathbb{E}}_{i\sim G} \Big(\mathop{\mathbb{E}}_{j\sim G|i} f_j\Big)^2 \\ &\leqslant \mathop{\mathbb{E}}_{i\sim G} \mathop{\mathbb{E}}_{j\sim G|i} f_j^2 \qquad \text{(by Cauchy–Schwarz)} \\ &= \mathop{\mathbb{E}}_{j\sim G} f_j^2 \qquad \text{(since } G \text{ is undirected)} \\ &= \|f\|^2 .\end{aligned}$$

*Item 3:*From the definition of K_G and the norm $\|\cdot\|_1$ on $L_2(V)$,

$$\begin{aligned}\|K_G f\|_1 &= \mathop{\mathbb{E}}_{i\sim G} \Big|\mathop{\mathbb{E}}_{j\sim G|i} f_j\Big| \\ &\leqslant \mathop{\mathbb{E}}_{i\sim G} \mathop{\mathbb{E}}_{j\sim G|i} |f_j| \qquad \text{(triangle inequality)} \\ &= \mathop{\mathbb{E}}_{j\sim G} |f_j| \qquad \text{(since } G \text{ is undirected)} \\ &= \|f\|_1 .\end{aligned}$$

□

A.2 Label-Extended Graph vs Constraint Graph

Let Γ be a d-to-1 game with vertex set V and alphabet $[k]$. Let $G = G(\Gamma)$ be its constraint graph and $\hat{G} = \hat{G}(\Gamma)$ be its label-extended graph.

Lemma (Restatement of Lemma 3.4)**.** *For every $\tau > 0$, we have* $\mathrm{rank}^*_\tau(\hat{G}) \leqslant k^2 d \cdot \mathrm{rank}^*_\tau(G)$.

Proof. Let $\hat{K} = K_{\hat{G}}$ be the Markov operator of the label-extended graph $\hat{G}$ and let $K = K_G$ be the Markov operator of the constraint graph G.

By the definition of rank^*_τ, it is enough to show that $\operatorname{Tr} \hat{K}^{2t} \leqslant k \cdot \operatorname{Tr} K^{2t}$ for all $t \in \mathbb{N}$. Let $\{f^{(i,a)}\}_{i\in V, a\in[k]}$ be an orthonormal basis of $L_2(\hat{V})$. (We will fix a specific basis later.) Then,

$$\operatorname{Tr} \hat{K}^{2t} = \sum_{i,a} \|\hat{K}^t f^{(i,a)}\|^2 = \sum_{i,a} \mathop{\mathbb{E}}_{j,b} \left(\mathop{\mathbb{E}}_{(\ell,c)\sim \hat{G}^t|(j,b)} f^{(i,a)}_{\ell,c} \right)^2$$

Next, we choose an appropriate basis. Let d_i be the fractional degree of i in the constraint graph, and let $d_{i,a}$ be the fractional degree of (i,a) in the label extended graph. Note that $d_i = \sum_a d_{i,a}$ for every $i \in V$. Consider the orthonormal basis $f^{(i,a)} = \sqrt{1/d_{i,a}} \mathbb{1}_{(i,a)}$ for $L_2(\hat{V})$ and the related basis $f^{(i)} = \sqrt{1/d_i}$ for $L_2(V)$. Then,

$$\begin{aligned} \|\hat{K}^t f^{(i,a)}\|^2 &= \mathop{\mathbb{E}}_{j,b} \left(\mathop{\mathbb{P}}_{(\ell,c)\sim \hat{G}^t|(j,b)} \{(\ell,c) = (i,a)\} \right)^2 / d_{i,a} \\ &\leqslant \mathop{\mathbb{E}}_{j} \left(\mathop{\mathbb{P}}_{\ell \sim G^t|j} \{\ell = i\} \right)^2 / d_{i,a} = \|K^t f^{(i)}\|^2 \cdot \tfrac{d_i}{d_{i,a}} \,. \end{aligned}$$

By our assumption on the structure of the constraint (that every label a participates in at least one satisfying configuration per constraint (i, j, P) of Γ), we can lower bound $d_{i,a} \geqslant d_i/(kd)$ for all $i \in V$ and $a \in [k]$. Hence, for $C \geqslant 1$, we can relate the upper bound the trace of $\hat{K}^{2t}$ by

$$\operatorname{Tr} \hat{K}^{2t} = \sum_{i,a} \|\hat{K}^t f^{(i,a)}\|^2 \leqslant k^2 d \sum_i \|K^t f^{(i)}\|^2 = k^2 d \operatorname{Tr} K^{2t} \,. \qquad \square$$