Conditional Hardness for Approximate Coloring*

Irit Dinur† Elchanan Mossel‡ Oded Regev§

September 12, 2008

Abstract

We study the ApproxColoring(q, Q) problem: Given a graph G, decide whether $\chi(G) \leq q$ or $\chi(G) \geq Q$. We present hardness results for this problem for any constants $3 \leq q < Q$. For $q \geq 4$, our result is based on Khot's 2-to-1 label cover, which is conjectured to be NP-hard [Khot02]. For $q = 3$, we base our hardness result on a certain '$\rhd\!\!<$ shaped' variant of his conjecture. Previously no hardness result was known for $q = 3$ and $Q \geq 6$. At the heart of our proof are tight bounds on generalized noise-stability quantities, which extend the recent work of Mossel et al. [MOO05] and should have wider applicability.

1 Introduction

The approximate graph coloring problem, which we describe next, is one of a few classical optimization problems whose approximability behavior is still quite mysterious, despite increasingly complex techniques developed in the past 15 years. For an undirected graph $G = (V, E)$, let $\chi(G)$ be its chromatic number, i.e., the smallest number of colors needed to color the vertices of G without monochromatic edges. Then the approximate graph coloring problem is defined as follows.

ApproxColoring(q, Q) : Given a graph G, decide between $\chi(G) \leq q$ and $\chi(G) \geq Q$.

This problem also has a natural search variant, which can be stated as follows: given a graph G with $\chi(G) \leq q$, color G with less than Q colors. It is easy to see that the search variant is not easier than the original decision variant, and hence for the purpose of showing hardness results it is enough to consider the decision variant.

It is easy to solve the problem ApproxColoring$(2, Q)$ for any $Q \geq 3$ in polynomial time as it amounts to checking bipartiteness. The situation with ApproxColoring$(3, Q)$ is much more interesting, as there is a huge gap between the value of Q for which an efficient algorithm is known and that for which a hardness result exists. Indeed, until not long ago, the best known polynomial-time algorithm was due to Blum and Karger [9], who solve the problem for $Q = \tilde{O}(n^{3/14})$ colors, where n is the number of vertices and the $\tilde{O}$ notation hides poly-logarithmic factors (as is often

*A preliminary version of this paper has appeared in the proceedings of STOC 2006.

†Hebrew University. Email: dinuri@cs.huji.ac.il. Supported by the Israel Science Foundation and by the Binational Science Foundation.

‡Statistics, U.C. Berkeley. Email: mossel@stat.berkeley.edu. Supported by a Miller fellowship in Computer Science and Statistics, a Sloan fellowship in Mathematics, the Binational Science Foundation, and NSF grants DMS-0504245 and DMS-0548249 (CAREER)

§Department of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Supported by an Alon Fellowship, by the Binational Science Foundation, by the Israel Science Foundation, by the European Commission under the Integrated Project QAP funded by the IST directorate as Contract Number 015848, and by a European Research Council (ERC) Starting Grant.

the case, their algorithm actually solves the search variant). Their work continues a long line of research [37, 8, 27] and is based on a semi-definite relaxation. Very recently, Arora et al. [2] were able to improve this to $Q = O(n^{0.207})$ colors, and the constant in the exponent can possibly be reduced even further if certain geometric conjectures are proven. However, there is some indication that this line of work is limited by $Q = n^{\delta}$ for some fixed $\delta > 0$, see [17]. In contrast, the strongest known hardness result shows that the problem is NP-hard for $Q = 5$ [28, 23]. Thus, the problem is open for all $5 < Q < O(n^{0.207})$. In this paper we prove a hardness result for any constant (i.e., independent of n) value of Q. As we shall explain later, our hardness result is based on the conjectured NP-hardness of certain instances of the label cover problem due to Khot [30].

The situation with $\text{APPROXCOLORING}(q, Q)$ for small values $q \geq 4$ is similar. The best known algorithm, due to Halperin et al. [24], solves $\text{APPROXCOLORING}(q, Q)$ for $Q = \tilde{O}(n^{\alpha_q})$ where $0 < \alpha_q < 1$ is some constant depending on q. For example, $\alpha_4 \approx 0.37$. On the other hand, there are several known NP-hardness results. One of the strongest is due to Khot [29], who improved on an earlier result of Fürer [20] by showing that for any large enough constant q and $Q = q^{\frac{\log q}{25}}$, $\text{APPROXCOLORING}(q, Q)$ is NP-hard. Notice that for any fixed q, Khot's result, as well as all other known hardness results, apply only up to some fixed Q. Our result holds for any $Q > q \geq 3$.

Hardness Results. One of the most successful approaches to deriving hardness proofs, which is also the one we shall take here, is by a reduction from a certain combinatorial problem known as the *label-cover problem* [3]. The PCP theorem [5, 4] says that this problem is NP-hard. In the label-cover problem, we are given an undirected graph and a number R. Each edge is associated with a binary relation on $\{1, \ldots, R\}$ and we refer to it as a *constraint*. The goal is to label the vertices with values from $\{1, \ldots, R\}$ such that the number of satisfied constraints is maximized, where a constraint is satisfied if the labels on the two incident vertices satisfy the relation associated with it.

Without going into the details of the reduction (these details are described in Section 4), we remark that for our reduction to work, the label-cover instances we use must have constraints of a very specific form. For example, we might require all constraints to be *bijections*, i.e., a binary relation in which any labelling of one vertex determines the other, and vice versa. We call this special case the $1{\leftrightarrow}1$-*label-cover*. The precise definition of this and other special cases will appear later.

Unfortunately, these special cases of the label-cover problem are *not* known to be NP-hard. Nevertheless, in his seminal work [30] Khot conjectured that such problems *are* in fact NP-hard, although the tools necessary to prove this conjecture seem to be beyond our current reach. This conjecture has since been heavily scrutinized [36, 13, 22, 14], and so far there is no evidence against the conjecture. This issue is currently one of the central topics in theoretical computer science.

Khot's conjecture is known to imply many strong, and often tight, hardness results. Two examples are the NP-hardness of approximating the VERTEXCOVER problem to within factors below 2 [32], which is tight by a simple greedy algorithm, and the NP-hardness of approximating the MAXCUT problem to within factors larger than roughly 0.878 [31], which is tight by the algorithm of Goemans and Williamson [21]. Our result continues this line of work by showing that (variants of) Khot's conjecture imply strong hardness results for another fundamental problem – that of approximate graph coloring. More specifically, we present three reductions, each from a different special case of the label-cover problem. An exact description of the three reductions will be given later. For now, we just state informally one implication of our reductions.

Theorem 1.1 (informal) *If a certain special case of the label-cover problem is NP-hard, then for any* $Q > 3$, $\text{APPROXCOLORING}(3, Q)$ *is NP-hard.*

Bounds on Bilinear Forms

At the heart of our hardness results are certain bounds on bilinear forms describing the correlation under noise between two functions $f, g : [q]^n \to \mathbb{R}$ where $[q] := \{1, \dots, q\}$. Since we believe these bounds might be useful elsewhere, we now describe them in some detail.

Let T be a symmetric Markov operator on $[q]$ (equivalently, T is the random walk on a regular undirected weighted graph on vertex set $[q]$ possibly with self-loops). We study the expectation $\mathbf{E}_{x,y}[f(x)g(y)]$ where $x \in [q]^n$ is chosen uniformly and $y \in [q]^n$ is obtained from x by applying T to each coordinate independently. Using notation we introduce later, this expectation can also be written as

$$\langle f, T^{\otimes n} g\rangle = \mathbf{E}_x[f(x) \cdot T^{\otimes n} g(x)]. \tag{1}$$

We are interested in the case where T and q are fixed, and n tends to infinity. Our main technical result provides tight bounds on the bilinear form $\langle f, T^{\otimes n} g\rangle$ for bounded functions f, g, in terms of $\mathbf{E}[f]$, $\mathbf{E}[g]$, and ρ, where ρ is the second largest eigenvalue in absolute value of T.

To motivate this result, consider the following concrete example. Take $q = 2$ and $f : \{0,1\}^n \to \{0,1\}$ satisfying $\mathbf{E}[f] = 1/2$ (i.e., f is balanced). Fix some $\rho \in (0,1)$, and let T_ρ be the operator that flips each bit with probability $(1-\rho)/2$,

$$T_\rho = \rho \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (1-\rho) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

We would like to know how high the *stability* of a balanced Boolean function f can be, where the stability of f (with parameter ρ) is defined as the probability that $f(x) = f(y)$ where x is chosen uniformly from $\{0,1\}^n$ and y is obtained from x by flipping each bit independently with probability $(1-\rho)/2$. It is easy to see that the stability of f can be written as

$$\Pr_{x,y}[f(x) = f(y)] = 2\mathbf{E}_{x,y}[f(x)f(y)] = 2\langle f, T_\rho^{\otimes n} f\rangle \tag{2}$$

and hence an upper bound on the stability of a balanced Boolean function would follow from an upper bound on (1).

If the function f depends on just one coordinate of its input, say $f(x_1, \dots, x_n) = x_1$, then its stability is simply $(1+\rho)/2$, and it can be shown that this is the highest possible for any balanced function f. We consider such cases degenerate and instead study functions that do not depend too strongly on any one coordinate of their input (this will be made precise soon). An example of such a function is the majority function, whose value is 1 if and only if more than half of its input bits are 1 (assume for simplicity n is odd). It can be shown that the stability of this function approaches $\frac{1}{2} + \arcsin(\rho)/\pi$ as n goes to infinity. But is majority the most stable function among those who do not depend too strongly on any one coordinate?

The results of [33] imply that the answer is essentially yes. In the work presented here we generalize such stability statements to cases where T is a general reversible Markov operator (and not just the specific operator defined above) and relax the assumptions on influences as discussed later.

Functions with low influences. The notion of the *influence* of a variable on a function defined in a product space [26] played a major role in recent developments in discrete mathematics, see for example [34, 35, 19, 18, 11, 7, 12]. Consider the space $[q]^n$ equipped with the uniform measure. Then the *influence* of the i'th variable on the function $f : [q]^n \to \mathbb{R}$ is defined by

$$I_i(f) := \mathbf{E}[\mathbf{Var}_{x_i}[f(x) | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]].$$

In recent years, starting with [7, 11], an effort has been made to study properties of functions all of whose variables have low influences. In addition to a natural mathematical interest in functions that do not depend strongly on any one coordinate, the study of such functions is essential for proofs of hardness of approximation results, see for example [16, 30, 32, 31].

Main Technical Result. Our main technical result is a bound on (1) for bounded functions f, g that have no common influential coordinate. The upper and lower bounds are stated in terms of inner products of the form $\langle F_\eta, U_\rho F_\nu \rangle_\gamma$ where γ denotes the standard Gaussian measure on $\mathbb{R}$, $F_\mu(s) = 1_{s<t}$ is an indicator function where t is chosen so that $\mathbf{E}_\gamma[F_\mu] = \mu$, and U_ρ is the Ornstein-Uhlenbeck operator, $U_\rho G(x) = \mathbf{E}_{y\sim\gamma}[G(\rho x + \sqrt{1-\rho^2}y)]$. These inner products can be written as certain double integrals, and we mention the easy bound $0 < \langle F_\eta, U_\rho F_\nu \rangle_\gamma < \min(\eta, \nu)$ for all ν, η and ρ strictly between 0 and 1.

Theorem 1.2 *Let T be some fixed symmetric Markov operator on a finite state space $[q]$ whose second largest eigenvalue in absolute value is $\rho = r(T) < 1$. Then for any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $f, g : [q]^n \to [0,1]$ are two functions satisfying*

$$\min\big(I_i(f), I_i(g)\big) < \delta$$

for all i, then it holds that

$$\langle F_{\mathbf{E}[f]}, U_\rho(1 - F_{1-\mathbf{E}[g]}) \rangle_\gamma - \varepsilon \le \langle f, T^{\otimes n} g \rangle \le \langle F_{\mathbf{E}[f]}, U_\rho F_{\mathbf{E}[g]} \rangle_\gamma + \varepsilon. \tag{3}$$

Stated in the contrapositive, this theorem says that if for some two functions f, g, (1) deviates from a certain range, then there must exist a coordinate i that is influential in *both* functions. The fact that we obtain a *common* influential coordinate is crucial in our applications, as well as in a recent application of Theorem 1.2 to the characterization of independent sets in graph powers [15]. We remark that for any T, the bounds in the theorem are essentially tight (see Appendix B).

Going back to our earlier example, consider a balanced function $f : \{0,1\}^n \to \{0,1\}$ all of whose influences are small. Applying the theorem with $q = 2$ and T_ρ, and using (2), gives us an upper bound of essentially $2\langle F_{0.5}, U_\rho F_{0.5} \rangle_\gamma$ on the stability of f. A straightforward calculation shows that this value equals $\frac{1}{2} + \arcsin(\rho)/\pi$, and hence we obtain as a special case of our main theorem that asymptotically, the majority function is the most stable among all balanced functions with low influences.

As mentioned before, this special case is not new to our work. It was originally presented as a conjecture in the work of [31] on the computational hardness of the MAXCUT problem, and has since been proven by Mossel et al [33], who refer to it as the "majority is stablest theorem". For the proof, [33] developed a very powerful *invariance principle*. This principle allows one to translate questions on low-influence functions in the discrete setting (such as the above question on $\{0,1\}^n$) to corresponding questions in other spaces, and in particular Gaussian space. The advantage of this is that one can then apply known (and powerful) results in Gaussian space (such as [10]).

Our proof of Theorem 1.2 also relies on this invariance principle, and can in fact be seen as an extension of the techniques in [33]. Our theorem improves on the one from [33] in the following two aspects:

- The analysis of [33] only considers a very particular noise operator known as the Beckner operator. We extend this to more general noise operators that are given by an arbitrary symmetric Markov operator. In the application to hardness of coloring we apply the result to three different operators. We remark that our main theorem can be easily extended to *reversible* Markov operators, with the stationary distribution taking the place of the uniform distribution.

- Perhaps more importantly, our Theorem 1.2 allows one to conclude about the existence of a common influential coordinate. A more direct application of [33] only implies the existence of an influential variable in *one* of the functions (in other words, one would have a $\max$ instead of the $\min$ in the theorem). As mentioned above, this difference is crucial for our application as well as to the recent results in [15].

Independent sets in graph powers

To demonstrate the usefulness of Theorem 1.2, let us consider a question in the study of independent sets in graph powers. Let $G = ([q], E)$ be a regular, connected, non-bipartite graph on q vertices. Consider the graph G^n on vertex set $[q]^n$ in which vertices $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ are connected if and only if x_i is connected to y_i in G for all i (this is known as the n-fold weak product of G with itself). Let T_G be the symmetric Markov operator corresponding to one step in a random walk in G. It is easy to verify that the operator $T_G^{\otimes n}$ corresponds to one step in G^n.

Let $f, g : [q]^n \to \{0, 1\}$ be two Boolean functions, and think of them as being the indicator functions of two subsets of $[q]^n$. Then, the bilinear form in (1) gives the fraction of edges that are spanned between these two subsets in the graph G^n. In particular, $\langle f, T_G^{\otimes n} f\rangle = 0$ if and only if f is the indicator function of an independent set in G^n. Using the lower bound in Theorem 1.2, we obtain that for any $\mu > 0$ there exists a $\delta > 0$ such that any independent set of measure μ (i.e., $\mathbf{E}[f] = \mu$) must have at least one coordinate with influence at least δ. Less formally, this says that *any reasonably big independent set in graph powers must have some "structure"* (namely, have an influential coordinate). Our hardness result for approximate graph coloring uses Theorem 1.2 in a similar fashion.

We remark that graph powers were studied in a similar context in [1], where a similar "structure" theorem was proved for the restricted case that the independent set has nearly maximal size. Moreover, Theorem 1.2 was recently used in [15] to show that every independent set in a graph power is contained (up to $o(1)$) in a nontrivial set described by a constant number of coordinates.

2 Preliminaries

2.1 Functions on the q-ary hypercube

Let $[q]$ denote the set $\{0, \ldots, q-1\}$. For an element x of $[q]^n$ write $|x|_a$ for the number of coordinates k of x such that $x_k = a$ and $|x| = \sum_{a \neq 0} |x|_a$ for the number of nonzero coordinates.

In this paper we are interested in functions from $[q]^n$ to $\mathbb{R}$. We define an inner product on this space by $\langle f, g\rangle = \frac{1}{q^n} \sum_x f(x)g(x)$. In our applications, we usually take q to be some constant (say, 3) and n to be large.

Definition 2.1 *Let $f : [q]^n \to \mathbb{R}$ be a function. The influence of the i'th variable on f, denoted $I_i(f)$ is defined by*

$$I_i(f) = \mathbf{E}[\mathbf{Var}_{x_i}[f(x)|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]]$$

where $x_1, \ldots, x_n$ are uniformly distributed.

Consider a sequence of vectors $\alpha_0 = \mathbf{1}, \alpha_1, \ldots, \alpha_{q-1} \in \mathbb{R}^q$ forming an orthonormal basis of $\mathbb{R}^q$. Equivalently, we can think of these vectors as functions from $[q]$ to $\mathbb{R}$. These vectors can be used to form an orthonormal basis of the space of functions from $[q]^n$ to $\mathbb{R}$, as follows.

Definition 2.2 *Let $\alpha_0 = \mathbf{1}, \alpha_1, \ldots, \alpha_{q-1}$ be an orthonormal basis of $\mathbb{R}^q$. For $x \in [q]^n$, write $\alpha_x \in \mathbb{R}^{q^n}$ for*

$$\alpha_{x_1} \otimes \alpha_{x_2} \otimes \cdots \otimes \alpha_{x_n}.$$

Equivalently, we can define α_x as the function mapping $y \in [q]^n$ to $\alpha_{x_1}(y_1)\alpha_{x_2}(y_2)\cdots\alpha_{x_n}(y_n)$.

Clearly, any function from $[q]^n$ to $\mathbb{R}$ can be written as a linear combination of α_x for $x \in [q]^n$. This leads to the following definition.

Definition 2.3 *For a function $f : [q]^n \to \mathbb{R}$, define $\hat{f}(\alpha_x) = \langle f, \alpha_x \rangle$ and notice that $f = \sum_x \hat{f}(\alpha_x)\alpha_x$.*

The following standard claim relates the influences of a function to its decomposition. Notice that the claim holds for any choice of orthonormal basis $\alpha_0, \ldots, \alpha_{q-1}$ as long as $\alpha_0 = \mathbf{1}$.

Claim 2.4 *For any function $f : [q]^n \to \mathbb{R}$ and any $i \in \{1, \ldots, n\}$,*

$$I_i(f) = \sum_{x : x_i \neq 0} \hat{f}^2(\alpha_x).$$

Proof: Let us first fix the values of $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$. Then

$$\mathbf{Var}_{x_i}[f] = \mathbf{Var}_{x_i}\Big[\sum_y \hat{f}(\alpha_y)\alpha_y\Big] = \mathbf{Var}_{x_i}\Big[\sum_{y : y_i \neq 0} \hat{f}(\alpha_y)\alpha_y\Big],$$

where the last equality follows from the fact that if $y_i = 0$ then α_y is a constant function of x_i. If $y_i \neq 0$, then the expected value of α_y with respect to x_i is zero. Therefore,

$$\mathbf{Var}_{x_i}\Big[\sum_{y : y_i \neq 0} \hat{f}(\alpha_y)\alpha_y\Big] = \mathbf{E}_{x_i}\left[\left(\sum_{y : y_i \neq 0} \hat{f}(\alpha_y)\alpha_y\right)^2\right] = \mathbf{E}_{x_i}\left[\sum_{y,z : y_i \neq 0, z_i \neq 0} \hat{f}(\alpha_y)\hat{f}(\alpha_z)\alpha_y\alpha_z\right].$$

Thus,

$$I_i(f) = \mathbf{E}_x\left[\sum_{y,z : y_i \neq 0, z_i \neq 0} \hat{f}(\alpha_y)\hat{f}(\alpha_z)\alpha_y\alpha_z\right] = \sum_{y,z : y_i \neq 0, z_i \neq 0} \hat{f}(\alpha_y)\hat{f}(\alpha_z)\mathbf{E}_x[\alpha_y\alpha_z] = \sum_{y : y_i \neq 0} \hat{f}^2(\alpha_y),$$

as needed. ■

We now define the notion of low-level influence.

Definition 2.5 *Let $f : [q]^n \to \mathbb{R}$ be a function, and let $k \leq n$. The low-level influence of the i'th variable on f is defined by*

$$I_i^{\leq k}(f) = \sum_{x : x_i \neq 0, |x| \leq k} \hat{f}^2(\alpha_x).$$

It is easy to see that for any function f,

$$\sum_i I_i^{\leq k}(f) = \sum_{x : |x| \leq k} \hat{f}^2(\alpha_x)\,|x| \leq k \sum_x \hat{f}^2(\alpha_x) = k\|f\|_2^2.$$

In particular, for any function f obtaining values in $[0,1]$, $\sum_i I_i^{\leq k}(f) \leq k$. Moreover, let us mention that $I_i^{\leq k}$ is in fact independent of the particular choice of basis $\alpha_0, \alpha_1, \ldots, \alpha_{q-1}$ as long as $\alpha_0 = \mathbf{1}$. This follows by noting that $I_i^{\leq k}$ is the squared length of the projection of f on the subspace spanned by all α_x with $x_i \neq 0, |x| \leq k$, and that this subspace can be equivalently defined in terms of tensor products of α_0 and $\alpha_0^{\perp}$.

There is a natural equivalence between $[q]^{2n}$ and $[q^2]^n$. As this equivalence is used often in this paper, we introduce the following notation.

Definition 2.6 *For any $x \in [q]^{2n}$ we denote by $\overline{x}$ the element of $[q^2]^n$ given by*

$$\overline{x} = ((x_1, x_2), \dots, (x_{2n-1}, x_{2n})).$$

For any $y \in [q^2]^n$ we denote by $\underline{y}$ the element of $[q]^{2n}$ given by

$$\underline{y} = (y_{1,1}, y_{1,2}, y_{2,1}, y_{2,2}, \dots, y_{n,1}, y_{n,2}).$$

For a function f on $[q]^{2n}$ we denote by $\overline{f}$ the function on $[q^2]^n$ defined by $\overline{f}(y) = f(\underline{y})$. Similarly, for a function f on $[q^2]^n$ we denote by $\underline{f}$ the function on $[q]^{2n}$ defined by $\underline{f}(x) = f(\overline{x})$.

Claim 2.7 *For any function $f : [q]^{2n} \to \mathbb{R}$, any $i \in \{1, \dots, n\}$, and any $k \geq 1$,*

$$I_i^{\leq k}(\overline{f}) \leq I_{2i-1}^{\leq 2k}(f) + I_{2i}^{\leq 2k}(f).$$

Proof: Fix some basis α_x of $[q]^{2n}$ as above and let $\alpha_{\overline{x}}$ be the basis of $[q^2]^n$ defined by $\alpha_{\overline{x}}(\overline{y}) = \alpha_x(y)$. Then, it is easy to see that $\hat{\overline{f}}(\alpha_{\overline{x}}) = \hat{f}(\alpha_x)$. Hence,

$$I_i^{\leq k}(\overline{f}) = \sum_{\overline{x}:\overline{x}_i \neq (0,0), |\overline{x}| \leq k} \hat{\overline{f}}^2(\alpha_{\overline{x}}) \leq \sum_{x: x_{2i-1} \neq 0, |x| \leq 2k} \hat{f}^2(\alpha_x) + \sum_{x: x_{2i} \neq 0, |x| \leq 2k} \hat{f}^2(\alpha_x) = I_{2i-1}^{\leq 2k}(f) + I_{2i}^{\leq 2k}(f)$$

where we used that $|x| \leq 2|\overline{x}|$. ∎

For the following definition, recall that we say that a Markov operator T is *symmetric* if it is reversible with respect to the uniform distribution, i.e., if the transition matrix representing T is symmetric.

Definition 2.8 *Let T be a symmetric Markov operator on $[q]$. Let $1 = \lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{q-1}$ be the eigenvalues of T. We define $r(T)$ to be the second largest eigenvalue in absolute value, that is,*

$$r(T) = \max\{|\lambda_1|, |\lambda_{q-1}|\}.$$

For T as above, we may define a Markov operator $T^{\otimes n}$ on $[q]^n$ in the standard way. Note that if T is symmetric then $T^{\otimes n}$ is also symmetric and $r(T^{\otimes n}) = r(T)$. If we choose $\alpha_0, \dots, \alpha_{q-1}$ to be an orthonormal set of eigenvectors for T with corresponding eigenvalues $\lambda_0, \dots, \lambda_{q-1}$ (so $\alpha_0 = \mathbf{1}$), we see that

$$T^{\otimes n} \alpha_x = \left(\prod_{a \neq 0} \lambda_a^{|x|_a} \right) \alpha_x.$$

and hence

$$T^{\otimes n} f = \sum_x \left(\prod_{a \neq 0} \lambda_a^{|x|_a} \right) \hat{f}(\alpha_x) \alpha_x.$$

holds for any function $f : [q]^n \to \mathbb{R}$.

We now describe two operators that we use in this paper. The first is the Beckner operator, T_ρ. For any $\rho \in [-\frac{1}{q-1}, 1]$, it is defined by $T_\rho(x \to x) = \frac{1}{q} + (1 - \frac{1}{q})\rho$ and $T_\rho(x \to y) = \frac{1}{q}(1 - \rho)$ for any $x \neq y$ in $[q]$. It can be seen that T_ρ is a Markov operator as in Definition 2.8 with $\lambda_1 = \dots = \lambda_{q-1} = \rho$ and hence $r(T_\rho) = |\rho|$.

Another useful operator is the averaging operator, A_S. For a subset $S \subseteq \{1, \dots, n\}$, it acts on functions on $[q]^n$ by averaging over coordinates in S, namely,

$$A_S(f) = \mathbf{E}_{x_S}[f].$$

Notice that the function $A_S(f)$ is independent of the coordinates in S.

2.2 Functions in Gaussian space

We let γ denote the standard Gaussian measure on $\mathbb{R}^n$ with density $(2\pi)^{-n/2}e^{-\|x\|_2^2/2}$. We denote by $\mathbf{E}_\gamma$ the expected value with respect to γ and by $\langle\cdot,\cdot\rangle_\gamma$ the inner product on $L^2(\mathbb{R}^n,\gamma)$. Notice that $\mathbf{E}_\gamma[f] = \langle f, \mathbf{1}\rangle_\gamma$ where $\mathbf{1}$ is the constant 1 function. For $\rho \in [-1,1]$, we denote by U_ρ the *Ornstein-Uhlenbeck operator*, which acts on $L^2(\mathbb{R},\gamma)$ by

$$U_\rho f(x) = \mathbf{E}_{y\sim\gamma}[f(\rho x + \sqrt{1-\rho^2}y)].$$

Since for $x, y \sim \gamma$ we have that $\rho x + \sqrt{1-\rho^2}y$ is also distributed according to the standard Gaussian distribution, $\mathbf{E}_{x\sim\gamma}[U_\rho f(x)] = \mathbf{E}_{x\sim\gamma}[f(x)]$.

Finally, for $0 < \mu < 1$, let $F_\mu : \mathbb{R} \to \{0,1\}$ denote the function $F_\mu(x) = 1_{x<t}$ where t is chosen in such a way that $\mathbf{E}_\gamma[F_\mu] = \mu$. One useful quantity that will appear later is $\langle F_\eta, U_\rho F_\nu\rangle_\gamma$, which by definition can also be written as

$$\langle F_\eta, U_\rho F_\nu\rangle_\gamma = \Pr_{x,y\sim\gamma}[x < s \text{ and } \rho x + \sqrt{1-\rho^2}y < t],$$

where s and t are such that $F_\eta(x) = 1_{x<s}$ and $F_\nu(x) = 1_{x<t}$. It is not difficult to see that for any $\nu, \eta > 0$, and any $\rho \in [-1,1]$, it holds that $\langle F_\eta, U_\rho F_\nu\rangle_\gamma = \langle F_\nu, U_\rho F_\eta\rangle_\gamma$ (say, since U_ρ is self-adjoint) and that

$$\langle F_\tau, U_\rho F_\tau\rangle_\gamma \le \langle F_\eta, U_\rho F_\nu\rangle_\gamma \le \tau,$$

where $\tau = \min(\eta,\nu)$. Moreover, for all $\tau > 0$ and $\rho > -1$ it holds that

$$\langle F_\tau, U_\rho F_\tau\rangle_\gamma > 0.$$

3 An Inequality for Noise Operators

The main analytic result of the paper, Theorem 3.1, is a generalization of the result of [33]. It shows that if the inner product of two functions f and g under some noise operator deviates from a certain range then there must exist an index i such that the low-level influence of the ith variable is large in both f and g. This range depends on the expected value of f and g, and on $r(T)$. Note in particular that Theorem 3.1 implies Theorem 1.2.

Theorem 3.1 *Let q be a fixed integer and let T be a symmetric Markov operator on $[q]$ such that $\rho = r(T) < 1$. Then for any $\varepsilon > 0$ there exist $\delta > 0$ and $k \in \mathbb{N}$ such that if $f, g : [q]^n \to [0,1]$ are two functions satisfying*

$$\min\left(I_i^{\le k}(f), I_i^{\le k}(g)\right) < \delta$$

for all i, then it holds that

$$\langle f, T^{\otimes n} g\rangle \ge \langle F_\mu, U_\rho(1 - F_{1-\nu})\rangle_\gamma - \varepsilon \tag{4}$$

and

$$\langle f, T^{\otimes n} g\rangle \le \langle F_\mu, U_\rho F_\nu\rangle_\gamma + \varepsilon \tag{5}$$

where $\mu = \mathbf{E}[f]$, $\nu = \mathbf{E}[g]$.

Note that (4) follows from (5). Indeed, apply (5) to $1 - g$ to obtain

$$\langle f, T^{\otimes n}(1-g)\rangle \le \langle F_\mu, U_\rho F_{1-\nu}\rangle_\gamma + \varepsilon$$

and then use

$$\langle f, T^{\otimes n}(1-g)\rangle = \langle f, 1\rangle - \langle f, T^{\otimes n} g\rangle = \mu - \langle f, T^{\otimes n} g\rangle = \langle F_\mu, U_\rho 1\rangle_\gamma - \langle f, T^{\otimes n} g\rangle.$$

From now on we focus on proving (5).

Following the approach of [33], the proof consists of two powerful techniques. The first is an inequality by Christer Borell [10] on continuous Gaussian space. The second is an invariance principle shown in [33] that allows us to translate our discrete question to the continuous Gaussian space.

Definition 3.2 (Gaussian analogue of an operator) *Let T be an operator as in Definition 2.8. We define its Gaussian analogue as the operator $\tilde{T}$ on $L^2(\mathbb{R}^{q-1}, \gamma)$ given by*

$$\tilde{T} = U_{\lambda_1} \otimes U_{\lambda_2} \otimes \cdots \otimes U_{\lambda_{q-1}}.$$

For example, the Gaussian analogue of T_ρ is $U_\rho^{\otimes(q-1)}$. We need the following powerful theorem by Borell [10]. It says that the functions that maximize the inner product under the operator U_ρ are the indicator functions of half-spaces.

Theorem 3.3 (Borell [10]) *Let $f, g : \mathbb{R}^n \to [0,1]$ be two functions and let $\mu = \mathbf{E}_\gamma[f], \nu = \mathbf{E}_\gamma[g]$. Then*

$$\langle f, U_\rho^{\otimes n} g\rangle_\gamma \le \langle F_\mu, U_\rho F_\nu\rangle_\gamma.$$

The above theorem only applies to the Ornstein-Uhlenbeck operator. In the following corollary we derive a similar statement for more general operators. The proof follows by writing a general operator as a product of the Ornstein-Uhlenbeck operator and some other operator.

Corollary 3.4 *Let $f, g : \mathbb{R}^{(q-1)n} \to [0,1]$ be two functions and define $\mu = \mathbf{E}_\gamma[f]$, $\nu = \mathbf{E}_\gamma[g]$. Let T be an operator as in Definition 2.8 and let $\rho = r(T)$. Then*

$$\langle f, \tilde{T}^{\otimes n} g\rangle_\gamma \le \langle F_\mu, U_\rho F_\nu\rangle_\gamma.$$

Proof: For $1 \le i \le q-1$, let $\delta_i = \lambda_i/\rho$. Note that $|\delta_i| \le 1$ for all i. Let S be the operator defined by

$$S = U_{\delta_1} \otimes U_{\delta_2} \otimes \cdots \otimes U_{\delta_{q-1}}.$$

Then,

$$U_\rho^{\otimes(q-1)} S = U_\rho U_{\delta_1} \otimes \cdots \otimes U_\rho U_{\delta_{q-1}} = U_{\rho\delta_1} \otimes \cdots \otimes U_{\rho\delta_{q-1}} = \tilde{T}$$

(this is often called the *semi-group property*). It follows that $\tilde{T}^{\otimes n} = U_\rho^{\otimes(q-1)n} S^{\otimes n}$. Since $S^{\otimes g}$ is an averaging operator, the function $S^{\otimes n} g$ obtains values in $[0,1]$ and satisfies $\mathbf{E}_\gamma[S^{\otimes n} g] = \mathbf{E}_\gamma[g]$. Thus the claim follows by applying Theorem 3.3 to the functions f and $S^{\otimes n} g$. ■

Definition 3.5 (Real analogue of a function) *Let $f : [q]^n \to \mathbb{R}$ be a function with decomposition*

$$f = \sum \hat{f}(\alpha_x)\alpha_x.$$

Consider the $(q-1)n$ variables $z_1^1, \ldots, z_{q-1}^1, \ldots, z_1^n, \ldots, z_{q-1}^n$ and let $\Gamma_x = \prod_{i=1, x_i \neq 0}^n z_{x_i}^i$. We define the real analogue of f to be the function $\tilde{f} : \mathbb{R}^{n(q-1)} \to \mathbb{R}$ given by

$$\tilde{f} = \sum \hat{f}(\alpha_x)\Gamma_x.$$

Claim 3.6 *For any two functions $f, g : [q]^n \to \mathbb{R}$ and operator T on $[q]^n$,*

$$\langle f, g\rangle = \langle \tilde{f}, \tilde{g}\rangle_\gamma$$
$$\langle f, T^{\otimes n} g\rangle = \langle \tilde{f}, \tilde{T}^{\otimes n}\tilde{g}\rangle_\gamma$$

where $\tilde{f}, \tilde{g}$ denote the real analogues of f, g respectively and $\tilde{T}$ denotes the Gaussian analogue of T.

Proof: Both α_x and Γ_x form an orthonormal set of functions hence both sides of the first equality are

$$\sum_x \hat{f}(\alpha_x)\hat{g}(\alpha_x).$$

For the second claim, notice that for every x, α_x is an eigenvector of $T^{\otimes n}$ and Γ_x is an eigenvector of $\tilde{T}^{\otimes n}$ and both correspond to the eigenvalue $\prod_{a\neq 0}\lambda_a^{|x|_a}$. Hence, both sides of the second equality are

$$\sum_x \left(\prod_{a\neq 0}\lambda_a^{|x|_a}\right)\hat{f}(\alpha_x)\hat{g}(\alpha_x).$$

■

Definition 3.7 *For any function f with range $\mathbb{R}$, define the function $\mathrm{chop}(f)$ as*

$$\mathrm{chop}(f)(x) = \begin{cases} f(x) & \text{if } f(x)\in[0,1] \\ 0 & \text{if } f(x) < 0 \\ 1 & \text{if } f(x) > 1 \end{cases}$$

The following theorem is proven in [33]. It shows that under certain conditions, if a function f obtains values in $[0,1]$ then $\tilde{f}$ and $\mathrm{chop}(\tilde{f})$ are close. Its proof is non-trivial and builds on the main technical result of [33], a result that is known as an invariance principle. In essence, it shows that the distribution of values obtained by f and that of values obtained by $\tilde{f}$ are close. In particular, since f never deviates from $[0,1]$, it implies that $\tilde{f}$ rarely deviates from $[0,1]$ and hence $\tilde{f}$ and $\mathrm{chop}(\tilde{f})$ are close. See [33] for more details.

Theorem 3.8 ([33, Theorem 3.20]) *There exists a function $\delta_{MOO}(\eta,\varepsilon)$ such that for any $\eta < 1$ and $\varepsilon > 0$ the following holds. For any function $f : [q]^n \to [0,1]$ such that*

$$\forall d \quad \sum_{x:|x|\geq d} |\hat{f}(\alpha_x)|^2 \leq \eta^d \qquad \text{and} \qquad \forall i \quad I_i(f) < \delta_{MOO}(\eta,\varepsilon),$$

it holds that

$$\|\tilde{f} - \mathrm{chop}(\tilde{f})\|_2 \leq \varepsilon.$$

We are now ready to prove the first step in the proof of Theorem 3.1. It is here that we use the invariance principle and Borell's inequality.

Lemma 3.9 *Let q be a fixed integer and let T be a symmetric Markov operator on $[q]$ such that $\rho = r(T) < 1$. Then for any $\varepsilon > 0$, $\eta < 1$, there exists a $\delta > 0$ such that for any functions $f, g : [q]^n \to [0,1]$ satisfying*

$$\forall i \qquad \max\left(I_i(f), I_i(g)\right) < \delta$$

and

$$\forall d \quad \sum_{x:|x|\geq d} |\hat{f}(\alpha_x)|^2 \leq \eta^d, \qquad \forall d \quad \sum_{x:|x|\geq d} |\hat{g}(\alpha_x)|^2 \leq \eta^d,$$

it holds that

$$\langle f, T^{\otimes n} g\rangle \leq \langle F_\mu, U_\rho F_\nu\rangle_\gamma + \varepsilon$$

where $\mu = \mathbf{E}[f]$, $\nu = \mathbf{E}[g]$.

Proof: Let $\mu' = \mathbf{E}_\gamma[\mathrm{chop}(\tilde{f})]$ and $\nu' = \mathbf{E}_\gamma[\mathrm{chop}(\tilde{g})]$. We note that $\langle F_\mu, U_\rho F_\nu\rangle_\gamma$ is a uniformly continuous function of μ and ν. Let ε_1 be chosen such that if $|\mu - \mu'| \le \varepsilon_1$ and $|\nu - \nu'| \le \varepsilon_1$ then it holds that

$$|\langle F_\mu, U_\rho F_\nu\rangle_\gamma - \langle F_{\mu'}, U_\rho F_{\nu'}\rangle_\gamma| \le \varepsilon/2.$$

Let $\varepsilon_2 = \min(\varepsilon/4, \varepsilon_1)$ and let $\delta = \delta_{MOO}(\eta, \varepsilon_2)$ be the value given by Theorem 3.8. Then, using the Cauchy-Schwartz inequality,

$$|\mu' - \mu| = |\mathbf{E}_\gamma[\mathrm{chop}(\tilde{f}) - \tilde{f}]| = |\langle \mathrm{chop}(\tilde{f}) - \tilde{f}, \mathbf{1}\rangle_\gamma| \le \|\mathrm{chop}(\tilde{f}) - \tilde{f}\|_2 \le \varepsilon_2 \le \varepsilon_1.$$

Similarly, we have $|\nu' - \nu| \le \varepsilon_1$. Now,

$$\begin{aligned}
\langle f, T^{\otimes n} g\rangle &= \langle \tilde{f}, \tilde{T}^{\otimes n}\tilde{g}\rangle_\gamma && \text{(Claim 3.6)}\\
&= \langle \mathrm{chop}(\tilde{f}), \tilde{T}^{\otimes n}\mathrm{chop}(\tilde{g})\rangle_\gamma + \\
&\quad\quad \langle \mathrm{chop}(\tilde{f}), \tilde{T}^{\otimes n}(\tilde{g} - \mathrm{chop}(\tilde{g}))\rangle_\gamma + \langle \tilde{f} - \mathrm{chop}(\tilde{f}), \tilde{T}^{\otimes n}\tilde{g}\rangle_\gamma \\
&\le \langle \mathrm{chop}(\tilde{f}), \tilde{T}^{\otimes n}\mathrm{chop}(\tilde{g})\rangle_\gamma + 2\varepsilon_2 \\
&\le \langle F_{\mu'}, U_\rho F_{\nu'}\rangle_\gamma + 2\varepsilon_2 && \text{(Corollary 3.4)}\\
&\le \langle F_\mu, U_\rho F_\nu\rangle_\gamma + \varepsilon/2 + 2\varepsilon_2 \le \langle F_\mu, U_\rho F_\nu\rangle_\gamma + \varepsilon
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwartz inequality together with the fact that $\mathrm{chop}(\tilde{f})$ and $\tilde{g}$ have L_2 norm at most 1 and that $\tilde{T}^{\otimes n}$ is a contraction on L_2. ∎

We complete the proof of Theorem 3.1 by proving:

Lemma 3.10 *Let q be a fixed integer and let T be a symmetric Markov operator on $[q]$ such that $\rho = r(T) < 1$. Then for any $\varepsilon > 0$, there exists a $\delta > 0$ and an integer k such that if $f, g : [q]^n \to [0,1]$ satisfy*

$$\forall i \quad \min\left(I_i^{\le k}(f), I_i^{\le k}(g)\right) < \delta \tag{6}$$

then

$$\langle f, T^{\otimes n} g\rangle \le \langle F_\mu, U_\rho F_\nu\rangle_\gamma + \varepsilon \tag{7}$$

where $\mu = \mathbf{E}[f]$, $\nu = \mathbf{E}[g]$.

Proof: Let $f_1 = T_\eta^{\otimes n} f$ and $g_1 = T_\eta^{\otimes n} g$ where $\eta < 1$ is chosen so that $\rho^j(1 - \eta^{2j}) < \varepsilon/4$ for all j. Then

$$\begin{aligned}
|\langle f_1, T^{\otimes n} g_1\rangle - \langle f, T^{\otimes n} g\rangle| &= \Big|\sum_x \hat{f}(\alpha_x)\hat{g}(\alpha_x)\prod_{a\neq 0}\lambda_a^{|x|_a}(1 - \eta^{2|x|})\Big| \\
&\le \sum_x \rho^{|x|}(1 - \eta^{2|x|})\Big|\hat{f}(\alpha_x)\hat{g}(\alpha_x)\Big| \le \varepsilon/4
\end{aligned}$$

where the last inequality follows from the Cauchy-Schwartz inequality. Thus, in order to prove (7) it suffices to prove

$$\langle f_1, T^{\otimes n} g_1\rangle \le \langle F_\mu, U_\rho F_\nu\rangle_\gamma + 3\varepsilon/4. \tag{8}$$

Let $\delta(\varepsilon/4, \eta)$ be the value given by Lemma 3.9 plugging in $\varepsilon/4$ for ε. Let $\delta' = \delta(\varepsilon/4, \eta)/2$. Let k be chosen so that $\eta^{2k} < \min(\delta', \varepsilon/4)$. Define $C = k/\delta'$ and $\delta = (\varepsilon/8C)^2 < \delta'$. Let

$$B_f = \{i : I_i^{\le k}(f) \ge \delta'\}, \quad B_g = \{i : I_i^{\le k}(g) \ge \delta'\}.$$

We note that B_f and B_g are of size at most $C = k/\delta'$. By (6), we have that whenever $i \in B_f$, $I_i^{\leq k}(g) < \delta$. Similarly, for every $i \in B_g$ we have $I_i^{\leq k}(f) < \delta$. In particular, B_f and B_g are disjoint.

Recall the averaging operator A. We now let

$$f_2 = A_{B_f}(f_1) = \sum_{x:x_{B_f}=0} \hat{f}(\alpha_x)\alpha_x \eta^{|x|},$$

$$g_2 = A_{B_g}(g_1) = \sum_{x:x_{B_g}=0} \hat{g}(\alpha_x)\alpha_x \eta^{|x|}.$$

Clearly, $\mathbf{E}[f_2] = \mathbf{E}[f]$ and $\mathbf{E}[g_2] = \mathbf{E}[g]$, and for all x, $f_2(x), g_2(x) \in [0,1]$. It is easy to see that $I_i(f_2) = 0$ if $i \in B_f$ and $I_i(f_2) \leq I_i^{\leq k}(f) + \eta^{2k} < 2\delta'$ otherwise and similarly for g_2. Thus, for any i, $\max(I_i(f_2), I_i(g_2)) < 2\delta'$. We also see that for any d, $\sum_{x:|x|\geq d} |\hat{f}_2(\alpha_x)|^2 \leq \eta^d$ and the same for g_2. Thus, we can apply Lemma 3.9 to obtain that

$$\langle f_2, T^{\otimes n} g_2 \rangle \leq \langle F_\mu, U_\rho F_\nu \rangle_\gamma + \varepsilon/4.$$

In order to show (8) and complete the proof, we show that

$$|\langle f_1, T^{\otimes n} g_1 \rangle - \langle f_2, T^{\otimes n} g_2 \rangle| \leq \varepsilon/2.$$

This follows by

$$\begin{aligned}
|\langle f_1, T^{\otimes n} g_1 \rangle - \langle f_2, T^{\otimes n} g_2 \rangle| &= \Big| \sum_{x:x_{B_f \cup B_g} \neq 0} \hat{f}(\alpha_x)\hat{g}(\alpha_x) \prod_{a \neq 0} \lambda_a^{|x|_a} \eta^{2|x|} \Big| \\
&\leq \eta^{2k} \sum_{x:|x|\geq k} \Big| \hat{f}(\alpha_x)\hat{g}(\alpha_x) \Big| + \sum \Big\{ \Big| \hat{f}(\alpha_x)\hat{g}(\alpha_x) \Big| : x_{B_f \cup B_g} \neq 0, |x| \leq k \Big\} \\
&\leq \varepsilon/4 + \sum_{i \in B_f \cup B_g} \sum \Big\{ \Big| \hat{f}(\alpha_x)\hat{g}(\alpha_x) \Big| : x_i \neq 0, |x| \leq k \Big\} \\
&\leq \varepsilon/4 + \sum_{i \in B_f \cup B_g} \sqrt{I_i^{\leq k}(f)} \sqrt{I_i^{\leq k}(g)} \\
&\leq \varepsilon/4 + \sqrt{\delta}(|B_f| + |B_g|) \\
&\leq \varepsilon/4 + 2C\sqrt{\delta} = \varepsilon/2,
\end{aligned}$$

where the next-to-last inequality holds because for each $i \in B_f \cup B_g$ one of $I_i^{\leq k}(f), I_i^{\leq k}(g)$ is at most δ and the other is at most 1. ∎

The final theorem of this section is needed only for the APPROXCOLORING$(3, Q)$ result. Here, the operator T acts on $[q^2]$ and is assumed to have an additional property. Before proceeding, it is helpful to recall Definition 2.6.

Theorem 3.11 *Let q be a fixed integer and let T be a symmetric Markov operator on $[q^2]$ such that $\rho = r(T) < 1$. Suppose moreover, that T has the following property. Given (x_1, x_2) chosen uniformly at random and (y_1, y_2) chosen according to T applied to (x_1, x_2) we have that (x_2, y_2) is distributed uniformly at random. Then for any $\varepsilon > 0$, there exists a $\delta > 0$ and an integer k such that for any functions $f, g : [q]^{2n} \to [0,1]$ satisfying that for $i = 1, \ldots, n$*

$$\min\left(I_{2i-1}^{\leq k}(f), I_{2i-1}^{\leq k}(g)\right) < \delta, \qquad \min\left(I_{2i-1}^{\leq k}(f), I_{2i}^{\leq k}(g)\right) < \delta, \text{ and } \qquad \min\left(I_{2i}^{\leq k}(f), I_{2i-1}^{\leq k}(g)\right) < \delta$$

it holds that

$$\langle \overline{f}, T^{\otimes n}\overline{g}\rangle \geq \langle F_\mu, U_\rho(1 - F_{1-\nu})\rangle_\gamma - \varepsilon \tag{9}$$

and

$$\langle \overline{f}, T^{\otimes n}\overline{g}\rangle \leq \langle F_\mu, U_\rho F_\nu\rangle_\gamma + \varepsilon \tag{10}$$

where $\mu = \mathbf{E}[f]$, $\nu = \mathbf{E}[g]$.

Proof: As in Theorem 3.1, (9) follows from (10) so it is enough to prove (10). Assume first that in addition to the three conditions above we also have that for all $i = 1, \ldots, n$,

$$\min\left(I_{2i}^{\leq k}(f), I_{2i}^{\leq k}(g)\right) < \delta. \tag{11}$$

Then it follows that for all i, either both $I_{2i-1}^{\leq k}(f)$ and $I_{2i}^{\leq k}(f)$ are smaller than δ or both $I_{2i-1}^{\leq k}(g)$ and $I_{2i}^{\leq k}(g)$ are smaller than δ. Hence, by Claim 2.7, we know that for all i we have

$$\min\left(I_i^{\leq k/2}(\overline{f}), I_i^{\leq k/2}(\overline{g})\right) < 2\delta$$

and the result then follows from Lemma 3.10. However, we do not have this extra condition and hence we have to deal with 'bad' coordinates i for which $\min(I_{2i}^{\leq k}(f), I_{2i}^{\leq k}(g)) \geq \delta$. Notice that for such i it must be the case that both $I_{2i-1}^{\leq k}(f)$ and $I_{2i-1}^{\leq k}(g)$ are smaller than δ. Informally, the proof proceeds as follows. We first define functions f_1, g_1 that are obtained from f, g by adding a small amount of noise. We then obtain f_2, g_2 from f_1, g_1 by averaging the coordinates $2i - 1$ for bad i. Finally, we obtain f_3, g_3 from f_2, g_2 by averaging the coordinate $2i$ for bad i. The point here is to maintain $\langle \overline{f}, T^{\otimes n}\overline{g}\rangle \approx \langle \overline{f}_1, T^{\otimes n}\overline{g}_1\rangle \approx \langle \overline{f}_2, T^{\otimes n}\overline{g}_2\rangle \approx \langle \overline{f}_3, T^{\otimes n}\overline{g}_3\rangle$. The condition in Equation 11 now applies to f_3, g_3 and we can apply Lemma 3.10, as described above. We now describe the proof in more detail.

We first define $\overline{f_1} = T_\eta^{\otimes n}\overline{f}$ and $\overline{g_1} = T_\eta^{\otimes n}\overline{g}$ where $\eta < 1$ is chosen so that $\rho^j(1 - \eta^{2j}) < \varepsilon/4$ for all j. As in the previous lemma it is easy to see that

$$|\langle \overline{f_1}, T^{\otimes n}\overline{g_1}\rangle - \langle \overline{f}, T^{\otimes n}\overline{g}\rangle| < \varepsilon/4$$

and thus it suffices to prove that

$$\langle \overline{f_1}, T^{\otimes n}\overline{g_1}\rangle \leq \langle F_\mu, U_\rho F_\nu\rangle_\gamma + 3\varepsilon/4.$$

Let $\delta(\varepsilon/2, \eta), k(\varepsilon/2, \eta)$ be the values given by Lemma 3.10 with ε taken to be $\varepsilon/2$. Let $\delta' = \delta(\varepsilon/2, \eta)/2$. Choose a large enough k so that $128k\eta^k < \varepsilon^2\delta'$ and $k/2 > k(\varepsilon/2, \eta)$. We let $C = k/\delta'$ and $\delta = \varepsilon^2/128C$. Notice that $\delta < \delta'$ and $\eta^k < \delta$. Finally, let

$$B = \left\{ i \;\middle|\; I_{2i}^{\leq k}(f) \geq \delta', I_{2i}^{\leq k}(g) \geq \delta' \right\}.$$

We note that B is of size at most C. We also note that if $i \in B$ then we have $I_{2i-1}^{\leq k}(f) < \delta$ and $I_{2i-1}^{\leq k}(g) < \delta$. We claim that this implies that $I_{2i-1}(f_1) \leq \delta + \eta^k < 2\delta$ and similarly for g. To see that, take any orthonormal basis $\beta_0 = 1, \beta_1, \ldots, \beta_{q-1}$ of $\mathbb{R}^q$ and notice that we can write

$$f_1 = \sum_{x \in [q]^{2n}} \hat{f}(\beta_x)\eta^{|\overline{x}|}\beta_x.$$

Hence,

$$I_{2i-1}(f_1) = \sum_{\substack{x \in [q]^{2n} \\ x_{2i-1} \neq 0}} \hat{f}(\beta_x)^2\eta^{2|\overline{x}|} < \delta + \eta^k \sum_{\substack{x \in [q]^{2n} \\ |x| > k}} \hat{f}(\beta_x)^2 \leq \delta + \eta^k$$

where we used that the number of nonzero elements in $\overline{x}$ is at least half of that in x.

Next, we define $f_2 = A_{2B-1}(f_1)$ and $g_2 = A_{2B-1}(g_1)$ where A is the averaging operator and $2B-1$ denotes the set $\{2i-1 \mid \ i \in B\}$. Note that

$$\|\overline{f_2} - \overline{f_1}\|_2^2 = \|f_2 - f_1\|_2^2 \leq \sum_{i \in B} I_{2i-1}(f_1) \leq 2C\delta.$$

and similarly,

$$\|\overline{g_2} - \overline{g_1}\|_2^2 = \|g_2 - g_1\|_2^2 \leq 2C\delta.$$

Thus

$$\begin{aligned}|\langle \overline{f_1}, T^{\otimes n}\overline{g_1}\rangle - \langle \overline{f_2}, T^{\otimes n}\overline{g_2}\rangle| &\leq |\langle \overline{f_1}, T^{\otimes n}\overline{g_1}\rangle - \langle \overline{f_1}, T^{\otimes n}\overline{g_2}\rangle| + |\langle \overline{f_1}, T^{\otimes n}\overline{g_2}\rangle - \langle \overline{f_2}, T^{\otimes n}\overline{g_2}\rangle| \\ &\leq 2\sqrt{2C\delta} = \varepsilon/4\end{aligned}$$

where the last inequality follows from the Cauchy-Schwartz inequality together with the fact that $\|\overline{f_1}\|_2 \leq 1$ and also $\|T^{\otimes n}\overline{g_2}\|_2 \leq 1$. Hence, it suffices to prove

$$\langle \overline{f_2}, T^{\otimes n}\overline{g_2}\rangle \leq \langle F_\mu, U_\rho F_\nu\rangle_\gamma + \varepsilon/2.$$

We now define $f_3 = A_{2B}(f_2)$ and $g_3 = A_{2B}(g_2)$. Equivalently, we have $\overline{f_3} = A_B(\overline{f_1})$ and $\overline{g_3} = A_B(\overline{g_1})$. We show that $\langle \overline{f_2}, T^{\otimes n}\overline{g_2}\rangle = \langle \overline{f_3}, T^{\otimes n}\overline{g_3}\rangle$. Let α_x, $x \in [q^2]^n$, be an orthonormal basis of eigenvectors of $T^{\otimes n}$. Then

$$\langle \overline{f_3}, T^{\otimes n}\overline{g_3}\rangle = \sum_{x,y \in [q^2]^n, x_B = y_B = 0} \hat{\overline{f_1}}(\alpha_x)\hat{\overline{g_1}}(\alpha_y)\langle \alpha_x, T^{\otimes n}\alpha_y\rangle. \tag{12}$$

Moreover, since A is a linear operator and f_1 can be written as $\sum_{x \in [q^2]^n} \hat{\overline{f_1}}(\alpha_x)\underline{\alpha_x}$ and similarly for g_1, we have

$$\langle \overline{f_2}, T^{\otimes n}\overline{g_2}\rangle = \sum_{x,y \in [q^2]^n} \hat{\overline{f_1}}(\alpha_x)\hat{\overline{g_1}}(\alpha_y)\langle \overline{A_{2B-1}(\underline{\alpha_x})}, T^{\otimes n}\overline{A_{2B-1}(\underline{\alpha_y})}\rangle. \tag{13}$$

First, notice that when $x_B = 0$, $\overline{A_{2B-1}(\underline{\alpha_x})} = \alpha_x$ since α_x does not depend on coordinates in B. Hence, in order to show that the expressions in (12) and (13) are equal, it suffices to show that

$$\langle \overline{A_{2B-1}(\underline{\alpha_x})}, T^{\otimes n}\overline{A_{2B-1}(\underline{\alpha_y})}\rangle = 0$$

unless $x_B = y_B = 0$. So assume without loss of generality that $i \in B$ is such that $x_i \neq 0$. The above inner product can be equivalently written as

$$\mathbf{E}_{z,z' \in [q^2]^n}[\overline{A_{2B-1}(\underline{\alpha_x})}(z) \cdot \overline{A_{2B-1}(\underline{\alpha_y})}(z')]$$

where z is chosen uniformly at random and z' is chosen according to $T^{\otimes n}$ applied to z. Fix some arbitrary values to $z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n$ and $z'_1, \ldots, z'_{i-1}, z'_{i+1}, \ldots, z'_n$ and let us show that

$$\mathbf{E}_{z_i, z'_i \in [q^2]}[\overline{A_{2B-1}(\underline{\alpha_x})}(z) \cdot \overline{A_{2B-1}(\underline{\alpha_y})}(z')] = 0.$$

Since $i \in B$, the two expressions inside the expectation do not depend on $z_{i,1}$ and $z'_{i,1}$ (where by $z_{i,1}$ we mean the first coordinate of $\underline{z_i}$). Moreover, by our assumption on T, $z_{i,2}$ and $z'_{i,2}$ are independent. Hence, the above expectation is equal to

$$\mathbf{E}_{z_i \in [q^2]}[\overline{A_{2B-1}(\underline{\alpha_x})}(z)] \cdot \mathbf{E}_{z'_i \in [q^2]}[\overline{A_{2B-1}(\underline{\alpha_y})}(z')].$$

Since $x_i \neq 0$, the first expectation is zero. This establishes that $\langle \overline{f_2}, T^{\otimes n} \overline{g_2} \rangle = \langle \overline{f_3}, T^{\otimes n} \overline{g_3} \rangle$.

The functions f_3, g_3 satisfy the property that for every $i = 1, \ldots, n$, either both $I_{2i-1}^{\leq k}(f_3)$ and $I_{2i}^{\leq k}(f_3)$ are smaller than δ' or both $I_{2i-1}^{\leq k}(g_3)$ and $I_{2i}^{\leq k}(g_3)$ are smaller than δ'. By Claim 2.7, we get that for $i = 1, \ldots, n$, either $I_i^{\leq k/2}(\overline{f_3})$ or $I_i^{\leq k/2}(\overline{g_3})$ is smaller $2\delta'$. We can now apply Lemma 3.10 to obtain

$$\langle \overline{f_3}, T^{\otimes n} \overline{g_3} \rangle \leq \langle F_\mu, U_\rho F_\nu \rangle_\gamma + \varepsilon/2.$$

■

4 Approximate Coloring

As mentioned in the introduction, one of the most successful approaches to deriving hardness proofs, which is also the one we shall take here, is by a reduction from a combinatorial problem known as the *label-cover problem*. To recall, in the label-cover problem we are given an undirected graph together with a constraint (i.e., a binary relation on $\{1, \ldots, R\}$) for each edge. The goal is to label the vertices with values from $\{1, \ldots, R\}$ such that the number of satisfied constraints is maximized, where a constraint is satisfied if the labels on the two incident vertices satisfy the relation associated with it. It is known that in this problem (as well as in many of its variants), it is NP-hard to tell whether there exists a way to label the vertices such that *all* constraints are satisfied, or whether any labeling satisfies at most, say, 0.01 fraction of the constraints.

Our reduction follows the general paradigm of [6, 25]. Each vertex of the label-cover instance is replaced with a block of vertices, often known as a *gadget*. In our case, the gadget is simply a set of q^R vertices, and we think of them as corresponding to elements of $[q]^R$. We then add edges between these gadgets in a way that "encodes" the label-cover constraints. For the reduction to work, we need to have two properties. First, if the label-cover is satisfiable, then the resulting graph is q-colorable (this is known as the *completeness* part). This property would follow immediately from our construction. The more difficult part is to show that if there is no way to satisfy more than 0.01 fraction of the constraints in the label-cover instance, then the resulting graph has chromatic number at least Q (this is known as the *soundness* part). The way this is shown is by assuming towards contradiction that there exists a coloring with less than Q colors, and then "decoding" it into a labeling of the label-cover instance that satisfies more than 0.01 of the constraints. It is this part that is usually the most difficult to establish. In our case, we will apply Theorem 1.2 to detect influential coordinates in each block based on the coloring given to it.

The above outline hides one very important fact: for our reduction to work, the label-cover instances we use must have constraints of a very specific form. For example, we might require all constraints to be *bijections*, i.e., a binary relation in which any labelling of one vertex determines the other, and vice versa. We call this special case $1{\leftrightarrow}1$-*label-cover*. We will also consider two other restrictions of the label-cover problem, which we call the $2{\leftrightarrow}2$-label-cover and the $\rtimes$-label-cover (read: alpha-label-cover). The precise definitions of these problems will appear later.

As already discussed in the introduction, these special cases of the label-cover problem are *not* known to be NP-hard. Nevertheless, Khot's "unique games conjecture" [30] asserts that such problems *are* in fact NP-hard. The conjecture has been heavily scrutinized [36, 13, 22, 14], and so far there is no evidence against the conjecture.

Our Hardness Results: We now describe our hardness results in more detail. In addition to ApproxColoring(q, Q), we consider the following computational problem, defined for any $\varepsilon > 0$.

ALMOST3COLORING$_\varepsilon$: Given a graph $G = (V, E)$, decide between

- There exists a set $V' \subseteq V$, $|V'| \geq (1 - \varepsilon)|V|$ such that $\chi(G|_{V'}) \leq 3$ where $G|_{V'}$ is the graph induced by V'.
- Every independent set $S \subseteq V$ in G has size $|S| \leq \varepsilon |V|$.

Observe that these two items are mutually exclusive for $\varepsilon < 1/4$.

We consider three conjectures: the $1\leftrightarrow 1$ conjecture, the $2\leftrightarrow 2$ conjecture, and the $\rhd\!\!<$ conjecture. Roughly speaking, each conjecture says that in the corresponding label-cover instances it is NP-hard to distinguish between completely satisfiable instances, and instances that are almost completely unsatisfiable. The only exception is the $1\leftrightarrow 1$ conjecture: it is easy to see that checking if a $1\leftrightarrow 1$-label-cover is *completely* satisfiable can be done in polynomial time. Hence the $1\leftrightarrow 1$ conjecture says that it is NP-hard to distinguish between *almost* completely satisfiable and almost completely unsatisfiable. This drawback of the $1\leftrightarrow 1$ conjecture, often known as 'imperfect completeness', prevents us from using it for proving the hardness of the approximate coloring problem. Instead, we use it to show hardness of the (somewhat harder) problem ALMOST3COLORING.

We present three reductions, each from a different special case of the label-cover problem. These reductions yield the following.

- For any constant $\varepsilon > 0$, the $1\leftrightarrow 1$ conjecture implies the NP-hardness of ALMOST3COLORING$_\varepsilon$.
- For any constant $Q > 4$, the $2\leftrightarrow 2$ conjecture implies that APPROXCOLORING$(4, Q)$ is NP-hard. This also holds for APPROXCOLORING(q, Q) for any $q \geq 4$.
- For any constant $Q > 3$, the $\rhd\!\!<$ conjecture implies that APPROXCOLORING$(3, Q)$ is NP-hard. This also holds for APPROXCOLORING(q, Q) for any $q \geq 3$.

We remark that Khot's original conjectures actually refer to slightly different variants of the label-cover problem. Most notably, his label-cover instances are bipartite. However, as we shall show later, Khot's unique-games conjecture implies our $1\leftrightarrow 1$ conjecture, and Khot's two-to-one conjecture implies our $2\leftrightarrow 2$ conjecture. The $\rhd\!\!<$ conjecture is, to the best of our knowledge, new, and seems to be not weaker than the $2\leftrightarrow 2$ conjecture.

Future work: Our constructions can be extended in several ways. First, using similar techniques, one can show hardness of APPROXCOLORING(q, Q) based on the d-to-1 conjecture of Khot for larger values of d (and not only $d = 2$ as we do here). It would be interesting to find out how q depends on d. Second, by strengthening the current conjectures to sub-constant values, one can obtain hardness for Q that depends on n, the number of vertices in the graph. Again, it is interesting to see how large Q can be. Finally, let us mention that in all our reductions we in fact show in the soundness case that there are no independent sets of relative size larger than ε for arbitrarily small constant ε (note that this is somewhat stronger than showing that there is no Q-coloring). In fact, a more careful analysis can be used to obtain the stronger statement that there are no 'almost-independent' sets of relative size larger than ε.

Organization: In Section 4.1, we describe the three conjectures along with some definitions. We then prove the three reductions mentioned above. The three reductions are very similar, each combining a conjecture with an appropriately constructed noise operator. In Section 4.2 we describe the three noise operators, and in Section 4.3 we spell out the reductions. Then, in Sections 4.4 and 4.5, we prove the completeness and soundness of the three reductions.

4.1 Label-cover problems

Definition 4.1 *A label-cover instance is a triple $G = ((V, E), R, \Psi)$ where (V, E) is a graph, R is an integer, and $\Psi = \left\{\psi_e \subseteq \{1, \ldots, R\}^2 \,\middle|\, e \in E\right\}$ is a set of constraints (relations), one for each edge. For a given labeling $L : V \to \{1, \ldots, R\}$, let*

$$\mathsf{sat}_L(G) = \Pr_{e=(u,v)\in E}[(L(u), L(v)) \in \psi_e], \qquad \mathsf{sat}(G) = \max_L(\mathsf{sat}_L(G)) .$$

For $t, R \in \mathbb{N}$ let $\binom{R}{\le t}$ denote the collection of all subsets of $\{1, \ldots, R\}$ whose size is at most t.

Definition 4.2 *A t-labeling is a function $L : V \to \binom{R}{\le t}$ that labels each vertex $v \in V$ with a subset of values $L(v) \subseteq \{1, \ldots, R\}$ such that $|L(v)| \le t$ for all $v \in V$. A t-labeling L is said to* satisfy *a constraint $\psi \subseteq \{1, \ldots, R\}^2$ over variables u and v iff there are $a \in L(u)$, $b \in L(v)$ such that $(a, b) \in \psi$. In other words, iff $(L(u) \times L(v)) \cap \psi \neq \emptyset$.*

In the special case of $t = 1$, a 1-labeling is essentially a labeling $L : V \to \{1, \ldots, R\}$ (except that some vertices might get no label).

Similar to the definition of $\mathsf{sat}(G)$, we also define $\mathsf{isat}(G)$ ("induced-sat") to be the relative size of the largest set of vertices for which there is a labeling that satisfies *all* of the induced edges.

$$\mathsf{isat}(G) = \max_S \left\{ \frac{|S|}{|V|} \,\middle|\, \exists L : S \to \{1, \ldots, R\} \text{ that satisfies all the constraints induced by } S \subseteq V \right\}.$$

Let $\mathsf{isat}_t(G)$ denote the relative size of the largest set of vertices $S \subseteq V$ for which there is a t-labeling that satisfies *all* the constraints induced by S.

$$\mathsf{isat}_t(G) = \max_S \left\{ \frac{|S|}{|V|} \,\middle|\, \exists L : S \to \binom{R}{\le t} \text{ that satisfies all the constraints induced by } S \subseteq V \right\}.$$

We next describe three conjectures on which our reductions are based. The main difference between the three conjectures is in the type of constraints that are allowed. The three types are defined next, and also illustrated in Figure 1.

Definition 4.3 ($1\leftrightarrow1$-constraint) *A $1\leftrightarrow1$ constraint is a relation $\{(i, \pi(i))\}_{i=1}^R$, where $\pi : \{1, \ldots, R\} \to \{1, \ldots, R\}$ is any arbitrary permutation. The constraint is satisfied by (a, b) iff $b = \pi(a)$.*

Definition 4.4 ($2\leftrightarrow2$-constraint) *A $2\leftrightarrow2$ constraint is defined by a pair of permutations $\pi_1, \pi_2 : \{1, \ldots, 2R\} \to \{1, \ldots, 2R\}$ and the relation*

$$2\leftrightarrow2 = \{(2i, 2i), (2i, 2i-1), (2i-1, 2i), (2i-1, 2i-1)\}_{i=1}^R .$$

The constraint is satisfied by (a, b) iff $(\pi_1^{-1}(a), \pi_2^{-1}(b)) \in 2\leftrightarrow2$.

Definition 4.5 ($\rhd\!\!<$-constraint) *An $\rhd\!\!<$ constraint is defined by a pair of permutations $\pi_1, \pi_2 : \{1, \ldots, 2R\} \to \{1, \ldots, 2R\}$ and the relation*

$$\rhd\!\!< = \{(2i-1, 2i-1), (2i, 2i-1), (2i-1, 2i)\}_{i=1}^R .$$

The constraint is satisfied by (a, b) iff $(\pi_1^{-1}(a), \pi_2^{-1}(b)) \in \rhd\!\!<$.

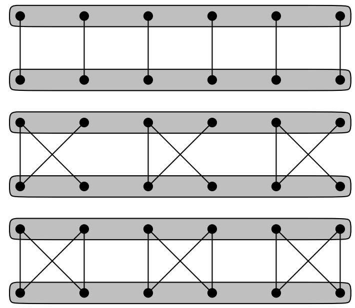

Figure 1: Three types of constraints (top to bottom): $1\leftrightarrow 1, \ltimes, 2\leftrightarrow 2$

Conjecture 4.6 ($1\leftrightarrow 1$ Conjecture) *For any $\varepsilon, \zeta > 0$ and $t \in \mathbb{N}$ there exists some $R \in \mathbb{N}$ such that given a label-cover instance $G = \langle (V, E), R, \Psi \rangle$ where all constraints are $1\leftrightarrow 1$-constraints, it is NP-hard to decide between*

- $\mathsf{isat}(G) \geq 1 - \zeta$
- $\mathsf{isat}_t(G) < \varepsilon$

It is easy to see that the above problem is in P when $\zeta = 0$.

Conjecture 4.7 ($2\leftrightarrow 2$ Conjecture) *For any $\varepsilon > 0$ and $t \in \mathbb{N}$ there exists some $R \in \mathbb{N}$ such that given a label-cover instance $G = \langle (V, E), 2R, \Psi \rangle$ where all constraints are $2\leftrightarrow 2$-constraints, it is NP-hard to decide between*

- $\mathsf{sat}(G) = 1$
- $\mathsf{isat}_t(G) < \varepsilon$

The above two conjectures are no stronger than the corresponding conjectures of Khot. Namely, our $1\leftrightarrow 1$ conjecture is not stronger than Khot's (bipartite) unique games conjecture, and our $2\leftrightarrow 2$ conjecture is not stronger than Khot's (bipartite) two-to-one conjecture. The former claim was already proven by Khot and Regev in [32]. The latter claim is proven in a similar way. For completeness, we include both proofs in Appendix A. We also make a third conjecture that is used in our reduction to APPROXCOLORING$(3, Q)$. This conjecture seems stronger than Khot's conjectures.

Conjecture 4.8 ($\ltimes$ Conjecture) *For any $\varepsilon > 0$ and $t \in \mathbb{N}$ there exists some $R \in \mathbb{N}$ such that given a label-cover instance $G = \langle (V, E), 2R, \Psi \rangle$ where all constraints are $\ltimes$-constraints, it is NP-hard to decide between*

- $\mathsf{sat}(G) = 1$
- $\mathsf{isat}_t(G) < \varepsilon$

Remark: The (strange-looking) $\ltimes$-shaped constraints have already appeared before in [16]. There, it is essentially proven that for all $\varepsilon, \zeta > 0$ given a label-cover instance G where all constraints are $\ltimes$-constraints, it is NP-hard to distinguish between

- $\mathsf{isat}(G) > 1 - \zeta$

- $\mathsf{isat}_{t=1}(G) < \varepsilon$

The main difference between their theorem and our conjecture is that in our conjecture we consider any constant t, while in their case t is 1. Another difference is that in our conjecture we assume perfect completeness (i.e., $\mathsf{sat}(G) = 1$).[1]

4.2 Noise operators

We now define the noise operators corresponding to the $1{\leftrightarrow}1$-constraints, $\rhd\!\!<$-constraints, and $2{\leftrightarrow}2$-constraints. The noise operator that corresponds to the $1{\leftrightarrow}1$-constraints is the simplest, and acts on $\{0,1,2\}$. For the other two cases, since the constraints involve pairs of coordinates, we obtain an operator on $\{0,1,2\}^2$ and an operator on $\{0,1,2,3\}^2$. See Figure 2 for an illustration.

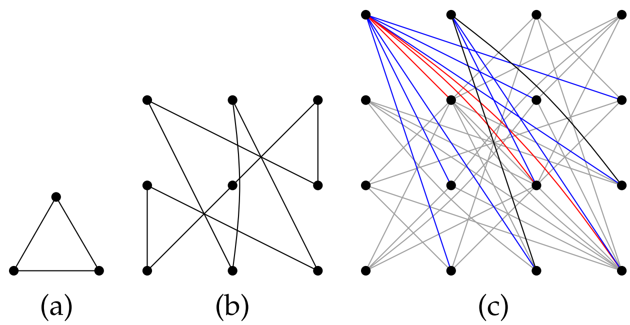

Figure 2: Three noise operators (edge weights not shown) corresponding to: (a) $1{\leftrightarrow}1$, (b) $\rhd\!\!<$, and (c) $2{\leftrightarrow}2$.

Lemma 4.9 *There exists a symmetric Markov operator T on $\{0,1,2\}$ such that $r(T) < 1$ and such that if $T(x \leftrightarrow y) > 0$ then $x \neq y$.*

Proof: Take the operator given by

$$T = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

See Figure 2(a). ■

Lemma 4.10 *There exists a symmetric Markov operator T on $\{0,1,2,3\}^2$ such that $r(T) < 1$ and such that if $T((x_1,x_2) \leftrightarrow (y_1,y_2)) > 0$ then $\{x_1,x_2\} \cap \{y_1,y_2\} = \emptyset$.*

Proof: Our operator has three types of transitions, with transitions probabilities β_1, β_2, and β_3.

[1] The main idea in their construction is to take an NP-hard label-cover as given by the parallel repetition theorem applied to the PCP theorem, and to construct a new $\rhd\!\!<$-label-cover with $\binom{R|X|}{l}$ variables corresponding to all subsets of size l of $X \times \{1,\ldots,R\}$, where $l = cR$ for some large constant c. The number of labels is equal to the number of binary strings of length l whose Hamming weight is at least $l/2R$. Constraints are placed between any pair of l-tuples for which (i) their intersection has size $l-1$, and (ii) the unique elements, one from each l-tuple, correspond to an inconsistency in the original label-cover. These constraints check for agreement on their intersection and that not both unique elements are 1, and are therefore essentially $\rhd\!\!<$-constraints.

- With probability β_1 we have $(x, x) \leftrightarrow (y, y)$ where $x \neq y$.
- With probability β_2 we have $(x, x) \leftrightarrow (y, z)$ where x, y, z are all different.
- With probability β_3 we have $(x, y) \leftrightarrow (z, w)$ where x, y, z, w are all different.

These transitions are illustrated in Figure 2(c). For T to be a symmetric Markov operator, we need that β_1, β_2 and β_3 are non-negative and

$$3\beta_1 + 6\beta_2 = 1, \quad 2\beta_2 + 2\beta_3 = 1.$$

It is easy to see that the two equations above have solutions bounded away from 0 and that the corresponding operator has $r(T) < 1$. For example, choose $\beta_1 = \frac{1}{12}$, $\beta_2 = \frac{1}{8}$, and $\beta_3 = \frac{3}{8}$. ■

Lemma 4.11 *There exists a symmetric Markov operator T on $\{0, 1, 2\}^2$ such that $r(T) < 1$ and such that if $T((x_1, x_2) \leftrightarrow (y_1, y_2)) > 0$ then $x_1 \notin \{y_1, y_2\}$ and $y_1 \notin \{x_1, x_2\}$. Moreover, the noise operator T satisfies the following property. Let (x_1, x_2) be chosen according to the uniform distribution and (y_1, y_2) be chosen according T applied to (x_1, x_2). Then the distribution of (x_2, y_2) is uniform.*

Proof: The proof resembles the previous proof. Again there are 3 types of transitions.

- With probability β_1 we have $(x, x) \leftrightarrow (y, y)$ where $x \neq y$.
- With probability β_2 we have $(x, x) \leftrightarrow (y, z)$ where x, y, z are all different.
- With probability β_3 we have $(x, y) \leftrightarrow (z, y)$ where x, y, z are all different.

For T to be a symmetric Markov operator we require β_1, β_2 and β_3 to be non-negative and

$$2\beta_1 + 2\beta_2 = 1, \quad \beta_2 + \beta_3 = 1.$$

For the uniformity property, assume (x_1, x_2) is chosen according to the uniform distribution and (y_1, y_2) is chosen according T applied to (x_1, x_2). It is not difficult to verify that each of the nine possible settings of (x_2, y_2) is obtained with probability either $2\beta_3/9$ (if $x_2 = y_2$) or $\beta_1/9 + 2\beta_2/9$ (otherwise). Therefore, the uniformity property amounts to the equation

$$\beta_1 + 2\beta_2 = 2\beta_3.$$

It is easy to see that $\beta_2 = \beta_3 = \frac{1}{2}$ and $\beta_1 = 0$ is the solution of all equations and that the corresponding operator has $r(T) < 1$. This operator is illustrated in Figure 2(b). ■

4.3 The three reductions

The basic idea in all three reductions is to take a label-cover instance and to replace each vertex with a block of q^R vertices, corresponding to the q-ary hypercube $[q]^R$. The intended way to q-color this block is by coloring $x \in [q]^R$ according to x_i where i is the label given to this block. One can think of this coloring as an encoding of the label i. We will essentially prove that any other coloring of this block that uses relatively few colors, can be "list-decoded" into at most t labels from $\{1, \ldots, R\}$. By properly defining edges connecting these blocks, we can guarantee that the lists decoded from two blocks can be used as t-labelings for the label-cover instance.

In the rest of this section, we use the following notation. For a vector $x = (x_1, \ldots, x_n)$ and a permutation π on $\{1, \ldots, n\}$, we define $x^\pi = (x_{\pi(1)}, \ldots, x_{\pi(n)})$.

ALMOST3COLORING: Let $G = ((V, E), R, \Psi)$ be a label-cover instance as in Conjecture 4.6. For $v \in V$ write $[v]$ for a collection of vertices, one per point in $\{0, 1, 2\}^R$. Let $e = (v, w) \in E$, and let ψ be the 1↔1-constraint associated with e. By Definition 4.3 there is a permutation π such that $(a, b) \in \psi$ iff $b = \pi(a)$. We now write $[v, w]$ for the following collection of edges. We put an edge (x, y) for $x = (x_1, \ldots, x_R) \in [v]$ and $y = (y_1, \ldots, y_R) \in [w]$ iff

$$\forall i \in \{1, \ldots, R\}, \quad T\left(x_i \leftrightarrow y_{\pi(i)}\right) \neq 0$$

where T is the noise operator from Lemma 4.9. In other words, x is adjacent to y whenever

$$T^{\otimes R}\left(x \leftrightarrow y^{\pi}\right) = \prod_{i=1}^{R} T\left(x_i \leftrightarrow y_{\pi(i)}\right) \neq 0.$$

The reduction outputs the graph $[G] = ([V], [E])$ where $[V]$ is the disjoint union of all blocks $[v]$ and $[E]$ is the disjoint union of all collections of edges $[v, w]$.

APPROXCOLORING$(4, Q)$: This reduction is nearly identical to the one above, with the following changes:

- The starting point of the reduction is an instance $G = ((V, E), 2R, \Psi)$ as in Conjecture 4.7.
- Each vertex v is replaced by a copy of $\{0, 1, 2, 3\}^{2R}$ (which we still denote $[v]$).
- For every $(v, w) \in E$, let ψ be the 2↔2-constraint associated with e. By Definition 4.4 there are two permutations π_1, π_2 such that $(a, b) \in \psi$ iff $(\pi_1^{-1}(a), \pi_2^{-1}(b)) \in$ 2↔2. We now write $[v, w]$ for the following collection of edges. We put an edge (x, y) for $x = (x_1, \ldots, x_{2R}) \in [v]$ and $y = (y_1, \ldots, y_{2R}) \in [w]$ if

  $$\forall i \in \{1, \ldots, R\}, \quad T((x_{\pi_1(2i-1)}, x_{\pi_1(2i)}) \leftrightarrow (y_{\pi_2(2i-1)}, y_{\pi_2(2i)})) \neq 0$$

  where T is the noise operator from Lemma 4.10. Equivalently, we put an edge if $T^{\otimes R}(\overline{x^{\pi_1}} \leftrightarrow \overline{y^{\pi_2}}) \neq 0$.

As before, the reduction outputs the graph $[G] = ([V], [E])$ where $[V]$ is the union of all blocks $[v]$ and $[E]$ is the union of collection of the edges $[v, w]$.

APPROXCOLORING$(3, Q)$: Here again the reduction is nearly identical to the above, with the following changes:

- The starting point of the reduction is an instance of label-cover, as in Conjecture 4.8.
- Each vertex v is replaced by a copy of $\{0, 1, 2\}^{2R}$ (which we again denote $[v]$).
- For every $(v, w) \in E$, let π_1, π_2 be the permutations associated with the constraint, as in Definition 4.5. Define a collection $[v, w]$ of edges, by including the edge $(x, y) \in [v] \times [w]$ iff

  $$\forall i \in \{1, \ldots, R\}, \quad T((x_{\pi_1(2i-1)}, x_{\pi_1(2i)}) \leftrightarrow (y_{\pi_2(2i-1)}, y_{\pi_2(2i)})) \neq 0$$

  where T is the noise operator from Lemma 4.11. As before, this condition can be written as $T^{\otimes R}(\overline{x^{\pi_1}} \leftrightarrow \overline{y^{\pi_2}}) \neq 0$.

As before, we look at the coloring problem of the graph $[G] = ([V], [E])$ where $[V]$ is the union of all blocks $[v]$ and $[E]$ is the union of collection of the edges $[v, w]$.

4.4 Completeness of the three reductions

ALMOST3COLORING: If $\mathsf{isat}(G) \geq 1 - \varepsilon$, then there is some $S \subseteq V$ of size $(1-\varepsilon)|V|$ and a labeling $\ell : S \to R$ that satisfies all of the constraints induced by S. We 3-color all of the vertices in $\cup_{v \in S}[v]$ as follows. Let $c : \cup_{v\in S}[v] \to \{0,1,2\}$ be defined as follows. For every $v \in S$, the color of $x = (x_1, \ldots, x_R) \in \{0,1,2\}^R = [v]$ is defined to be $c(x) := x_i$, where $i = \ell(v) \in \{1, \ldots, R\}$.

To see that c is a legal coloring on $\cup_{v\in S}[v]$, observe that if $x \in [v]$ and $y \in [w]$ share the same color, then $x_i = y_j$ for $i = \ell(v)$ and $j = \ell(w)$. Since ℓ satisfies every constraint induced by S, it follows that if (v, w) is a constraint with an associated permutation π, then $j = \pi(i)$. Since $T(z \leftrightarrow z) = 0$ for all $z \in \{0,1,2\}$, there is no edge between x and y.

APPROXCOLORING$(4, Q)$: Let $\ell : V \to \{1, \ldots, 2R\}$ be a labeling that satisfies all the constraints in G. We define a legal 4-coloring $c : [V] \to \{0,1,2,3\}$ as follows. For a vertex $x = (x_1, \ldots, x_{2R}) \in \{0,1,2,3\}^{2R} = [v]$ set $c(x) := x_i$, where $i = \ell(v) \in \{1, \ldots, 2R\}$.

To see that c is a legal coloring, fix any $2{\leftrightarrow}2$ constraint $(v, w) \in E$ and let π_1, π_2 be the permutations associated with it. Let $i = \ell(v)$ and $j = \ell(w)$, so by the assumption on ℓ we have that $(\pi_1^{-1}(i), \pi_2^{-1}(j)) \in 2{\leftrightarrow}2$. In other words there is some $k \in \{1, \ldots, R\}$ such that $i \in \{\pi_1(2k-1), \pi_1(2k)\}$ and $j \in \{\pi_2(2k-1), \pi_2(2k)\}$. If $x \in [v]$ and $y \in [w]$ share the same color, then $x_i = c(x) = c(y) = y_j$. Since

$$x_i \in \left\{x^{\pi_1}_{2k-1}, x^{\pi_1}_{2k}\right\} \quad \text{and} \quad y_j \in \left\{y^{\pi_2}_{2k-1}, y^{\pi_2}_{2k}\right\}$$

we have that the above sets intersect. This, by Lemma 4.10, implies that $T^{\otimes R}(\overline{x^{\pi_1}} \leftrightarrow \overline{y^{\pi_2}}) = 0$. So the vertices x, y cannot be adjacent, hence the coloring is legal.

APPROXCOLORING$(3, Q)$: Here the argument is nearly identical to the above. Let $\ell : V \to \{1, \ldots, 2R\}$ be a labeling that satisfies all of the constraints in G. We define a legal 3-coloring $c : [V] \to \{0,1,2\}$ like before: $c(x) := x_i$, where $i = \ell(v) \in \{1, \ldots, 2R\}$. To see that c is a legal coloring, fix any edge $(v, w) \in E$ and let π_1, π_2 be the permutations associated with the $\bowtie$-constraint. Let $i = \ell(v)$ and $j = \ell(w)$, so by the assumption on ℓ we have that $(\pi_1^{-1}(i), \pi_2^{-1}(j)) \in \bowtie$. In other words there is some $k \in \{1, \ldots, R\}$ such that $i \in \{\pi_1(2k-1), \pi_1(2k)\}$ and $j \in \{\pi_2(2k-1), \pi_2(2k)\}$ and not both $i = \pi_1(2k)$ and $j = \pi_2(2k)$. Assume, without loss of generality, that $i = \pi_1(2k-1)$, so $x_i = x^{\pi_1}_{2k-1}$ and $y_j \in \left\{y^{\pi_2}_{2k-1}, y^{\pi_2}_{2k}\right\}$.

If $x \in [v]$ and $y \in [w]$ share the same color, then $x_i = c(x) = c(y) = y_j$, so

$$x^{\pi_1}_{2k-1} = x_i = y_j \in \left\{y^{\pi_2}_{2k-1}, y^{\pi_2}_{2k}\right\} .$$

By Lemma 4.11 this implies $T((x^{\pi_1}_{2k-1}, x^{\pi_1}_{2k}) \leftrightarrow (y^{\pi_2}_{2k-1}, y^{\pi_2}_{2k})) = 0$, which means there is no edge between x and y.

4.5 Soundness of the three reductions

Before presenting the soundness proofs, we need the following corollary. It is simply a special case of Theorem 3.1 stated in the contrapositive, with ε playing the role of ν and μ. Here we use the fact that $\langle F_\varepsilon, U_\rho(1 - F_{1-\varepsilon})\rangle_\gamma > 0$ whenever $\varepsilon > 0$.

Corollary 4.12 *Let q be a fixed integer and let T be a symmetric Markov operator on $[q]$ such that $r(T) < 1$. Then for any $\varepsilon > 0$ there exist $\delta > 0$ and $k \in \mathbb{N}$ such that the following holds. For any $f, g : [q]^n \to [0,1]$, if $E[f] \geq \varepsilon$, $E[g] \geq \varepsilon$, and $\langle f, T^{\otimes n} g\rangle = 0$, then*

$$\exists i \in \{1, \ldots, n\}, \quad I_i^{\leq k}(f) \geq \delta \quad \textit{and} \quad I_i^{\leq k}(g) \geq \delta \,.$$

ALMOST3COLORING: We will show that if $[G]$ has an independent set $S \subseteq [V]$ of relative size $\geq 2\varepsilon$, then $\mathsf{isat}_t(G) \geq \varepsilon$ for a fixed constant $t > 0$ that depends only on ε. More explicitly, we will find a set $J \subseteq V$, and a t-labeling $L : J \to \binom{R}{\leq t}$ such that $|J| \geq \varepsilon |V|$ and L satisfies all the constraints of G induced by J. In other words, for every constraint ψ over an edge $(u, v) \in E \cap J^2$, there are values $a \in L(u)$ and $b \in L(v)$ such that $(a, b) \in \psi$.

Let J be the set of all vertices $v \in V$ such that the fraction of vertices belonging to S in $[v]$ is at least ε. Then, since $|S| \geq 2\varepsilon\,|[V]|$, Markov's inequality implies $|J| \geq \varepsilon\,|V|$.

For each $v \in J$ let $f_v : \{0,1,2\}^R \to \{0,1\}$ be the characteristic function of S restricted to $[v]$, so $\mathbf{E}[f_v] \geq \varepsilon$. Select δ, k according to Corollary 4.12 with ε and the operator T of Lemma 4.9, and set

$$L(v) = \left\{ i \in \{1, \dots, R\} \;\middle|\; I_i^{\leq k}(f_v) \geq \delta \right\} .$$

Clearly, $|L(v)| \leq k/\delta$ because $\sum_{i=1}^{R} I_i^{\leq k}(f) \leq k$. Thus, L is a t-labeling for $t = k/\delta$. The main point to prove is that for every edge $e = (v_1, v_2) \in E \cap J^2$ induced on J, there is some $a \in L(v_1)$ and $b \in L(v_2)$ such that $(a, b) \in \psi_e$. This would imply that $\mathsf{isat}_t(G) \geq |J| / |V| \geq \varepsilon$.

Fix $(v_1, v_2) \in E \cap J^2$, and let π be the permutation associated with the $1{\leftrightarrow}1$ constraint on this edge. (It may be easier to first think of $\pi = id$.) Recall that the edges in $[v_1, v_2]$ were defined based on π, and on the noise operator T defined in Lemma 4.9. Let $f = f_{v_1}$, and define g by $g(x^\pi) = f_{v_2}(x)$. Since S is an independent set, $f(x) = f_{v_1}(x) = 1$ and $g(y^\pi) = f_{v_2}(y) = 1$ implies that x, y are not adjacent, so by construction $T^{\otimes R}(x \leftrightarrow y^\pi) = 0$. Therefore,

$$\langle f, T^{\otimes R} g\rangle = 3^{-R} \sum_x f(x) T^{\otimes R} g(x) = 3^{-R} \sum_x f(x) \sum_{y^\pi} T^{\otimes R}(x \leftrightarrow y^\pi) g(y^\pi) = \sum_{x, y^\pi} 0 = 0 .$$

Also, by assumption, $E[g] \geq \varepsilon$ and $E[f] \geq \varepsilon$. Corollary 4.12 implies that there is some index $i \in \{1, \dots, R\}$ for which both $I_i^{\leq k}(f) \geq \delta$ and $I_i^{\leq k}(g) \geq \delta$. By definition of L, $i \in L(v_1)$. Since the i-th variable in g is the $\pi(i)$-th variable in f_{v_2}, $\pi(i) \in L(v_2)$. It follows that there are values $i \in L(v_1)$ and $\pi(i) \in L(v_2)$ such that $(i, \pi(i))$ satisfies the constraint on (v_1, v_2). This means that $\mathsf{isat}_t(G) \geq |J| / |V| \geq \varepsilon$.

APPROXCOLORING$(4, Q)$: We outline the argument and emphasize only the modifications. Assume that $[G]$ contains an independent set $S \subseteq [V]$ whose relative size is at least $1/Q$ and set $\varepsilon = 1/2Q$.

- Let $f_v : \{0,1,2,3\}^{2R} \to \{0,1\}$ be the characteristic function of S in $[v]$. Define the set $J \subseteq V$ as before and for all $v \in J$, define

  $$L(v) = \left\{ i \in \{1, \dots, 2R\} \;\middle|\; I_i^{\leq 2k}(f_v) \geq \frac{\delta}{2} \right\}$$

  where k, δ are the values given by Corollary 4.12 with ε and the operator T of Lemma 4.10. As before, $|J| \geq \varepsilon\,|V|$ and $\mathbf{E}[f_v] \geq \varepsilon$ for $v \in J$. Now L is a t-labeling with $t = 4k/\delta$. Fix an edge $(v, w) \in E \cap J^2$ and let π_1, π_2 be the associated permutations. Define f, g by $f(x^{\pi_1}) {:=} f_{v_1}(x)$ and $g(y^{\pi_2}) {:=} f_{v_2}(y)$.

- Since S is an independent set, $f(x^{\pi_1}) = f_{v_1}(x) = 1$ and $g(y^{\pi_2}) = f_{v_2}(y) = 1$ implies that x, y are not adjacent, so by construction $T(x^{\pi_1} \leftrightarrow y^{\pi_2}) = 0$. Therefore, $\langle f, Tg\rangle = 0$.

- Now, recalling Definition 2.6, consider the functions $\overline{f}, \overline{g} : (\{0,1,2,3\}^2)^R \to \{0,1\}$. Applying Corollary 4.12 on $\overline{f}, \overline{g}$ we may deduce the existence of an index $i \in \{1,\ldots,R\}$ for which both $I_i^{\le k}(\overline{f}) \ge \delta$ and $I_i^{\le k}(\overline{g}) \ge \delta$. By Claim 2.7, $\delta \le I_i^{\le k}(\overline{f}) \le I_{2i-1}^{\le 2k}(f) + I_{2i}^{\le 2k}(f)$, so either $I_{2i-1}^{\le 2k}(f) \ge \delta/2$ or $I_{2i}^{\le 2k}(f) \ge \delta/2$. Since the j-th variable in f is the $\pi_1(j)$-th variable in f_{v_1}, this puts either $\pi_1(2i)$ or $\pi_1(2i-1)$ in $L(v_1)$. Similarly, at least one of $\pi_2(2i), \pi_2(2i-1)$ is in $L(v_2)$. Thus, there are $a \in L(v_1)$ and $b \in L(v_2)$ such that $(\pi_1^{-1}(a), \pi_2^{-1}(b)) \in 2{\leftrightarrow}2$ so L satisfies the constraint on (v_1, v_2).

We have shown that L satisfies every constraint induced by J, so $\mathsf{isat}_t(G) \ge \varepsilon$.

APPROXCOLORING$(3, Q)$: The argument here is similar to the previous one. The main difference is in the third step, where we replace Corollary 4.12 by the following corollary of Theorem 3.11. The corollary follows by letting ε play the role of μ and ν, and using the fact that $\langle F_\varepsilon, U_\rho(1 - F_{1-\varepsilon})\rangle_\gamma > 0$ whenever $\varepsilon > 0$.

Corollary 4.13 *Let T be the operator on $\{0,1,2\}^2$ defined in Lemma 4.11. For any $\varepsilon > 0$, there exists $\delta > 0, k \in \mathbb{N}$, such that for any functions $f, g : \{0,1,2\}^{2R} \to [0,1]$ satisfying $\mathbf{E}[f] \ge \varepsilon, \mathbf{E}[g] \ge \varepsilon$, there exists some $i \in \{1,\ldots,R\}$ such that either*

$$\min\left(I_{2i-1}^{\le k}(f), I_{2i-1}^{\le k}(g)\right) \ge \delta \quad \textit{or} \quad \min\left(I_{2i-1}^{\le k}(f), I_{2i}^{\le k}(g)\right) \ge \delta \quad \textit{or} \quad \min\left(I_{2i}^{\le k}(f), I_{2i-1}^{\le k}(g)\right) \ge \delta.$$

Now we have functions $f_v : \{0,1,2\}^{2R} \to \{0,1\}$, and J is defined as before. Define a labeling

$$L(v) = \left\{ i \in \{1,\ldots,2R\} \;\middle|\; I_i^{\le k}(f_v) \ge \delta \right\}$$

where k, δ are the values given by Corollary 4.13 with ε. Then L is a t-labeling with $t = k/\delta$.

Let us now show that L is a satisfying t-labeling. Let (v_1, v_2) be a $\rhd\!\!<$-constraint with associated permutations π_1, π_2. Define $f(x^{\pi_1}) = f_{v_1}(x), g(x^{\pi_2}) = f_{v_2}(x)$. We apply Corollary 4.13 on f, g, and obtain an index $i \in \{1,\ldots,R\}$. Since the j-th variable in f is the $\pi_1(j)$-th variable in f_{v_1}, this puts either $\pi_1(2i)$ or $\pi_1(2i-1)$ in $L(v_1)$. Similarly, at least one of $\pi_2(2i), \pi_2(2i-1)$ is in $L(v_2)$. Moreover, we are guaranteed that either $\pi_1(2i-1) \in L(v_1)$ or $\pi_2(2i-1) \in L(v_2)$. Thus, there are $a \in L(v_1)$ and $b \in L(v_2)$ such that $(\pi_1^{-1}(a), \pi_2^{-1}(b)) \in \rhd\!\!<$ so L satisfies the constraint on (v_1, v_2).

Acknowledgments

We thank Uri Zwick for his help with the literature and the anonymous referees and Igor Shinkar for their helpful comments.

References

[1] N. Alon, I. Dinur, E. Friedgut, and B. Sudakov. Graph products, fourier analysis and spectral techniques. *GAFA*, 14(5):913–940, 2004.

[2] S. Arora, E. Chlamtac, and M. Charikar. New approximation guarantee for chromatic number. In *Proc. 38th Annual ACM Symposium on Theory of computing*, pages 215–224, 2006.

[3] S. Arora and C. Lund. Hardness of approximations. In D. S. Hochbaum, editor, *Approximation Algorithms for NP-Hard Problems*, chapter 10. PWS, 1995.

[4] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *J. of the ACM*, 45(3):501–555, 1998.

[5] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. *J. of the ACM*, 45(1):70–122, 1998.

[6] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs, and nonapproximability—towards tight results. *SIAM J. Comput.*, 27(3):804–915, 1998.

[7] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[8] A. Blum. New approximation algorithms for graph coloring. *J. of the ACM*, 41(3):470–516, 1994.

[9] A. Blum and D. Karger. An $\tilde{O}(n^{3/14})$-coloring algorithm for 3-colorable graphs. *Inform. Process. Lett.*, 61(1):49–53, 1997.

[10] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete*, 70(1):1–13, 1985.

[11] J. Bourgain. An appendix to *Sharp thresholds of graph properties, and the k-SAT problem*, by E. Friedgut. *J. American Math. Soc.*, 12(4):1017–1054, 1999.

[12] J. Bourgain. On the distribution of the Fourier spectrum of boolean functions. *Israel Journal of Mathematics*, 131:269–276, 2002.

[13] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for unique games. In *Proc. 38th Annual ACM Symposium on Theory of Computing*, pages 205–214, 2006.

[14] E. Chlamtac, K. Makarychev, and Y. Makarychev. How to play unique games using embeddings. In *Proc. 47th IEEE Symp. on Foundations of Computer Science*, pages 687–696, 2006.

[15] I. Dinur, E. Friedgut, and O. Regev. Independent sets in graph powers are almost contained in juntas. *GAFA*, 18(1):77–97, 2008.

[16] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1):439–485, 2005.

[17] U. Feige, M. Langberg, and G. Schechtman. Graphs with tiny vector chromatic numbers and huge chromatic numbers. *SIAM Journal on Computing*, 33(6):1338–1368, 2004.

[18] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain.

[19] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124:2993–3002, 1996.

[20] M. Fürer. Improved hardness results for approximating the chromatic number. In *Proc. 36th IEEE Symp. on Foundations of Computer Science*, pages 414–421, 1995.

[21] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. of the ACM*, 42:1115–1145, 1995.

[22] A. Gupta and K. Talwar. Approximating unique games. In *Proc. 17th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 99–106, 2006.

[23] V. Guruswami and S. Khanna. On the hardness of 4-coloring a 3-colorable graph. *SIAM Journal on Discrete Mathematics*, 18(1):30–40, 2004.

[24] E. Halperin, R. Nathaniel, and U. Zwick. Coloring k-colorable graphs using relatively small palettes. *J. Algorithms*, 45(1):72–90, 2002.

[25] J. Håstad. Some optimal inapproximability results. *J. of the ACM*, 48(4):798–859, 2001.

[26] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[27] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *J. of the ACM*, 45(2):246–265, 1998.

[28] S. Khanna, N. Linial, and S. Safra. On the hardness of approximating the chromatic number. *Combinatorica*, 20(3):393–415, 2000.

[29] S. Khot. Improved inapproximability results for MaxClique, chromatic number and approximate graph coloring. In *Proc. 42nd IEEE Symp. on Foundations of Computer Science*, pages 600–609, 2001.

[30] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. of 34th Annual ACM Symposium on Theory of computing*, pages 767–775, 2002.

[31] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.

[32] S. Khot and O. Regev. Vertex cover might be hard to approximate to within $2 - \varepsilon$. *Journal of Computer and System Sciences (JCSS)*, 74(3):335–349, 2008.

[33] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*. To appear. arXiv:math/0503503v2.

[34] M. Talagrand. On Russo's approximate 0-1 law. *Annals of Probability*, 22:1576–1587, 1994.

[35] M. Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.

[36] L. Trevisan. Approximation algorithms for unique games. In *Proc. 46th IEEE Symp. on Foundations of Computer Science*, pages 197–205, 2005.

[37] A. Wigderson. Improving the performance for approximate graph coloring. *J. of the ACM*, 30:729–735, 4 1983.

A Comparison with Khot's Conjectures

Let us first state Khot's original conjectures. For $d \geq 1$, an instance of the weighted bipartite d-to-1 label cover problem is given by a tuple $\Phi = (X, Y, \Psi, W)$. We often refer to variables in X as *left* variables and to variables in Y as *right* variables. The set Ψ consists of one d-to-1 relation ψ_{xy}

for each $x \in X$ and $y \in Y$. More precisely, $\psi_{xy} \subseteq \{1, \ldots, R\} \times \{1, \ldots, R/d\}$ is such that for any $b \in \{1, \ldots, R/d\}$ there are precisely d elements $a \in \{1, \ldots, R\}$ such that $(a, b) \in \psi_{xy}$. The set W includes a non-negative weight $w_{xy} \geq 0$ for each $x \in X$, $y \in Y$. We denote by $w(\Phi, x)$ the sum $\sum_{y \in Y} w_{xy}$ and by $w(\Phi)$ the sum $\sum_{x \in X, y \in Y} w_{xy}$. A *labeling* is a function L mapping X to $\{1, \ldots, R\}$ and Y to $\{1, \ldots, R/d\}$. A constraint ψ_{xy} is *satisfied* by a labeling L if $(L(x), L(y)) \in \psi_{xy}$. Also, for a labeling L, the weight of satisfied constraints, denoted by $w_L(\Phi)$, is $\sum w_{xy}$ where the sum is taken over all $x \in X$ and $y \in Y$ such that ψ_{xy} is satisfied by L. Similarly, we define $w_L(\Phi, x)$ as $\sum w_{xy}$ where the sum is now taken over all $y \in Y$ such that ψ_{xy} is satisfied by L. The following conjectures were presented in [30].

Conjecture A.1 (Bipartite 1-to-1 Conjecture) *For any $\zeta, \gamma > 0$ there exists a constant R such that the following is NP-hard. Given a 1-to-1 label cover instance Φ with label set $\{1, \ldots, R\}$ and $w(\Phi) = 1$ distinguish between the case where there exists a labeling L such that $w_L(\Phi) \geq 1 - \zeta$ and the case where for any labeling L, $w_L(\Phi) \leq \gamma$.*

In the following conjecture, d is any fixed integer greater than 1.

Conjecture A.2 (Bipartite d-to-1 Conjecture) *For any $\gamma > 0$ there exists a constant R such that the following is NP-hard. Given a bipartite d-to-1 label cover instance Φ with label sets $\{1, \ldots, R\}, \{1, \ldots, R/d\}$ and $w(\Phi) = 1$ distinguish between the case where there exists a labeling L such that $w_L(\Phi) = 1$ and the case where for any labeling L, $w_L(\Phi) \leq \gamma$.*

The theorem we prove in this section is the following.

Theorem A.3 *Conjecture 4.6 follows from Conjecture A.1 and Conjecture 4.7 follows from Conjecture A.2 for $d = 2$.*[2]

The first part of the theorem was already proven in [32], and the second part is proven similarly. For completeness, we include here the entire proof of the theorem.

The proof follows by combining Lemmas A.4, A.5, A.7, and A.9. Each lemma presents an elementary transformation between variants of the label cover problem. The first transformation modifies a bipartite label cover instance so that all X variables have the same weight. When we say below that Φ' has the same type of constraints as Φ we mean that the transformation only duplicates existing constraints and hence if Φ consists of d-to-1 constraints for some $d \geq 1$, then so does Φ'.

Lemma A.4 *There exists an efficient procedure that given a weighted bipartite label cover instance $\Phi = (X, Y, \Psi, W)$ with $w(\Phi) = 1$ and a constant ℓ, outputs a weighted bipartite label cover instance $\Phi' = (X', Y, \Psi', W')$ on the same label sets and with the same type of constraints with the following properties:*

- *For all $x \in X'$, $w(\Phi', x) = 1$.*
- *For any $\zeta \geq 0$, if there exists a labeling L to Φ such that $w_L(\Phi) \geq 1 - \zeta$ then there exists a labeling L' to Φ' in which $1 - \sqrt{(1 + \frac{1}{\ell - 1})\zeta}$ of the variables x in X' satisfy that $w_{L'}(\Phi', x) \geq 1 - \sqrt{(1 + \frac{1}{\ell - 1})\zeta}$. In particular, if there exists a labeling L such that $w_L(\Phi) = 1$ then there exists a labeling L' in which all variables satisfy $w_{L'}(\Phi', x) = 1$.*

[2] We in fact show that for any $d \geq 2$, the natural extension of Conjecture 4.7 to d-to-d constraints follows from Conjecture A.2 with the same value of d.

- *For any $\beta, \gamma > 0$, if there exists a labeling L' to Φ' in which β of the variables x in X' satisfy $w_{L'}(\Phi', x) \geq \gamma$, then there exists a labeling L to Φ such that $w_L(\Phi) \geq (1 - \frac{1}{\ell})\beta\gamma$.*

Proof: Given Φ as above, we define $\Phi' = (X', Y, \Psi', W')$ as follows. The set X' includes $k(x)$ copies of each $x \in X$, $x^{(1)}, \dots, x^{(k(x))}$ where $k(x)$ is defined as $\lfloor \ell \cdot |X| \cdot w(\Phi, x) \rfloor$. For every $x \in X$, $y \in Y$ and $i \in \{1, \dots, k(x)\}$ we define $\psi'_{x^{(i)}y}$ as ψ_{xy} and the weight $w'_{x^{(i)}y}$ as $w_{xy}/w(\Phi, x)$. Notice that $w(\Phi', x) = 1$ for all $x \in X'$ and that $(\ell - 1)|X| \leq |X'| \leq \ell|X|$. Moreover, for any $x \in X$, $y \in Y$, the total weight of constraints created from ψ_{xy} is $k(x)w_{xy}/w(\Phi, x) \leq \ell|X|w_{xy}$.

We now prove the second property. Given a labeling L to Φ that satisfies constraints of weight at least $1-\zeta$, consider the labeling L' defined by $L'(x^{(i)}) = L(x)$ and $L'(y) = L(y)$. By the property mentioned above, the total weight of unsatisfied constraints in Φ' is at most $\ell|X|\zeta$. Since the total weight in Φ' is at least $(\ell - 1)|X|$, we obtain that the fraction of unsatisfied constraints is at most $(1 + \frac{1}{\ell-1})\zeta$. Hence, by a Markov argument, we obtain that for at least $1 - \sqrt{(1 + \frac{1}{\ell-1})\zeta}$ of the X' variables $w_{L'}(\Phi', x) \geq 1 - \sqrt{(1 + \frac{1}{\ell-1})\zeta}$.

We now prove the third property. Assume we are given a labeling L' to Φ' for which β of the variables have $w_{L'}(\Phi', x) \geq \gamma$. Without loss of generality we can assume that for every $x \in X$, the labeling $L'(x^{(i)})$ is the same for all i. This holds since the constraints between $x^{(i)}$ and the Y variables are the same for all $i \in \{1, \dots, k(x)\}$. We define the labeling L as $L(x) = L'(x^{(1)})$. The weight of constraints satisfied by L is:

$$\begin{aligned}
\sum_{x \in X} w_L(\Phi, x) &\geq \frac{1}{\ell|X|} \sum_{x \in X} k(x) \cdot w_L(\Phi, x)/w(\Phi, x) \\
&= \frac{1}{\ell|X|} \sum_{x \in X'} w_{L'}(\Phi', x) \\
&\geq \frac{1}{\ell|X|} \beta |X'| \gamma \geq \Big(1 - \frac{1}{\ell}\Big)\beta\gamma
\end{aligned}$$

where the first inequality follows from the definition of $k(x)$. ■

The second transformation creates an *unweighted* label cover instance. Such an instance is given by a tuple $\Phi = (X, Y, \Psi, E)$. The multiset E includes pairs $(x, y) \in X \times Y$ and we can think of (X, Y, E) as a bipartite graph (possibly with parallel edges). For each $e \in E$, Ψ includes a constraint, as before. The instances created by this transformation are left-regular, in the sense that the number of constraints $(x, y) \in E$ incident to each $x \in X$ is the same.

Lemma A.5 *There exists an efficient procedure that given a weighted bipartite label cover instance $\Phi = (X, Y, \Psi, W)$ with $w(\Phi, x) = 1$ for all $x \in X$ and a constant ℓ, outputs an unweighted bipartite label cover instance $\Phi' = (X, Y, \Psi', E')$ on the same label sets and with the same type of constraints with the following properties:*

- *All left degrees are equal to $\alpha = \ell|Y|$.*
- *For any $\beta, \zeta > 0$, if there exists a labeling L to Φ such that $w_L(\Phi, x) \geq 1 - \zeta$ for at least $1 - \beta$ of the variables in X, then there exists a labeling L' to Φ' in which for at least $1 - \beta$ of the variables in X, at least $1 - \zeta - 1/\ell$ of their incident constraints are satisfied. Moreover, if there exists a labeling L such that $w_L(\Phi, x) = 1$ for all x then there exists a labeling L' to Φ' that satisfies all constraints.*

- *For any $\beta, \gamma > 0$, if there exists a labeling L' to Φ' in which β of the variables in X have γ of their incident constraints satisfied, then there exists a labeling L to Φ such that for β of the variables in X, $w_L(\Phi, x) > \gamma - 1/\ell$.*

Proof: We define the instance $\Phi' = (X, Y, \Psi', E')$ as follows. For each $x \in X$, choose some $y_0(x) \in Y$ such that $w_{xy_0(x)} > 0$. For every $x \in X$, $y \neq y_0(x)$, E' contains $\lfloor \alpha w_{xy} \rfloor$ edges from x to y associated with the constraint ψ_{xy}. Moreover, for every $x \in X$, E' contains $\alpha - \sum_{y \in Y \setminus \{y_0(x)\}} \lfloor \alpha w_{xy} \rfloor$ edges from x to $y_0(x)$ associated with the constraint $\psi_{xy_0(x)}$. Notice that all left degrees are equal to α. Moreover, for any $x, y \neq y_0(x)$, we have that the number of edges between x and y is at most αw_{xy} and the number of edges from x to $y_0(x)$ is at most $\alpha w_{xy_0(x)} + |Y| = \alpha(w_{xy_0(x)} + 1/\ell)$.

Consider a labeling L to Φ and let $x \in X$ be such that $w_L(\Phi, x) > 1 - \zeta$. Then, in Φ', the same labeling satisfies that the number of incident constraints to x that are satisfied is at least $(1 - \zeta - 1/\ell)\alpha$. Moreover, if $w_L(\Phi, x) = 1$ then all its incident constraints in Φ' are satisfied (this uses that $w_{xy_0(x)} > 0$). Finally, consider a labeling L' to Φ' and let $x \in X$ have γ of its incident constraints satisfied. Then, $w_{L'}(\Phi, x) > \gamma - \frac{1}{\ell}$. ■

In the third lemma we modify a left-regular unweighted label cover instance so that it has the following property: if there exists a labeling to the original instance that for many variables satisfies many of their incident constraints, then the resulting instance has a labeling that for many variables satisfies *all* their incident constraints. But first, we prove a combinatorial claim.

Claim A.6 *For any integers ℓ, d, R and real $0 < \gamma < \frac{1}{\ell^2 d}$, let $\mathcal{F} \subseteq P(\{1, \ldots, R\})$ be a multiset containing subsets of $\{1, \ldots, R\}$ each of size at most d with the property that no element $i \in \{1, \ldots, R\}$ is contained in more than γ fraction of the sets in $\mathcal{F}$. Then, the probability that a sequence of sets $F_1, F_2, \ldots, F_\ell$ chosen uniformly from $\mathcal{F}$ (with repetitions) is pairwise disjoint is at least $1 - \ell^2 d\gamma$.*

Proof: Note that by the union bound it suffices to prove that $\Pr[F_1 \cap F_2 \neq \emptyset] \leq d\gamma$. This follows by fixing F_1 and using the union bound again:

$$\Pr[F_1 \cap F_2 \neq \emptyset] \leq \sum_{x \in F_1} \Pr[x \in F_2] \leq d\gamma.$$

■

Lemma A.7 *There exists an efficient procedure that given an unweighted bipartite d-to-1 label cover instance $\Phi = (X, Y, \Psi, E)$ with all left-degrees equal to some α, and a constant ℓ, outputs an unweighted bipartite d-to-1 label cover instance $\Phi' = (X', Y, \Psi', E')$ on the same label sets with the following properties:*

- *All left degrees are equal to ℓ.*
- *For any $\beta, \zeta \geq 0$, if there exists a labeling L to Φ such that for at least $1-\beta$ of the variables in X $1-\zeta$ of their incident constraints are satisfied, then there exists a labeling L' to Φ' in which $(1-\zeta)^\ell(1-\beta)$ of the X' variables have all their ℓ constraints satisfied. In particular, if there exists a labeling L to Φ that satisfies all constraints then there exists a labeling L' to Φ' that satisfies all constraints.*
- *For any $\beta > 0$, $0 < \gamma < \frac{1}{\ell^2 d}$, if in any labeling L to Φ at most β of the variables have γ of their incident constraints satisfied, then in any labeling L' to Φ', the fraction of satisfied constraints is at most $\beta + \frac{1}{\ell} + (1-\beta)\ell^2 d\gamma$.*

Proof: We define $\Phi' = (X', Y, \Psi', E')$ as follows. For each $x \in X$, consider its neighbors $(y_1, \ldots, y_\alpha)$ listed with multiplicities. For each sequence $(y_{i_1}, \ldots, y_{i_\ell})$ where $i_1, \ldots, i_\ell \in \{1, \ldots, \alpha\}$ we create a variable in X'. This variable is connected to $y_{i_1}, \ldots, y_{i_\ell}$ with the same constraints as x, namely $\psi_{xy_{i_1}}, \ldots, \psi_{xy_{i_\ell}}$. Notice that the total number of variables created from each $x \in X$ is α^ℓ. Hence, $|X'| = \alpha^\ell |X|$.

We now prove the second property. Assume that L is a labeling to Φ such that for at least $1-\beta$ of the variables in X, $1-\zeta$ of their incident constraints are satisfied. Let L' be the labeling to Φ' assigning to each of the variables created from $x \in X$ the value $L(x)$ and for each $y \in Y$ the value $L(y)$. Consider a variable $x \in X$ that has $1-\zeta$ of its incident constraints satisfied and let Y_x denote the set of variables $y \in Y$ such that ψ_{xy} is satisfied. Then among the variables in X' created from x, the number of variables that are connected only to variables in Y_x is at least $\alpha^\ell(1-\zeta)^\ell$. Therefore, the total number of variables all of whose constraints are satisfied by L' is at least

$$\alpha^\ell (1-\zeta)^\ell (1-\beta)|X| = (1-\zeta)^\ell (1-\beta)|X'|.$$

We now prove the third property. Assume that in any labeling L to Φ at most β of the X variables have γ of their incident constraints satisfied. Let L' be an arbitrary labeling to Φ'. For each $x \in X$ define $\mathcal{F}_x \subseteq P(\{1, \ldots, R\})$ as the multiset that contains for each constraint incident to x the set of labels to x that, together with the labeling to the Y variables given by L', satisfy this constraint. So $\mathcal{F}_x$ contains α sets, each of size d. Moreover, our assumption above implies that for at least $1-\beta$ of the variables $x \in X$, no element $i \in \{1, \ldots, R\}$ is contained in more than γ fraction of the sets in $\mathcal{F}_x$. By Claim A.6, for such x, at least $1 - \ell^2 d\gamma$ fraction of the variables in X' created from x have the property that it is impossible to satisfy more than one of their incident constraints simultaneously. Hence, the number of constraints in Φ' satisfied by L' is at most

$$\begin{aligned}
&\alpha^\ell \cdot \beta \cdot |X| \cdot \ell + \alpha^\ell (1-\beta)|X| \Big((1-\ell^2 d\gamma) + (\ell^2 d\gamma) \cdot \ell \Big) \\
&= |X'| \left(\beta\ell + (1-\beta)(1-\ell^2 d\gamma) + (1-\beta)(\ell^2 d\gamma)\ell \right) \\
&\le |E'| \left(\beta + \frac{1}{\ell} + (1-\beta)\ell^2 d\gamma \right).
\end{aligned}$$

■

The last lemma transforms a bipartite label cover into a non-bipartite label cover. This transformation no longer preserves the constraint type: d-to-1 constraints become d-to-d constraints. We first prove a simple combinatorial claim.

Claim A.8 *Let $A_1, \ldots, A_N$ be pairwise intersecting sets of size at most T. Then there exists an element contained in at least N/T of the sets.*

Proof: All sets intersect A_1 in at least one element. Since $|A_1| \le T$, there exists an element of A_1 contained in at least N/T of the sets. ■

For the following lemma, recall from Definition 4.2 that a t-labeling labels each variable with a set of at most t labels. Recall also that a constraint on x, y is satisfied by a t-labeling L if there are labels $a \in L(x)$ and $b \in L(y)$ such that (a, b) satisfies the constraint.

Lemma A.9 *There exists an efficient procedure that given an unweighted bipartite d-to-1 label cover instance $\Phi = (X, Y, \Psi, E)$ on label sets $\{1, \ldots, R\}, \{1, \ldots, R/d\}$, with all left-degrees equal to some ℓ, outputs an unweighted d-to-d label cover instance $\Phi' = (X, \Psi', E')$ on label set $\{1, \ldots, R\}$ with the following properties:*

- *For any $\beta \geq 0$, if there exists a labeling L to Φ in which $1-\beta$ of the X variables have all their ℓ incident constraints satisfied, then there exists a labeling to Φ' and a set of $1-\beta$ of the variables of X such that all the constraints between them are satisfied. In particular, if there exists a labeling L to Φ that satisfies all constraints then there exists a labeling L' to Φ' that satisfies all constraints.*

- *For any $\beta > 0$ and integer t, if there exists a t-labeling L' to Φ' and a set of β variables of X such that all the constraints between them are satisfied, then there exists a labeling L to Φ that satisfies at least β/t^2 of the constraints.*

Proof: For each pair of constraints $(x_1, y), (x_2, y) \in E$ that share a Y variable we add one constraint $(x_1, x_2) \in E'$. This constraint is satisfied when there exists a labeling to y that agrees with the labeling to x_1 and x_2. More precisely,

$$\psi'_{x_1x_2} = \Big\{(a_1, a_2) \in \{1, \ldots, R\} \times \{1, \ldots, R\} \;\Big|\; \exists b \in \{1, \ldots, R/d\}\ (a_1, b) \in \psi_{x_1y} \wedge (a_2, b) \in \psi_{x_2y}\Big\}.$$

Notice that if the constraints in Ψ are d-to-1 then the constraints in Ψ' are d-to-d.

We now prove the first property. Let L be a labeling to Φ and let $C \subseteq X$ be of size $|C| \geq (1-\beta)|X|$ such that all constraints incident to variables in C are satisfied by L. Consider the labeling L' to Φ' given by $L'(x) = L(x)$. Then, we claim that L' satisfies all the constraints in Φ' between variables of C. Indeed, take any constraint between two variables $x_1, x_2 \in C$. Assume the constraint is created as a result of some $y \in Y$. Then, since $(L(x_1), L(y)) \in \psi_{x_1y}$ and $(L(x_2), L(y)) \in \psi_{x_2y}$, we also have $(L(x_1), L(x_2)) \in \psi'_{x_1x_2}$.

It remains to prove the second property. Let L' be a t-labeling to Φ' and let $C \subseteq X$ be a set of variables of size $|C| \geq \beta|X|$ with the property that any constraint between variables of C is satisfied by L'. We first define a t-labeling L'' to Φ as follows. For each $x \in X$, we define $L''(x) = L(x)$. For each $y \in Y$, we define $L''(y) \in \{1, \ldots, R/d\}$ as the label that maximizes the number of satisfied constraints between C and y. We claim that for each $y \in Y$, L'' satisfies at least $1/t$ of the constraints between C and y. Indeed, for each constraint between C and y consider the set of labels to y that satisfy it. These sets are pairwise intersecting since all constraints in Φ' between variables of C are satisfied by L'. Moreover, since Φ is a d-to-1 label cover, these sets are of size at most t. Claim A.8 asserts the existence of a labeling to y that satisfies at least $1/t$ of the constraints between C and y. Since at least β of the constraints in Φ are incident to C, we obtain that L'' satisfies at least β/t of the constraints in Φ.

To complete the proof, we define a labeling L to Φ by $L(y) = L''(y)$ and $L(x)$ chosen uniformly from $L''(x)$. Since $|L''(x)| \leq t$ for all x, the expected number of satisfied constraints is at least β/t^2, as required. ■

B Tightness of Theorem 1.2

Let v be an eigenvector of T whose eigenvalue λ satisfies $|\lambda| = \rho$, normalized so that $\sum_{i=1}^{q} v_i^2/q = 1$. Assume that $\lambda > 0$ (the proof for the case $\lambda < 0$ is similar). For any $n \geq 1$, define two indicator functions

$$f(x_1, \ldots, x_n) = \begin{cases} 1, & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{x_i} < \mu \\ 0, & \text{o.w.} \end{cases} \quad , \quad g(x_1, \ldots, x_n) = \begin{cases} 1, & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{x_i} < \nu \\ 0, & \text{o.w.} \end{cases}$$

where μ and ν are some arbitrary constants. The functions f and g have all of their influences of order $n^{-1/2}$. Moreover, by the central limit theorem, if $(x_1, \ldots, x_n)$ is chosen uniformly and

$(y_1, \ldots, y_n)$ is obtained from it through $T^{\otimes n}$, then the joint distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{x_i}$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{y_i}$ converges to that of two standard normal variables with correlation $\lambda = \rho$. From this it follows that $E[f]$, $E[g]$, and $\langle f, T^{\otimes n} g \rangle$ converge to μ, ν, and $\langle F_\mu, U_\rho F_\nu \rangle_\gamma$ respectively. A similar argument holds for the lower bound.