On the Optimality of Semidefinite Relaxations for Average-Case and Generalized Constraint Satisfaction

Boaz Barak* Guy Kindler† David Steurer‡

November 21, 2012

Abstract

This work studies several questions about the optimality of semidefinite programming (SDP) for constraint satisfaction problems (CSPs).

First we propose the hypothesis that the well known Basic SDP relaxation is actually optimal for random instances of constraint satisfaction problems for *every* predicate. This unifies several conjectures proposed in the past, and suggests a unifying principle for the average-case complexity of CSPs. We provide several types of indirect evidence for the truth of this hypothesis, and also show that it (and its variants) imply several conjectures in hardness of approximation including polynomial factor hardness for the densest k subgraph problem and hard instances for the Sliding Scale Conjecture of Bellare, Goldwasser, Lund and Russell (1993).

Second, we observe that for every predicate P, the basic SDP relaxation achieves the same approximation guarantee for the CSP for P and for a more general problem (involving not just Boolean but constrained vector assignments), which we call the *Generalized CSP* for P. Raghavendra (2008) showed that it is Unique Games-hard to approximate the CSP for P better than this guarantee. We show that it is **NP**-hard to approximate the Generalized CSP for P better than this guarantee.

*Microsoft Research New England, Cambridge, MA.

†The Hebrew University of Jerusalem.

‡Microsoft Research New England, Cambridge, MA.

Contents

1 Introduction

Some of the most appealing results in algorithms and computational complexity are *meta algorithms* or *meta reductions* that characterize the complexity not just of a single problem but a whole family of them. Some canonical examples for meta algorithms include the Ellipsoid algorithm, that applies for all convex problems admitting efficient separation and membership oracles [GLS81], and Robertson and Seymour's algorithm deciding every minor-closed graph property in polynomial time [RS95, RS04]. One example for a meta reduction is the work of Lewis and Yannakakis [Yan78, LY80] showing that for large a class of properties, finding the largest subgraph satisfying the property is **NP** hard.

Perhaps the nicest case is when we have a complimentary meta-algorithm and meta-reduction, thus obtaining a *meta characterization* that determines exactly the complexity of any computational problem inside some family. The *Dichotomy Conjecture* of Feder and Vardi [FV98] is aimed at achieving exactly such a characterization for the determining the satisfiability of constraint satisfaction problems (CSPs), and much exciting progress has been made on it [Sch78, HN90, Bul02, BK09]. In the context of approximation algorithms, Raghavendra [Rag08] gave, assuming Khot's Unique Games Conjecture (UGC) [Kho02b], a meta-characterization of the approximation threshold of all CSPs. In particular he showed that for every CSP it is Unique Games-hard to beat the approximation guarantee achieved by a simple semidefinite programming relaxation known as Basic Sdp.

Recent works, however, have raised some doubts concerning the UGC. In particular, while the UGC asserts that a certain computational problem known as Unique Games is **NP**-hard, it was shown that this problem can be solved in subexponential time [ABS10], and in fact there is a candidate algorithm that might solve it much faster than that [BBH+12]. Indeed, showing a distribution of plausibly hard instances for Unique Games appears to be a challenging problem [AKK+08, BHHS11, KMM11]. These works form a strong motivation for basing optimality results on conjectures that are weaker than, or at least different from, the UGC.[1] Moreover, as all of the above meta-characterization results are in the context of *worst-case* complexity, and the question of whether such characterizations can be proven, or are even true, in the context of *average-case complexity*, has not been addressed to our knowledge.

This paper shows some results in the above mentioned directions. Specifically, we study optimality results in two settings:

Average-case complexity We put forward the hypothesis that among polynomial time approximation algorithms, Basic Sdp is optimal for *random CSPs*. We show that this conjecture (and its extentions) would have some interesting consequences for hardness of approximation, and give some indirect evidence supporting it. While proving this hypothesis is beyond current

[1] We note such optimality results will have to differ from those of Raghavendra's theorem since his meta-characterization also implies the UGC.

techniques in complexity, there are many plausible ways to *refute* it, and we argue that studying this and similar conjectures is a good approach to make progress on average-case complexity.

Generalized CSPs We show that, assuming $\mathbf{P} \neq \mathbf{NP}$, Basic Sdp is in fact optimal for a class of problem that we call *generalized CSPs*. At a very high level, given a predicate P, the *generalized CSP* corresponding to P involves applying P not just on variables and their negations but an arbitrary linear functions of the variables (which are allowed now to be not just bits but certain real vectors). The definition of generalized CSP is motivated by the observation that Basic Sdp does not really distinguish between generalized CSPs and real ones. Thus in particular for every predicate P, Basic Sdp yields the same approximation on the generalized CSP for P that it yields for the CSP of P. (For example, Basic Sdp yields the same factor 0.87... approximation for Generalized Max-Cut as the one it achieves for the standard Max-Cut, and while it is an open question whether this factor can be improved for Max-Cut, our results imply that it is **NP**-hard to beat for the generalized version.)

We now give more details on our results.

1.1 Average-case complexity

For a predicate $P\colon \{\pm 1\}^K \to \{0,1\}$, an instance $\mathfrak{I}$ of CSP(P) is a set of K-tuples of literals over the variables $x_1, \ldots, x_n$. The *value* of $\mathfrak{I}$, denoted val($\mathfrak{I}$), is the maximum, over all assignments to the variables, of the fraction of tuples that satisfy P. We say that an algorithm A is a *relaxation* for CSP(P) if $\mathrm{val}(\mathfrak{I}) \leqslant A(\mathfrak{I})$ for all instances $\mathfrak{I}$ of CSP(P). Some examples for relaxations are obtained by linear and convex programs for CSP(P), including the canonical algorithm Basic Sdp considered by [Rag08], but in general a relaxation is any algorithm that always upper bounds the true value. A *random instance* of CSP(P) is obtained by selecting m random K-tuples of literals for some parameter $m = m(n)$.[2] The following conjecture states that Basic Sdp is essentially the optimal polynomial time upper-certificate on *random* instances of CSP(P).

RCSP Hypothesis (SDP optimality hypothesis for refuting CSPs, informal)**.** *Let $P\colon \{\pm 1\}^K \to \{0,1\}$ be a predicate and let $\varepsilon > 0$. For every polynomial time relaxation A for* CSP(P)*, it holds that for random instance $\mathfrak{I}$ of* CSP(P)*, almost always $A(\mathfrak{I}) \geqslant$ Basic Sdp$(\mathfrak{I}) - \varepsilon$.*

(The R in RCSP can stand for either "refutation" or "random".)

While it is clear what a *disproof* of the RCSP Hypothesis would consist of, given the state of art in computational complexity it seems unlikely that it will

[2] Unless specified otherwise, m will equal cn for some sufficiently large constant c. Also note that our definition considers tuples of *literals* (i.e., variables or their negations) and not just un-negated variables.

be *proven* (or even derived from well studied worst-case conjectures such as **P** ≠ **NP** [BT06]) in the foreseeable future. Nevertheless, there are still several approaches to obtain *evidence* for its veracity:

– *(Reductions)* It may be possible to derive the RCSP Hypothesis from a more well-studied *average case* conjectures/hypothesis such as Feige's Hypothesis [Fei02] about refuting random 3SAT, or conjectures about the hardness of learning parity with noise [GKL88, BFKL93].

  We show a partial result in this direction, observing that Feige's results can be restated as showing that his hypothesis implies the RCSP Hypothesis for the case of predicates of arity at most 3.

– *(Integrality gaps)* If the RCSP Hypothesis is true, then in particular stronger mathematical relaxations than Basic Sdp such as its extension to various hierarchies will not give better results than Basic Sdp on random instances. Confirming this prediction can be thought of as evidence for the RCSP Hypothesis.

  We observe that the results of Benabbas, Georgiou, Magen, and Tulsiani [BGMT12] can be adapted to show this prediction for the relaxation obtained by augmenting Basic Sdp with $\Omega(n)$ rounds of the Sherali-Adams hierarchy, while Tulsiani's work [Tul09] shows that for particular families of predicates, the same holds for $\Omega(n)$ rounds of the Lasserre hierarchy.

– *(Worst-case hardness of approximation results)* As was observed in the past, average-case assumptions such as the RCSP Hypothesis imply worst-case hardness of approximation results. Confirming such predictions again can be thought of as providing evidence for the hypothesis.

  One can view the classic result of Håstad [Hås01] as some hardness of approximation on this form, for the case of predicates such as k-XOR and k-SAT. Our results on generalized CSP's (see below) are also in this flavor.

Beside being an attractive hypothesis in its own right, we also show that the RCSP Hypothesis implies the conjecture that predicates supporting pairwise independent distributions are approximation resistant.[3] This was shown based on the Unique Games Conjecture by [AM09], so our results gives an alternative basis for that statement. We also consider generalizations of the RCSP Hypothesis to larger alphabet and broader sets of parameters, and show that these generalizations yield hard instances for the "Sliding Scale" conjecture of [BGLR93] (including its projection game variant [Mos11]), as well as instances for the problem of obtaining a polynomial approximation for the densest κ-subgraph problem (see also [BCC+10, AAM+11]).

[3] We say that a distribution D over a product set S^k is *pairwise independent* if for every $i \neq j$ in $[k]$, the projection $D_{i,j}$ is the uniform distribution over $S \times S$. This notion is also sometimes known as "pairwise uniform".

Remark 1.1. Traditionally in computational complexity it is not very common to consider "meta-conjectures" such as the RCSP Hypothesis, that posit the hardness of a large family of problems. Complexity theorists naturally prefer the more pleasing "meta-reductions" showing that hardness of a single problem implies the hardness of a large family. This approach has been particularly successful in the study of *worst-case* complexity, where researchers have been able to base the difficulty of an astounding number of problems based on the single assumption that 3SAT can't be solved efficiently. However, such reductions have been much more rare in average case complexity. Indeed, it seems that reductions are inherently problematic in this context, as they tend to use various "gadgets" and other transformations that result in the output of the reduction not being the natural distribution over the instances of the target problem. Thus in the context of average-case complexity we propose that such meta-conjectures, positing the optimality of a certain type of algorithm for a large family of problems, may be the right way to go forward in establishing at least some more intuition as to where lies the boundary between hardness and easiness.

On a broader level, there are two general approaches in theoretical Computer Science to cope with the fact that many our basic questions remain unsolved. One very useful approach is to try to make the weakest possible conjectures, with the hope of eventually getting rid of them altogether. The other approach, which the RCSP Hypothesis belongs to, is to start with the most simple and broad hypotheses possible and to see what they imply, thereby giving a "large target" for attempts at refutations. Finding refutations for such hypotheses can obviously be very instructive, directing research to more plausible directions. On the other hand, if many refutations attempts fail, this can also be quite useful, as it may reveal a general principle of nature which could be true even if a proof for it is very far from our reach. Finally, we note that as bold as the RCSP Hypothesis seems, it can be strengthened even further, both quantitatively and qualitatively; see Section 3.4.

1.2 Generalized CSP

Raghavendra [Rag08] showed how an efficient approximation can be attained for every CSP (contraint satisfaction problem), using an associated SDP which we call the Basic Sdp. In this paper we define a generalized version of each such CSP, for which the same Basic Sdp can still be applied, achieving the same approximation factor. For this generalized CSP, we show that beating the performance of the Basic Sdp relaxation is **NP**-hard. This is in contrast to Raghavendra's result [Rag08] mentioned above, who showed a much stronger conclusion —tightness of Basic Sdp on *standard* CSPs— but under a much stronger assumption, namely the unique games conjecture. Our work draws on the techniques of [KKMO04, MOO05, Rag08] developed in the context of the unique games conjecture, and thus shows that even if this conjecture turns out to be false, both the techniques and at least some parts of the results obtained by works relying on it can still be salvaged. Adapting the above techniques to

the case of **NP**-hardness is similar both in flavor and in techniques to the work of [GRSW12], where UGC-hardness was replaced by **NP**-hardness for a class of geometric problems rather than CSPs.[4]

Let $P : \{\pm 1\}^K \to \{0,1\}$ be a predicate. Our starting point is the observation that the Basic Sdp algorithm for CSP(P) actually obtains the same guarantee to a more general problem that we call the *generalized CSP* for P, denoted GCSP(P). The definition of GCSP(P) is a bit subtle (see more below), but roughly speaking it involves placing additional constraints on the assignment that can be enforced by semidefinite programming. Given the way we tailor the definition to Basic Sdp, it is perhaps not surprising that this relaxation achieves the exact same approximation for GCSP(P) as CSP(P). What is more surprising is that now we are able to prove that this approximation guarantee is *optimal*, and cannot be improved upon by any efficient algorithm unless **P** = **NP**. (See §4 details and Section 4.5 for proofs.)

Theorem 1.2. *For every $c, s \in \mathbb{R}$ with $0 < s < c < 1$ and predicates $P\colon \{\pm 1\}^K \to \{0,1\}$, the promise problem (c,s)-Gap* GCSP(P) *is either **NP**-hard or solvable in polynomial time (by* Basic Sdp*). Furthermore,* Basic Sdp *solves (c,s)-Gap* GCSP(P) *if and only if it solves (non-generalized) (c,s)-Gap* CSP(P)*.*

Defining generalized CSP. We now give more details about the definition of generalized CSPs. The definition is closely tied to the Basic Sdp algorithm (see Section 2), and so we start by reviewing it. Let $P : \{\pm 1\}^K \to \{0,1\}$ be a predicate, and let $\mathfrak{I}$ be an instance of CSP(P) over n variables. An *assignment* to $\mathfrak{I}$ is a Boolean vector $x \in \{\pm 1\}^n$ and the *value* of x, denoted $\mathrm{val}(x)$, is the fraction of the constraints it satisfies. The Basic Sdp algorithm optimizes over a larger convex set $\mathcal{X}$, (which we call the set of *SDP assignments*) which embeds inside it all the Boolean vectors. For an SDP assignment X, we denote by $\mathrm{sval}(X)$ be the value that the SDP outputs on X. One can define a canonical rounding algorithm that maps every SDP assignment $X \in \mathcal{X}$ into a Boolean assignment $x \in \{\pm 1\}^n$ [Rag08], and we define $\mathrm{rval}(X)$ to be the value of this Boolean assignment. (If X was already *integral*, i.e., (isomorphic to) a single Boolean assignment x then the rounding algorithm just returns x, and hence $\mathrm{rval}(X) = \mathrm{sval}(X) = \mathrm{val}(x)$.) So, an equivalent way to phrase the problem CSP(P) is that we want to find the maximum of $\mathrm{rval}(X)$ over $X \in \mathcal{X}$, while the Basic Sdp algorithm can find the maximum of $\mathrm{sval}(X)$ over the same set. That is, instead of considering the goal of CSP(P) as finding a good Boolean assignment, we think of it as finding a "well roundable" / near-integral SDP assignment.

For every $X \in \mathcal{X}$, we have that

$$\alpha\, \mathrm{sval}(X) \leqslant \mathrm{rval}(X) \leqslant \mathrm{sval}(X)\ , \tag{1.1}$$

[4] We note that in [Rag08], the term "generalized CSP" was used for constraints which are real-valued rather than Boolean valued. Our notion is different, as we consider Boolean valued constraints where the variables which appear in them are real-valued.

where α is called the *integrality gap ratio* of the program. Clearly, Basic Sdp is an α-approximation for CSP(P). But because (1.1) holds pointwise for every $X \in \mathcal{X}$, one can see that even if we placed additional linear and semidefinite constraints on X, thus restricting the set $\mathcal{X}$ into some subset $\mathcal{X}'$, then Basic Sdp still yields an α approximation for this more general program. This general problem is the generalized CSP problem. That is, an instance to GCSP(P) consists of a set of K-tuples of literals, as well as some additional constraints that we place on the SDP solutions X, and the goal is to find a "well roundable" / near-integral SDP solution meeting those constraints.

Are generalized CSP natural? We think that the mere fact that GCSP(P) has the same approximation as CSP(P) but it is **NP**-hard to beat already shows that this problem is non-trivial. But one may wonder if the generalized CSP problem is also interesting on its own right. While the definition of the problem is indeed closely tied to the Basic Sdp algorithm, we believe that it also of some independent interest. One way to think about SDP assignment is as distributions over Boolean assignment, while the additional constraints posit certain correlations between the random variables comprising this distribution (e.g., a typical such constraint for generalized Max Cut will require that if i and j are in the left side of the cut, then k is also likely to be on that side). One can imagine that solving a CSP with such weak side constraints could be useful in some applications.[5]

Evidence for the Unique Games Conjecture? The UGC implies Basic Sdp is optimal for Generalized CSPs. Our results confirm this prediction of this conjecture, and hence can be interpreted as giving evidence for the truth of the UGC. However, there is also evidence that Generalized CSPs are strictly harder computationally than (non-generalized) CSPs: Our **NP**-hardness reductions from smooth Label Cover to Generalized CSPs have linear running time. There is some evidence that smooth Label Cover (with the parameters that we need) is exponentially hard (see Remark 3.8), which would imply that it is exponentially hard to beat the approximation guarantee of Basic Sdp for Generalized CSPs. In contrast, for some (non-generalized) CSPs, it is known how to beat the approximation of Basic Sdp in subexponential time [ABS10].

1.3 Our techniques

In this section we outline some of the techniques used in the proof of Theorem 1.2—optimality of Basic Sdp for generalized CSP's— which is our most technical result. We focus on the case of the *generalized* Max Cut problem, which we also deal with more formally later, in Subsection 4.3. Going from generalized Max Cut to generalized CSPs of other types follows quite closely the transition made in [Rag08] from (non-generalized) Max Cut to other (non-generalized) CSPs.

[5] Note that the constraints are weak rather than precise linear relations between the variables, since the rounding function can introduce some errors.

Both our construction and the proof are generalizations of the Unique Games-hardness of the Max Cut problem [KKMO07, MOO10], and readers familiar with that result as well as [Rag08] should see the similarities. We also use techniques similar to the ones from [GRSW12].

We begin by describing the (c, s)-gap generalized Max Cut problem. As described above, this is the problem of finding an SDP solution which satisfies not just the constraints that result from the Max Cut instance but also additional linear constraints, and has a high expected value when rounded. The goal in generalized Max Cut is to find such an SDP solution which is rounded to the best possible integral solution by the standard rounding procedure. Specifically, the (c, s)-gap problem is that of distinguishing between the case where there exists an SDP solution that is rounded to a integral solution with value c, and the case where no SDP solution which satisfies the additional linear constraints get expected integral value greater than s.

The Basic Sdp algorithm for max-cut is the well known Goemans-Williamson SDP [GW95], which can be thought of as maximizing the sum of $\mathbb{E}\frac{1-y_i y_j}{2}$ over all edges (i, j) in the input graph, where (if the graph has m vertices) $y_1, \ldots, y_m$ are Gaussian random variables satisfying $\mathbb{E}|y_i|^2 = 1$ for all i.[6] The rounding function of this SDP is simply obtained by taking the sign of the Gaussians (which would of course not lose anything if they were degenerate random variables with variance zero), and hence one way to phrase the (non generalized) Max Cut problem is that the aim is to find Gaussian variables $y_1, \ldots, y_m$ maximizing the sum of $\mathbb{E}\frac{1-\mathrm{sign}(y_i)\,\mathrm{sign}(y_j)}{2}$ over all edges (i, j).

For the (c, s)-gap generalized Max Cut problem, we want to optimize the SDP solution under additional constraints, and in particular we allow the instance to require some linear constraints on the correlations between these Gaussians. A natural way in which this can be achieved is that the instance specifies that each gaussian y_i must be a linear combination of some other gaussian variables. Another way to say the same thing is that an instance of generalized Max Cut will consists of a graph on m vertices where each vertex is identified with some linear function from $\mathbb{R}^n$ to $\mathbb{R}$. Thus, now we aim to find Gaussians $x_1, \ldots, x_n$ that maximize the sum of $\mathbb{E}\frac{1-\mathrm{sign}(a(x))\,\mathrm{sign}(b(x))}{2}$ over all edges (a, b). This leads us to the following definition:

Definition 1.3. (c, s)-gap generalized Max Cut is the following promise problem: Given a distribution over pairs of linear functions (a, b) over $\mathbb{R}^n$, distinguish between the following cases:

YES: there exists $x \in \{0,1\}^n$ such that $a(x) \in \{\pm 1\}$ for all linear forms a appearing in $\mathfrak{I}$ and $\mathbb{P}_{(a,b)\sim\mathfrak{I}}\{a(x) \neq b(x)\} \geqslant c$,

[6] The above description is equivalent to the more common description of the program with vectors, since every vector v_i can be associated with a Gaussian random variable y_i obtained by taking $\langle v_i, g\rangle$ for a random Gaussian, and visa versa. (To generate non-symmetric Gaussian variables one can introduce an additional vector v_0 and identify it with the Gaussian variable that is identically 1.)

NO: for every n-dimensional random vector x whose coordinates are jointly Gaussian, which satisfies $\mathbb{E}_x\, xx^T \succcurlyeq 0$ and $\mathbb{E}\, a(x)^2 = 1$ for all linear forms a appearing in $\mathfrak{I}$, we have $\mathbb{E}_x\, \mathbb{P}_{(a,b)\sim\mathfrak{I}} \{\mathrm{sign}(a(x)) \neq \mathrm{sign}(b(x))\} \leqslant s$.[7]

To make the connection to other CSPs more apparent, we state our result for generalized Max Cut in a generic way as follows.

Theorem 1.4. *Let $c, s \in \mathbb{R}$ with $0 < s < c < 1$ and $c \geqslant 0.845$.[8] If there exists a (c,s) SDP-gap instance (with respect to the Goemans-Williamson* Max Cut *SDP) for (non-generalized)* Max Cut*, then (c,s)-gap generalized* Max Cut *is **NP**-hard. On the other hand, if (c,s)-gap* Max Cut *is solvable by the Goemans-Williamson* Max Cut *algorithm, then so is the (c,s) generalized* Max Cut *gap problem.*

As the generalized Max Cut problem is obtained by linear constraints on an SDP solution, it's not hard to show that for pairs (c,s) where the Goemans-Williamson algorithm solved Max Cut it will also solve generalized Max Cut. Let us therefore see how a (c,s) integrality gap for Max Cut translates to an **NP**-hardness result.

The projection test. In [Rag08], it is explained how to transform an integrality gap instance of a CSP into a long-code test for a function $f : \{-1,1\}^R \to \{0,1\}$. For the specific case of Max Cut, this reduction was shown by [KKMO04] and works as follows: Pick a random edge (i,j) from the instance, and choose ρ to be the correlation between the Gaussian vectors x_i and x_j supplied by the SDP solution for the instance. Then pick $a, b \in \{-1,1\}^k$ to be random Boolean vectors with correlation ρ between the coordinates a_i and b_i for each i, and accepts if $f(a) \neq f(b)$. To get **NP**-hardnes, though, it is not enough to consider a dictatorship test – we need a test for projections, namely one that works for two functions on *different domains* $f : \{-1,1\}^{d\cdot R} \to \{0,1\}$ and $g : \{-1,1\}^R \to \{0,1\}$, with a given projection function $\pi : [d \cdot R] \to [R]$. Oversimplifying, the test should verify that f is an i dictatorship for some i, that g is a j dictatorship, and that $\pi(i) = j$. However, because the functions have so different domains, natural attempt at a test makes the distribution of the larger input a be very far from the uniform distribution, and instead depend strongly on the particular projection π. On an intuitive level this is a bad thing, since the function f might be very far from a dictator globally, but agree with a dictator on this particular distribution. Indeed, no such test is known.

Our solution to this problem, which follows a similar approach to [GRSW12], is to use the additional linear constraints to bypass this issue. Roughly speaking, one can define a simple linear map L_π that takes a function $f : \{\pm 1\}^{dR} \to \mathbb{R}$ and maps it into a function $f' : \{\pm 1\}^R \to \mathbb{R}$ such that if f is an i dictatorship then f' is

[7] Note that the requirement that $\mathbb{E}_x\, xx^T \succcurlyeq 0$ is another linear constraint on the correlation matrix. For showing the problem is easy we can generalize to all such linear constraints, but our hardness proof only requires this one.

[8] In this range of c the Goemans-Williamson rounding is optimal for Basic Sdp of Max Cut

a $\pi(i)$ dictatorship. Indeed, the map L_π is defined by simply requiring it to satisfy this property for the dictatorships, and extending it linearly to all other functions. In our problem we can enforce this kind of linear relation between f and f', and so at least are in the position where we can syntactically apply the same test on f' and g, which are now on the same domain.

However, there is a host of technical issues that must be overcome to show the test actually works. To prove soundness for this test, we need to show a that (very roughly speaking) the map L_π has the property that if $L_\pi f$ is close to a dictator (or more accurately, has an influential coordinate) then so is the original function f. It's very easy to show that this is not generally true, and that for every π one can come up with a function f that is far from a dictator but $L_\pi f$ is very close to one. Indeed, one example would be to fix some $j_0 \in [R]$ and take f to be the sum over all $i \in \pi^{-1}(j_0)$ of the i^{th} dictator. However, here is where (as in [GRSW12]) the technical property of *smoothness* comes to the rescue. This property means that for every function f which is a sum of dictatorships, with non-negative weight w_i for the i^{th} dictatorship, if we look at a "typical" projection π that will arise in our reduction, then the the sum of the weights in each of the "buckets" $\pi^{-1}(j)$ will be roughly the same. Now some additional complications arise since in our cases these "weights" may be negative, and in fact are not even numbers but actually vectors. Here we use some properties of moments of Gaussians, together with the condition on non-negative correlations, to argue that cancelations will not be an issue. The above is clearly only a very rough outline of the proof, which is given in Section 4.

2 Preliminaries

A *Gaussian vector* (over some $\mathbb{R}$-vector space) is a vector-valued random variable with coordinates drawn from a joint Gaussian distribution. We do not assume symmetry. In particular, the mean of a Gaussian vector is not necessarily the 0 vector.

Boolean CSPs. Let $P\colon \{\pm 1\}^K \to \{0,1\}$ be a Boolean predicate. A CSP(P) instance $\mathfrak{I}$ with variable set $V = [n]$ is specified by a collection of ordered K-tuples of literals. We identity K-tuple of literals with functions $S\colon \{\pm 1\}^V \to \{\pm 1\}^K$ such that each output bit $S(x)_i$ depends on at most one coordinate of $x \in \{\pm 1\}^V$. The value of an assignment $x \in \{\pm 1\}^V$ is defined as $\mathfrak{I}(x) = \mathbb{E}_{S\in\mathfrak{I}}\, P(S(x))$. We denote the maximum value of an assignment by $\mathrm{opt}(\mathfrak{I}) = \max_{x\in\{\pm 1\}^V} \mathfrak{I}(x)$.

Basic SDP Relaxation. For every CSP(P) instance $\mathfrak{I}$, we associate a semidefinite program Basic Sdp($\mathfrak{I}$). A *solution* for Basic Sdp($\mathfrak{I}$) consists of an n-dimensional Gaussian vector x and a collection of distributions $\{\mu_S\}_{S\in\mathfrak{I}}$ over $\{\pm 1\}^K$. The solution is *feasible* if for every $S \in \mathfrak{I}$, the first two moments of the distributions $S(x)$ and μ_S match. (Here, we extend S to $\mathbb{R}^n$ in a multilinear way.) Concretely,

every K-variate quadratic[9] polynomial Q satisfies $\mathbb{E}_x Q(S(x)) = \mathbb{E}_{\mu_S} Q$. The *value* of such a solution is defined as $\mathbb{E}_{S\in\mathfrak{I}} \mathbb{E}_{\mu_S} P \circ S$. We denote the maximum value of Basic Sdp($\mathfrak{I}$) by opt(Basic Sdp($\mathfrak{I}$)). For brevity, we will often drop opt$(\cdot)$ and use Basic Sdp($\mathfrak{I}$) also to refer to the optimal value of the SDP relaxation.

Since a feasible Gaussian vector x is supposed to model (a distribution over) points in $\{\pm 1\}^n$, its diagonal second moments are $\mathbb{E}_x x_1^2 = \ldots = \mathbb{E}_x x_n^2 = 1$. We refer to this constraint as *normalization condition*.

For $0 < s < c < 1$, we say that Basic Sdp achieves a (c, s)-approximation for CSP(P) if every CSP(P) instance $\mathfrak{I}$ with Basic Sdp($\mathfrak{I}$) $\geqslant c$ also satisfies opt($\mathfrak{I}$) $\geqslant s$.

3 The SDP optimality hypothesis for random CSPs

In this section we formulate more precisely the RCSP Hypothesis, and show some of the evidence supporting it, as well as its implications. We also consider some extensions of the hypothesis to higher alphabet, and non-constant parameters, and discuss their implications.

3.1 Random CSPs

For every n, m and $P\colon \{\pm 1\}^K \to \{0,1\}$, we let $\mathrm{CSP}_{n,m}(P)$ denote the distribution over instances of CSP(P) obtained by taking m random K-tuples over n variables. We let $\underline{\mathrm{val}}(P)$ denote the expected value a random assignment gives to a CSP(P) instance, namely $\underline{\mathrm{val}}(P) = \mathbb{E}\, P(U_K)$ where U_K is the uniform distribution over $\{\pm 1\}^K$. We note the simple fact that the optimal value of random instances of CSP(P) is asymptotically equal to the value obtained by a random assignment:

Lemma 3.1. *For every K, ε, δ there is $c = c(K, \varepsilon, \delta)$ such that for all $P\colon \{\pm 1\}^K \to \{0,1\}$ and $m \geqslant cn$, if $\mathfrak{I}$ is chosen at random from $\mathrm{CSP}_{n,m}(P)$ then a.s. $\underline{\mathrm{val}}(\mathfrak{I}) \leqslant \mathrm{val}(\mathfrak{I}) \leqslant \underline{\mathrm{val}}(\mathfrak{I}) + \varepsilon$.*

Proof. Since val($\mathfrak{I}$) is the maximum value over all assignments, clearly $\underline{\mathrm{val}}(\mathfrak{I}) \leqslant \mathrm{val}(\mathfrak{I})$ for all $\mathfrak{I}$. Now for the other direction, we claim that for sufficiently large c, with high probability over the choice of the m tuples $S_1, \ldots, S_m$, we will have that for *every* $x \in \{\pm 1\}^n$, the distribution $D = D_{\{S_i\},x}$ obtained by taking $i \leftarrow_R [m]$ and outputting $(y_1, \ldots, y_K) = S_i(x)$ is ε-close to the uniform distribution in statistical distance, thus concluding the proof. This will follow from a simple concentration+union bound argument. Using the Vazirani XOR Lemma, it suffices to show that for every fixed vector $a = (a_1, \ldots, a_K) \in \{0,1\}^K$ and fixed x, the probability over the choice of the S_i's that **(*)** $|\sum_{i\in[m]} \prod_{j=1}^K S_i(x)_j^{a_j} - m/2| \geqslant m\varepsilon/2^K$ is less than $\delta 2^{-n}$. But because the S_i's are chosen independently at random, and in particular the negation pattern added to them is random, for every x the random variables $b_1, \ldots, b_m$ where $b_i = \prod_{j=1}^K S_i(x)_j^{a_j}$ are unbiased and independent, and hence the probability of **(*)** is at most $\exp(-\varepsilon^2 m/4)$ which can be made sufficiently small by taking c large enough. □

[9] Here, quadratic means degree *at most* 2.

3.2 Characterizing the SDP value of CSP(P)

We now make a simple, but very useful observation, that has been implicitly made in other works as well [AM09]— the value that Basic Sdp returns on random instances of CSP(P) is equal to the maximum of $\mathbb{E}\, P(D)$ taken over all pairwise independent distributions D over $\{\pm 1\}^K$.

Theorem 3.2. *Let $P\colon \{\pm 1\}^K \to \{0,1\}$, and let $\overline{\mathrm{val}}(P)$ equal the maximum of $\mathbb{E}\, P(D)$ over all pairwise independent distributions D over $\{\pm 1\}^K$. Then:*

1. *For every instance $\mathfrak{I}$ of* CSP(P), Basic Sdp$(\mathfrak{I}) \geqslant \overline{\mathrm{val}}(P)$.
2. *For every $\varepsilon, \delta > 0$, there is $c = c(K, \varepsilon, \delta)$ such that if $\mathfrak{I}$ is a random instance of* CSP(P) *with n variables and at least cn tuples, then with probability $1 - \delta$,* Basic Sdp$(\mathfrak{I}) \leqslant \overline{\mathrm{val}}(P) + \varepsilon$.

Proof. Let $\mathfrak{I}$ be a CSP(P) instance, consisting of m tuples of literals $S_1, \ldots, S_m$ over the variables $x_1, \ldots, x_n$. A Basic Sdp *assignment* for $\mathfrak{I}$ consists of n random variables $X_1, \ldots, X_n$ (that can be thought of as real vectors over some sample space Ω), and m distributions $\mu_1, \ldots, \mu_m$, each over $\{0,1\}^K$ (and hence can be given by 2^K numbers in $[0,1]$ summing to 1). The distribution μ_i is "supposed" to correspond to the restriction of the random variables $X_1, \ldots, X_n$ to the literals in S_i, but the only constraints that it needs to satisfy is that it agrees with this distribution up to the first two moments. That is, if we let $(Y_1, \ldots, Y_K)$ be the distribution $S_i(X_1, \ldots, X_n)$, and let $(Z_1, \ldots, Z_K)$ be distributed according to μ_i, then for all $j, j' \in [K]$, $\mathbb{E}\, Y_j Y_{j'} = \mathbb{E}\, Z_j Z_{j'}$, and $\mathbb{E}\, Y_j = \mathbb{E}\, Z_j$. The value of the assignment is $\mathbb{E}_i\, \mathbb{E}\, P(\mu_i)$, and the value Basic Sdp$(\mathfrak{I})$ is the maximum of this value over all assignments.

1. Let D be the pairwise independent distribution that achieves $\overline{\mathrm{val}}(P)$, letting $X_1, \ldots, X_n$ be orthogonal unit vectors, and let $\mu_i = D$ for all i, then we get a valid Basic Sdp assignment that demonstrates that Basic Sdp$(P) \geqslant \overline{\mathrm{val}}(P)$.
2. Suppose that there is a Basic Sdp assignment $X_1, \ldots, X_n$, $\mu_1, \ldots, \mu_m$ that achieves $\overline{\mathrm{val}}(P) + \varepsilon$. Let μ be the distribution over $\{\pm 1\}^K$ obtained by taking $i \in_r [m]$ and $(y_1, \ldots, y_K) \in_r \mu_i$. We denote the resulting distribution by $(Y_1, \ldots, Y_K)$. Then $\mathbb{E}\, P(\mu) = \overline{\mathrm{val}}(P) + \varepsilon$ which means that μ is *not* pairwise independent. In particular we get that there must be either **(i)** $j \in [K]$ such that $|\mathbb{E}\, Y_j| > \varepsilon'$ or **(ii)** some $j \neq j' \in [K]$ such that $|\mathbb{E}\, Y_j Y_{j'}| > \varepsilon'$ for some $e' = \Omega(\varepsilon/k^2) > 0$. But, if c is big enough then both **(i)** and **(ii)** are highly unlikely. First, it's easy to see for c large enough we would have that all but small fraction of the vertices, the number of times they appear in the j^{th} coordinate positively will be up to some $1 \pm \varepsilon'$ factor the same as the number of times in that coordinate negatively, thus ruling out **(i)**. Second, if we consider the 2XOR game obtained by taking for every tuple S_i the constraint that the product of the j^{th} and j'^{th} literals is 1 (or -1), then it is a

random 2XOR game, and it is known that (since the underlying graph will be an expander) the Basic Sdp value of this game will tend to 0 with c, thus ruling out **(ii)**.

□

We can now restate the RCSP Hypothesis:

RCSP Hypothesis (SDP optimality hypothesis for refuting CSP's, formal version)**.** *Let $P\colon \{\pm 1\}^K \to \{0,1\}$ be a predicate, let $\Delta > 0$ and let $m(n)$ be some function such that $m(n) \leqslant \Delta n$. Then for every polynomial time relaxation A for $\mathrm{CSP}(P)$ and $\varepsilon, \delta > 0$, we have that for all sufficiently large n*

$$\mathbb{P}[A(\mathfrak{I}) \leqslant \overline{\mathrm{val}}(\mathfrak{I}) - \varepsilon] < \delta \,,$$

where this probability is over $\mathfrak{I}$ chosen from $\mathrm{CSP}_{n,m(n)}(P)$.

Combining Lemma 3.1 with Theorem 3.2, we see that the RCSP Hypothesis implies that predicates supporting a pairwise independent distribution are *approximation resistant* in the sense of [AH12]. That is, for every predicate $P\colon \{\pm 1\}^K \to \{0,1\}$, if there exists a pairwise independent distribution D whose support is contained in $P^{-1}(1)$, then for every $\varepsilon > 0$ there is no polynomial-time algorithm B to distinguish, given an instance $\mathfrak{I}$ of $\mathrm{CSP}(P)$, between the case **YES** that $\mathrm{val}(\mathfrak{I}) = 1 - \varepsilon$ and the case **NO** that $\mathrm{val}(\mathfrak{I}) \leqslant \overline{\mathrm{val}}(\mathfrak{I}) + \varepsilon$. Indeed, if there was such an algorithm B then we could construct a relaxation algorithm A contradicting the RCSP Hypothesis by simply having $A(\mathfrak{I})$ output $1 - \varepsilon$ if $B(\mathfrak{I})$ ="NO", and $A(\mathfrak{I})$ output 1 otherwise.

3.3 Some evidence for the SDP optimality hypothesis

The RCSP Hypothesis is fairly bold, in the sense that it posits average-case hardness of a large family of problems, and so we would like to investigate whether it can actually be true. There are two types of evidence for such an hypothesis— (a) we could show that the hypothesis (or at least variants of it) is implied by seemingly weaker or more well studied conjectures, and (b) we could verify some of the *predictions* it makes, that is independently prove some the hypothesis' implications. As discussed in Section 1.1, we offer evidence of both types for this hypothesis. While we believe more investigations are merited, we believe these results do suggest that the RCSP Hypothesis may be true.

3.3.1 Relation to Feige's Hypothesis

Feige made the hypothesis [Fei02, Hypothesis 2] that for every $\varepsilon > 0$ there is no algorithm that can certify that the value of a random 3SAT instance (i.e., taken from $\mathrm{CSP}_{n,O(n)}(OR)$ in our notation) is smaller than $1 - \varepsilon$. Though he didn't phrase it in those terms, Feige's results show that his Hypothesis 2 implies the special

case of the RCSP Hypothesis for 3-ary predicates. Specifically, Theorem 2 in [Fei02] shows that under his Hypothesis 2, for every predicate $P\colon \{\pm 1\}^3 \to \{0,1\}$, it is hard, given a random $\mathfrak{I} \in \mathrm{CSP}_{n,O(n)}(P)$ to certify that $\mathrm{val}(\mathfrak{I}) \leqslant b(P)/4$, where b is the number of $x \in P^{-1}(1)$ such that $x_1x_2x_3 = +1$ (or -1, if that number is greater). However, it is not hard to show that for a 3-ary predicate P, $\overline{\mathrm{val}}(P) = b(P)/4$.

In fact, Feige's work can be used to show that for *every* non-trivial predicate $P\colon \{\pm 1\}^3 \to \{0,1\}$ (i.e., where $\underline{\mathrm{val}}(P) < \overline{\mathrm{val}}(P)$), the special case of the RCSP Hypothesis for P implies the hypothesis for all 3-ary predicates. For completeness we show a full proof of this result for the case of XOR.

Theorem 3.3. *Suppose that RCSP holds for the 3XOR predicate then it holds also for every other predicate $P\colon \{\pm 1\}^3 \to \{\pm 1\}$.*

Proof. For this proof it will be convenient to extend our notion of predicates to general functions $P\colon \{\pm 1\}^3 \to \mathbb{R}$ (and not just those that have output in $\{0,1\}$). Note that the SDP value, as well as the notions of $\underline{\mathrm{val}}, \overline{\mathrm{val}}, \mathrm{val}$ easily extend to this case. In particular, it will be convenient to define the predicate 3XOR as simply outputting the product $x_1x_2x_3$ of its inputs. (The 0/1 definition of 3XOR is $P_0(x_1,x_2,x_3) = (x_1x_2x_3 - 1)/2$, but clearly if you can refute one then you can refute the other and so the two definitions are equivalent in computational difficulty.)

Let $P\colon \{\pm 1\}^3 \to \{0,1\}$ be some predicate, and write P as a multilinear polynomial in the variables x_1, x_2, x_3 of the form $ax_1x_2x_3 + P'(x_1,x_2,x_3) + b$ where $a, b \in \mathbb{R}$ and P' has no constant term and degree at most 2. Suppose that there is an algorithm A that can certify that a random instance $\mathfrak{I}$ of $\mathrm{CSP}(P)$ has $\mathrm{val}_P(\mathfrak{I}) \leqslant \overline{\mathrm{val}}(P) - \varepsilon$ (where $\mathrm{val}_P(\mathfrak{I})$ denotes the value of the set of tuples $\mathfrak{I}$ when thought of as an instance of $\mathrm{CSP}(P)$). We'll construct an algorithm B that can certify that $\mathrm{val}_{3XOR}(\mathfrak{I}) \leqslant 1 - \varepsilon'$ for some $\varepsilon' > 0$ that tends to zero with ε. The algorithm B will first certify that $|\mathrm{val}_{P'}(\mathfrak{I})| < \varepsilon/2$ using Basic Sdp (which can be done since since for every $f : \{\pm 1\}^3 \to \mathbb{R}$ of degree at most 2 without a constant term, $\underline{\mathrm{val}}(f) = \overline{\mathrm{val}}(f) = 0$). Now this means that for every assignment x,

$$\mathop{\mathbb{E}}_{S \leftarrow_R \mathfrak{I}} P(S(x)) \in a \mathop{\mathbb{E}}_{S \leftarrow_R \mathfrak{I}} 3XOR(S(x)) + c \pm \varepsilon/2 \,.$$

Assume $a < 0$ (the other case is symmetric). Thus, if we want to certify that $\mathrm{val}_{3XOR}(\mathfrak{I}) \leqslant 1 - \varepsilon/2$, it suffices to certify that $\mathrm{val}_P(-\mathfrak{I}) \leqslant |a| + c - \varepsilon$, where $-\mathfrak{I}$ is obtained by flipping all signs in the literals of $\mathfrak{I}$. Indeed, otherwise, if $\mathrm{val}_{3XOR}(\mathfrak{I}) \geqslant 1 - \varepsilon/2$, then $\mathrm{val}_P(-\mathfrak{I})$ would have been at least $|a|(1-\varepsilon/2) + c - \varepsilon/2 \geqslant |a| + c - \varepsilon$, since it's easy to see that $c \leqslant 1$. So the result follows by showing that $\overline{\mathrm{val}}(P) \leqslant |a| + c$, which holds since for every pairwise independent distribution D over $\{\pm 1\}^3$, $EP'(D) = 0$, while $|E_{x \leftarrow_R D} x_1x_2x_3| \leqslant 1$. (In fact, $\overline{\mathrm{val}}(P) = |a| + c$, as demonstrated by the distribution $(x_1, x_2, -x_1x_2)$.) □

3.3.2 Integrality gaps

If the RCSP Hypothesis is true, then on random CSP instances, Basic Sdp can't be beat by even much more powerful algorithms. Thus, one way to get

more evidence for our hypothesis is to verify this prediction for some particular algorithms. Specifically, we will be interested in semidefinite programming *hierarchies*, which are systematic way of strengthening Basic Sdp to obtain tighter relaxations (see the survey [Lau03]). We observe that prior works, though not phrased in those terms, establish particular predictions of the RCSP Hypothesis:

Theorem 3.4 (Implicit in [BGMT12])**.** *For every $P\colon \{\pm 1\}^K \to \{0,1\}$, $\varepsilon > 0, \Delta \in \mathbb{N}$, there exists $\delta > 0$ such that for sufficiently large n, if $\mathfrak{I}$ is drawn randomly from $\mathrm{CSP}_{n,\Delta n}(P)$, with probability at least $1-\varepsilon$, the program obtained by augmenting* Basic Sdp *with δn levels of the Sherali-Adams hierarchy outputs at least $\overline{\mathrm{val}}(P) - \varepsilon$.*

Proof. Benabbas Georgiou Magen Tulsiani [BGMT12, Theorem 4.3] proved that this holds for the case that $\overline{\mathrm{val}}(P) = 1$ (i.e., when there is a pairwise independent distribution supported on $P^{-1}(1)$). However, the proof easily extends to the general case. □

For the stronger SDP hierarchy of Lasserre, this is not yet currently known. However, it is known for the special case of *subspace predicates*, which are predicates $P\colon \{\pm 1\}^K \to \{0,1\}$ such that if we identify $\{\pm 1\}^K$ with $\mathbb{F}_2^K$ in the obvious way, then P^{-1} is an affine subspace. For such a predicate, let $d(P)$ be the minimum weight of the dual subspace to $P^{-1}(1)$. In other words, this is minimum number of variables in a linear equation over $x_1, \ldots, x_K$ that is satisfied by all $x \in P^{-1}(1)$. It's not hard to see that if $d(P) \geqslant 3$ then $\overline{\mathrm{val}}(P) = 1$. Tulsiani [Tul09, Theorem 4.3] showed for such P, $\Omega(n)$ rounds of the Lasserre hierarchy give value 1 on random instances of CSP(P).

3.3.3 Hardness of approximation results

Another consequence of the RCSP Hypothesis would be some *worst-case* hardness of approximation results. Since a random instance of CSP(P) has value roughly $\underline{\mathrm{val}}(P)$, any approximation algorithm that can distinguish between instances with this value and instances with value $\overline{\mathrm{val}}(P) - o(1)$ would refute the hypothesis. Thus, another way to increase confidence in the hypothesis would be to derive these predictions based on more standard and widely believed assumptions such as **P** ≠ **NP**. We were not able to prove this result, which is of course interesting in its own right. However our results on generalized CSP can be viewed as making some progress for confirming these type of predictions.

Remark 3.5. Note that if the Unique Games Conjecture was true, then Raghavendra's Theorem of course implies that it's **NP**-hard to beat Basic Sdp. However, we don't consider this as strong evidence for the RCSP Hypothesis. First, as mentioned, it is not at all clear that the UGC is true. Second, even if true, the UGC seems not to say much about *random* or even *pseudorandom* instances, for which algorithms beating the UGC's predictions are known (indeed Basic Sdp itself can sometime work better than the worst case on such instances [AKK+08]). Furthermore, it is known that Unique Games-hardness cannot establish $\exp(\Omega(n))$

or $\exp(n^{1-o(1)})$ hardness for a problem, while the integrality gaps suggest that it may even be hard to beat Basic Sdp on random instances in subexponential time.

3.4 Extensions of the RCSP Hypothesis and their applications

While there is more evidence for "vanilla" version of the RCSP Hypothesis, there are several natural ways to quantitatively strengthen it that are not known to be false, and some of these, if true, will have interesting applications.

Super polynomial running time. While the hypothesis is stated equating "efficient" with polynomial time, it seems just as valid to assume that there is no $\exp(n^{o(1)})$-time algorithm that beats Basic Sdp. In fact, current integrality gaps support even the conjecture that beating Basic Sdp on random instances will take at least $\exp(\Omega(n))$ time, where the constant in the Ω notation may depend on the parameters (accuracy parameter ε, predicate arity parameter K, clause to variable ratio parameter Δ).

Non binary alphabet. A very natural extension is to consider predicates over alphabet $q > 2$. One way to represent them is as $P\colon R^K \to \{0,1\}$, where R is a ring of size q (or perhaps a field if q is a prime or prime power). Following prior works on integrality gaps and proof complexity [Tul09], we'll use the convention that a *literal* y_i would correspond to not just a variable x_i or its negation (which has no meaning here) but it would be a *shift* of x_i by some arbitrary $a \in R$ (where for $R = \mathbb{F}_2$, $a = 0$ corresponds to taking the variable, and $a = 1$ corresponds to negating it). This preserves the property that a random instance of CSP(P) would have value roughly $\underline{\mathrm{val}}(P)$.

Super constant parameters. While the conjecture is stated for parameters (such as accuracy ε, arity K, and clause to variable ratio Δ) that are constant, it may very well hold for values of these parameters (as well as the alphabet size parameter q mentioned above) that are *super-constant* functions of n. In particular, current knowledge seems consistent with the conjecture being true for $K, \Delta, q, 1/\varepsilon$ all being up to some value n^δ for some small $\delta > 0$. Such versions of the conjecture can be used to derive much stronger hardness of approximation results.

Remark 3.6 (*Determining thresholds for parameters*). It is an interesting question whether SDP's can be used to predict the precise hardness thresholds for these parameters. A case in point is the value of the clause to variable ratio Δ for refuting random 3XOR (or 3SAT) instances. The best polynomial-time algorithms known work for Δ that is at least (roughly) $\sqrt{n}$.[10] This is not done by Basic Sdp

[10] Interestingly, Feige, Kim and Ofek [FKO06] gave *non-deterministic* algorithm for refuting random 3XOR or 3SAT instances with $\Delta \sim n^{0.4}$. However, there is no efficient algorithm that matches their performance, and moreover their algorithm was for the case that ε is smaller than $n^{-0.2}$.

(that always outputs at least $\overline{\text{val}}(P)$ for any predicate P) but can be achieved by a polynomial-time extension of Basic Sdp. (One example is an algorithm that would work for $\Delta > \sqrt{n}$ is an SDP we call "local Lasserre" where one adds for every subset S of variables that appear in some tuple/clause, a vector that is supposed to correspond to $\prod_{i \in S} x_i$, with the appropriate consistency constraints.) Thus, one could hope to make a stronger SDP optimality conjecture that will pinpoint the right value of Δ. A more tricky question is whether SDP's can help pinpoint the minimum possible value for our "completeness parameter" ε. While setting ε to be an arbitrary small constant seems to keep the problem hard, and maybe we can even set it equal to $n^{-\delta}$ for some small $\delta > 0$, we do not know of a very principled way to hypothesize the right hardness threshold for ε. Consider again the case of 3XOR: although for $\varepsilon = 0$ the problem can be easily solved via Gaussian elimination, even very strong SDP's such as $\Omega(n)$ rounds of the Lasserre hierarchy cannot recover this algorithm, and hence do not yield a prediction on the right bound for ε. A related question is in what cases taking $\varepsilon = 0$ (i.e., "perfect completeness") still keeps the problem hard. Clearly having the predicate be *non-linear* (under any encoding of the alphabet) is a necessary requirement, but it is not clear that it is a sufficient one.

Decision version of the RCSP Hypothesis. There is also a natural way to *qualitatively* strengthen the RCSP Hypothesis. For any **NP** promise set $(\mathcal{Y}, \mathcal{N})$ (where there is always a certificate/witness to certify that $x \notin \mathcal{N}$), given distributions $\mathcal{D}_Y, \mathcal{D}_N$ over these sets, one can define three variants of average-case computational problems: **(i)** The *decision problem* is to distinguish, given an input x drawn randomly from either $\mathcal{D}_Y$ or $\mathcal{D}_N$, which is the case with success noticeably larger than 1/2. **(ii)** The *search problem* is to find, given x drawn from $\mathcal{D}_Y$ the witness that certifies that it is at least not in $\mathcal{N}$. **(iii)** The *refutation problem* is to certify, given x drawn from $\mathcal{D}_N$, that x is not in $\mathcal{Y}$, by giving an algorithm that never outputs "NO" on $x \in \mathcal{Y}$, and typically outputs "NO" on x from $\mathcal{D}_N$.

The decision problem is obviously easier than the refutation and search problem, which are in general incomparable, although in some cases search to decision reductions are also known. We formulate now a decisional version of the RCSP hypothesis, which posits that its hard to distinguish between random CSPinstances, and instances in which we "planted" a solution:

DCSP Hypothesis (SDP optimality hypothesis for deciding CSPs, informal)**.** *Let D_0 be a pairwise independent distribution over $\{\pm 1\}^K$ and let $\varepsilon, \delta > 0$ and $\Delta > 1$. Let $\text{CSP}_{n,m,k}$ be the uniform distribution of taking m random K-tuples of literals over n variables, and let $\mathcal{D}_Y(n, m, k, D_0, \varepsilon)$ be the distribution over such tuples that is output by the following process:*

1. *Pick $x_0 \leftarrow_R \{\pm 1\}^n$.*
2. *For $i = 1 \ldots m$, the i^{th} tuple S is chosen as follows. With probability ε, S is chosen to be a random K-tuple of literals. With probability $1 - \varepsilon$, we pick random $d \leftarrow_R D_0$, and pick S to be a random K-tuple of literals* conditioned on $S(x_0) = d$.

Then for every polynomial-time algorithm A, if n is sufficiently large and $m = \Delta n$ it holds that

$$\left|\mathbb{E}\, A(\mathrm{CSP}_{n,m,k}) - \mathbb{E}\, A(\mathcal{D}_Y(n, m, k, D_0, \varepsilon)\right| < \delta\ .$$

Note that the DCSP Hypothesis talks about pairwise independent distributions over $\{\pm 1\}^K$, instead of predicates. This is in some sense necessary, since naive planting of a satisfying assignment, without ensuring that the induced distribution over tuples is pairwise independent, would make it easy even for Basic Sdp to detect the difference between the NO and YES cases. We show in the appendix that the DCSP Hypothesis implies the RCSP Hypothesis (as is the expected relation between decisional and refutation assumption). We also show that if one restricts to the case of "atomic" pairwise independent distributions (i.e., those that are not convex combinations of other pairwise independent distributions), then the variant of the DCSP for such distributions is equivalent to the variant of the RCSP Hypothesis for *non-Boolean* predicates P that output a value in $[0, 1]$.

Other distributions. Another potential approach to strengthen the RCSP Hypothesis is to consider distributions other than the uniform one on instances. The integrality gap results do not need the instances to be random, but only require them to have sufficient expansion properties. However, since these integrality gaps can also apply to a distribution concentrated on a single input (that thus can never be hard), they cannot be blindly taken to be evidence of hardness. Still, while we do not explore this direction in the current work, it is possible that a much more general version of the RCSP Hypothesis holds, where every input satisfying a particular condition, randomly perturbing it would result in a distribution on hard instances. Indeed, conjectures of a somewhat similar type for Max-3XOR and learning parity with noise were raised by Alekhnovich, see [Ale03, Conjecture 2], though to our knowledge they have not received much investigation. Such an hypothesis, if true, would be an interesting dual to the notion of "smoothed complexity" [ST04]— here perturbing an input would result in a hard instance, rather than an easy one.

3.5 Applications of the RCSP Hypothesis

The RCSP Hypothesis on its own is a very interesting statement about average-case complexity, but it also has some implications for hardness of approximation, and perhaps cryptography. Here are some examples:

Max K-AND. Feige [Fei02, Hypothesis 3] hypothesized that it is hard to certify that a random Max-K-AND formula has value less than $2^{-c\sqrt{K}}$ for some $c > 0$. He showed that this hypothesis implies some hardness of approximation results for the 2-catalog problem, while [AAM+11] showed it implies hardness of approximation for the densest K subgraph problem. The RCSP Hypothesis

posits much bolder hardness of approximation for this problem, since it's easy to see that $\overline{\mathrm{val}}(K-AND)$ is equal to the maximum, over any $d \in \{\pm 1\}^K$ and pairwise distribution D over $\{\pm 1\}^K$, of the probability that $D = d$. In particular, if $K = 2^\ell - 1$ then, if we identify every $a \in [K]$ with a nonzero linear function over $\mathbb{F}_(2)^\ell$, then the pairwise independent distribution obtained by choosing $b \leftarrow_R \mathbb{F}_(2)^\ell$ and outputting the vector $(a(x))_{a\in[K]}$ demonstrates that $\overline{\mathrm{val}}(K-AND) \geqslant 2^{-\ell} = 1/(K+1)$. Thus, the RCSP Hypothesis posits that it is hard to distinguish random K-AND instances, which have value 2^{-K}, from instances with value $1/(K+1)$. These are much stronger parameters than those hypothesized by Feige (though of course, such worst-case hardness of approximation does hold if the Unique Games Conjecture is true).

Label cover, sliding scale, densest κ-subgraph. The large alphabet variant of the RCSP Hypothesis also implies hardness for random instances of the label cover problem. Recall that in this problem one is given a collection of triples (i, j, π) where $i \in [n']$, $j \in [n]$ and π is a function from $[Q]$ to $[q]$, and the value of an assignment $y \in [Q]^{n'}$ and $x \in [q]^n$ is the fraction of triples (i, j, π) such that $\pi(y_i) = x_j$. In particular, for every set of projections $\pi_1, \ldots, \pi_K$ mapping $[Q]$ to $[q]$, one can consider the predicate $P\colon [q]^K \to \{0, 1\}$ such that $P(x_1, \ldots, x_K)$ if there exists $y \in [Q]$ such that $\pi_i(y) = x_i$ for all i. A random instance of CSP(P) corresponds to a random instance of the label cover problem, and it is not hard to show that (assuming the parameters are chosen appropriately) every assignment of this label cover instance will have polynomially small value (see Lemma 3.7 below).[11] The RCSP hypothesis implies that it is hard to distinguish such instances from those where one can satisfy $1 - \varepsilon$ fraction of the triples. In particular, if we take K, q, Δ that can grow as large as n^δ for some small $\delta > 0$, we get the hardness of approximation predicted by (an imperfect completeness variant of) the "sliding scale conjecture" of Bellare et al. [BGLR93] and in fact, since the corresponding instances are projections, this also applies for the projection game variant proposed by Moshkovitz [Mos11]. Similarly, we can show that this variant of the RCSP Hypothesis also implies it's hard to approximate the densest κ-subgraph problem to some polynomial factor.

Lemma 3.7. *Let $\mathfrak{I}$ be a label cover instance chosen as above with parameters $Q, q, K, \Delta \geqslant 100$ satisfying that for some $0 < \mu < 1/(20K^{1/4})$, $Q < q^{\mu K/4}$, $2^{16\log(1/\mu)} < q$. Then with high probability, every assignment to $\mathfrak{I}$ satisfies at most 2μ fraction of the triples.*

Proof. Let us consider for a fixed assignment $x \in [q]^n$, what is the probability over the random choices of the Δn tuples (and in fact just the shifts) there exists some $y \in [Q]^{\Delta n}$ that such that for at least a μ fraction of the $i \in [\Delta n]$, there will be at least a μ fraction of the $j \in [K]$ with the constraint between y_i and the j^{th} member of the tuple satisfied. The probability that it is possible to satisfy a μ fraction of

[11] Note that this is a stronger statement than the fact that every assignment $x \in [q]^n$ will satisfy at most this fraction of the constraints corresponding to the K-sary predicate P.

any particular K-tuple is upper bounded by

$$\binom{K}{\mu k} Q q^{-\mu K} \leqslant 2^{2\log(1/\mu)\mu K} q^{-\mu K/4} \leqslant q^{-\mu K/8} ,$$

under our assumptions. Thus the probability that more than $\mu\Delta n$ of these tuples will be μ-satisfied is bounded by

$$2^{\Delta n} q^{-\mu^2 K \Delta n/8} \leqslant q^{-\sqrt{K}\Delta n} ,$$

since $\sqrt{K}\Delta \gg 1$, we can do a union bound over all q^n possible assignments. □

Remark 3.8. Because a random CSP over the alphabet $[q]$ involves random shifts, the corresponding label cover instance will be *smooth* with a smoothness parameter that tends to zero as K tends to infinity. Also, if the image of the projections $\pi_1, \ldots, \pi_K$ is a linear code over $\mathbb{F}_q^K$ with dual distance at least 3, then the results of Tulsiani [Tul09] imply that this instance is hard to certify for $\Omega(n)$ rounds of the Lasserre program, just supporting the conjecture that it is in fact *exponentially hard* to certify.

Lemma 3.9. *Suppose that the* RCSP *Hypothesis holds with* $\Delta = \text{polylog}(n)$ *and* $q, k = n^\delta$ *for some* $\delta > 0$*, then there is some* $\delta' > 0$ *such that there is no* $n^{\delta'}$ *polynomial-time algorithm for the densest* κ*-subgraph problem.*

Proof sketch. The proof closely follows the work of Bhaskara, Charikar, Guruswami, Vijayaraghavan and Zhou [BCV+12] who gave Lasserre integrality gaps for the densest κ-subgraph problem with similar parameters. Their proof was obtained in fact by reducing a random instance of a particular CSP into an instance of the densest subgraph problem (the CSP is of the same form as the one used to obtain random instances of label cover above, and the graph is essentially the label-extended graph of the label cover instance). In fact in our case the proof is easier since we don't need to carry out the completeness proof for vector assignments but only for actual assignments. They showed that for a random CSP(P) with for an appropriately chosen K-CSP P over the alphabet q, the resulting graph will have the property that every κ-sized subgraph has average degree $\tilde{O}(K/q)$, while if P had a satisfying assignment, there will be a κ-sized subgraph with average degree $\tilde{\Omega}(K)$. The only thing we need to do to complete the proof is to note that the same still holds even if P had an assignment that satisfied, say, 0.9 fraction of its clauses. □

Public key cryptography? The DCSP Hypothesis readily yields a one way function and hence private key cryptography. An intriguing question is whether it can yield stronger cryptographic primitives and in particular *public key cryptography*. This is of significance since we currently have very few candidates for public key systems, and the most well studied of them can be broken efficiently by quantum computers, and thus increasing the public key crypto "gene pool" has

been recognized as an important problem (e.g., see the discussion in [ABW10]). A public key cryptosystem based on the DCSP Hypothesis would arguably be the first scheme for which we have at least somewhat "principled" reasons for arguing about its security. Applebaum, Barak and Wigderson [ABW10] made a step in that direction by giving public key cryptosystems based on hardness of random CSPs and the hardness of planted version of the densest k subgraph problem. In fact, their assumptions can be seen as an instantiation of the RCSP Hypothesis but with very strong parameters, in particular requiring not just polynomially large parameters, but also control on the particular polynomial and relation between them. Reducing those assumptions, or at least finding principled ways to justify them, would be welcome progress.

4 Optimality of Basic SDP for Generalized CSPs

In this section we formally state and prove our harndess-of-approximation results for generalized Max Cut and generalized distribution matching, and also for any generalized CSP. The results are stated in Subsection 4.1, in Subsection 4.2 we define the label-cover problem and the generic rounding scheme of [Rag08] in a language that suits our needs. We also define there a dictatorship test and its properties. We then prove **NP**-hardness for the generalized Max-Cut and generalized distribution matching problems in Subsection 4.3 and Subsection 4.4 respectively, and in Subsection 4.5 we deal with all other generalized CSPs. The hardness proofs all rely on a theorem we prove in Subsection 4.6, which shows how to decode an assignment for a smooth label-cover instance from an assignment for its composition with a dictatorship test which has non-negligible influences, which in turn relies on an observation concerning smooth projections that appears in Subsection 4.7. Finally in Subsection 4.8 we deal with a technical issue of the sign function not having the Lipchitz property, and in Subsection 4.9 we cite the special case of the Invariance Principle that we use in Subsection 4.4.

4.1 Results

Our general approach for proving meta-characterizations of generalizations of CSPs is quite versatile. We restrict ourselves to three representative examples (listed in order of increasing generality).

Generalized Max Cut. This problem is motivated by the Basic Sdp for Max Cut and the Goemans-Williamson (GW) rounding algorithm [GW95]. A solution to the Basic Sdp for Max Cut assigns a Gaussian variable with second moment[12] 1 to every vertex of the graph. (The variables have a jointly Gaussian distribution.) The GW rounding algorithm draws a sample from this Gaussian distribution and assigns to each vertex the sign of its sampled value. This rounding algorithm

[12]Here, we mean raw moment $\mathbb{E}\, X^2$ (as opposed to central moment $\mathbb{E}\, X^2 - (\mathbb{E}\, X)^2$).

achieves an approximation factor $\alpha_{\mathrm{GW}} \approx 0.878$ in the worst case. However, for integral Basic Sdp solution (i.e., the Gaussian variables are constant), the GW rounding algorithm has no loss. Hence, the Max Cut problem is equivalent to finding a Basic Sdp solution that maximizes of the expected value of output of the GW rounding algorithm. In Generalized Max Cut, we have this same objective but with additional restrictions on the Basic Sdp solutions (satisfying certain linear inequalities). Algorithmically, it is still easy to optimize over Basic Sdp solutions that satisfy the additional restrictions. Since the GW rounding applies to any Basic Sdp solution (as opposed to just optimal ones, say), it has the same approximation guarantee for Generalized Max Cut as for (non-generalized) Max Cut. We will show that for all $\varepsilon > 0$, it is **NP**-hard to find Basic Sdp solutions with these restrictions such that the GW rounding achieves for this solution an approximation factor of at least $\alpha_{\mathrm{GW}} + \varepsilon$.

The kind of restrictions on Basic Sdp solutions we need for our hardness reductions are fairly simple: We require that the Gaussian variables for the vertices are linear forms $a(x)$ of an auxiliary Gaussian vector x with nonnegative second-moment matrix $\mathbb{E}\, xx^T \succcurlyeq 0$. At this point, it is convenient to identify the vertices of the graph with these linear forms. The following summarizes the formal statement of Generalized Max Cut.

Problem 4.1 ((c,s)-Gap Generalized Max Cut)**.** Given a collection $\mathfrak{I}$ of pairs (a,b) of linear forms on $\mathbb{R}^n$, distinguish the cases,

YES: there exists $x \in \mathbb{R}^n_+$ such that $a(x) \in \{\pm 1\}$ for all linear forms a appearing in $\mathfrak{I}$ and

$$\mathop{\mathbb{P}}_{(a,b)\sim\mathfrak{I}} \{a(x) \neq b(x)\} \geqslant c\,,$$

NO: for every n-dimensional Gaussian vector x with $\mathbb{E}_x\, xx^T \succcurlyeq 0$ and $\mathbb{E}\, a(x)^2 = 1$ for all linear forms a appearing in $\mathfrak{I}$, we have

$$\mathop{\mathbb{E}}_{x} \mathop{\mathbb{P}}_{(a,b)\sim\mathfrak{I}} \{\mathrm{sign}\circ a(x) \neq \mathrm{sign}\circ b(x)\} \leqslant s\,.$$

We can think of the promise in the **YES** case as an integral Basic Sdp solution with value c (for $\mathfrak{I}$ as a graph on linear forms) generated by the nonnegative vector x in a linear way. In this case, the expected value of the GW rounding would be c.

The promise in the **NO** case is that for every Basic Sdp solution (for $\mathfrak{I}$ as a graph on linear forms) that is linearly generated by a Gaussian vector x with nonnegative second moment matrix $\mathbb{E}_x\, xx^T \succcurlyeq 0$, the expected value of the GW rounding algorithm is at most s.

The following theorem shows a strong dichotomy for the complexity of (c,s)-Gap Generalized Max Cut, determined exactly by the approximation guarantee of the GW rounding algorithm.

Theorem 4.2. *For all $c, s \in \mathbb{R}$ with $0 < s < c < 1$ and $c \geqslant 0.845$,*[13] *the promise problem (c, s)-Gap Generalized* Max Cut *is either* **NP***-hard or solvable in polynomial time (by* Basic Sdp*). Furthermore,* Basic Sdp *succeeds if and only if GW rounding achieves a (c, s)-approximation for* Max Cut*, i.e., $s \leqslant \arccos(1-2c)/\pi$.*

We like to mention an interesting known **NP**-hardness result in the context of Max Cut and the GW rounding. This result follows from techniques in [GRSW12] and was communicated to us by Prasad Raghavendra. (Here, we view SDP solutions as assignments of unit vectors to vertices and we think of the GW rounding as partitioning the vertices/vectors via a random hyperplane through the origin. This view is equivalent to ours.) The result is that given a Basic Sdp solution for a Max Cut instance, it is **NP**-hard to find a hyperplane that achieves a strictly better approximation factor for this Basic Sdp solution than a random hyperplane. We are not aware of any formal relations between this result and ours. Since for this result, the Basic Sdp solution is part of the input, the corresponding computational problem is not a generalization of Max Cut.

Generalized Distribution Matching. For this example, we think about CSPs in a predicate-independent way (similar to the DCSP hypothesis in Section 3.4). Given a collection K-tuples of literals, the question is what kind of distributions over $\{\pm 1\}^K$ can be generated by assigning Boolean values to the variables and outputting the values of a random K-tuple of literals from the collection.

For a distribution D over $\{\pm 1\}^K$, the following formalizes this problem in the generalized setting. (Here, a collection of K-tuples of literals is generalized to a collection of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^K$.)

Problem 4.3 (Generalized D-Distribution Matching)**.** Given a collection $\mathfrak{I}$ of linear maps $A \colon \mathbb{R}^n \to \mathbb{R}^K$, distinguish the cases,

YES: there exists $x \in \mathbb{R}^n_+$ such that D is $o(1)$-close to the distribution $\{Ax\}$, where A is drawn at random from $\mathfrak{I}$. (Furthermore, $Ax \in \{\pm 1\}^K$ for all $A \in \mathfrak{I}$.)

NO: for all Gaussian distributions over vectors $x \in \mathbb{R}^n$ with $\mathbb{E}_x\, xx^T \succcurlyeq 0$ and $\mathbb{E}_x \|Ax\|^2 \leqslant 1$ for all $A \in \mathfrak{I}$, we have $\mathbb{E}_x\, Q(Ax) \leqslant o(1)$ for all normalized[14] multilinear polynomials Q with no constant term.

(Here, $o(1)$ is a function going sufficiently slowly to 0 with the instance size, for concreteness say $1/\log\log\log n$. We could introduce another parameter to absorb the $o(1)$ term. This change would have little effect on our results.)

The **NO** promise roughly says that only distributions close to uniform can be generated by the collection $\mathfrak{I}$ (no matter the assignment x).

We show the following characterization of the complexity of D-distribution matching. (Assuming the Unique Games Conjecture, this characterization is known even for non-generalized distribution matching [AH12].)

[13] In this range of c, GW rounding is optimal for Basic Sdp of Max Cut.

[14] Since the dimension of the space is constant, the choice of the norm is not important.

Theorem 4.4. *For every distribution D over $\{\pm 1\}^K$, the promise problem D-distribution matching is either **NP**-hard or solvable in polynomial time (by* Basic Sdp*). Furthermore,* Basic Sdp *succeeds if and only if D is not pairwise-uniform.*

Generalized Boolean CSPs. Let $P\colon \{\pm 1\}^K \to \{0,1\}$ be a Boolean predicate. For $c, s \in \mathbb{R}$ with $0 < s < c < 1$, we consider a generalization of (c,s)-Gap CSP(P). The definition of the problem is based on the generic rounding algorithm for Basic Sdp (see §4.2 for details). Given an Gaussian vector x as part of a Basic Sdp solution for a CSP(P) instance $\mathfrak{I}$, the rounding scheme (with accuracy parameter η) computes an assignment for $\mathfrak{I}$ with value $\mathrm{rval}_{\mathfrak{I},\eta}(x)$. If Basic Sdp achieves a (c,s)-approximation for CSP(P), then for every feasible Basic Sdp solution x with value at least c for $\mathfrak{I}$, we have $\mathrm{rval}_{\mathfrak{I},\eta}(x) \geqslant s - \eta$ (see Theorem 4.8). For all fixed accuracy parameters $\eta > 0$, the running time of the rounding scheme is polynomial. In fact, the running time stays polynomial even for slightly subconstant η, say $\eta = 1/\log\log\log n$. We will use this choice later to eliminate η as a parameter.

In Generalized CSP(P), the goal is (roughly speaking) to find a Basic Sdp solution x satisfying certain simple additional constraints so as to maximize $\mathrm{rval}_\eta(x)$ (the value achieved by the generic rounding scheme applied to x).

Problem 4.5 ($(c,s)_\eta$-Gap Generalized CSP(P))**.** Given a CSP(P) instance $\mathfrak{I}$ with variable set $[n]$ and a simple cone[15] C of n-by-n matrices, distinguish the cases,

YES there exists a Basic Sdp solution x with value at least c for $\mathfrak{I}$ and second-moment matrix $\mathbb{E}_x xx^T \in C$ such that achieves $\mathrm{rval}_{\mathfrak{I},\eta}(x) \geqslant c$.

NO there exists *no* Basic Sdp solution x for $\mathfrak{I}$ with second-moment matrix $\mathbb{E}_x xx^T \in C$ such that the rounding scheme achieves $\mathrm{rval}_{\mathfrak{I},\eta}(x) \geqslant s - \eta$.

By design, Basic Sdp solves $(c,s)_\eta$-Gap Generalized CSP(P) for all $\eta > 0$ if it achieves a (c,s)-approximation for (non-generalized) CSP(P). The following shows a strong dichotomy for the complexity of $(c,s)_\eta$-Gap Generalized CSP(P).

Theorem 4.6. *For all $c, s \in \mathbb{R}$ with $0 < s < c < 1$ and every predicate $P\colon \{\pm 1\}^K \to \{0,1\}$, the promise problem $(c,s)_\eta$-Gap Generalized* CSP(P) *is either **NP**-hard, for small enough $\eta > 0$, or solvable in polynomial time for all $\eta > 0$ (by* Basic Sdp*). Furthermore,* Basic Sdp *solves $(c,s)_\eta$-Gap Generalized* CSP(P) *for all $\eta > 0$ if and only if* Basic Sdp *solves (non-generalized) (c,s)-Gap* CSP(P)*.*

We remark that the theorem above also holds for slightly subconstant η, say $\eta = 1/\log\log\log n$. For this choice of η, the quantification of the parameter η in the theorem statement above is not needed.

[15] Here, we say a cone $C \subseteq \mathbb{R}^{n\times n}$ is *simple* if it is the projection of a cone $C' \subseteq R^{n'\times n'}$ with $C' = \{Y \mid Y \succeq 0,\ \langle A_1, Y\rangle \geqslant 0,\ \ldots,\ \langle A_m, Y\rangle \geqslant 0\}$ and $n' + m \leqslant \mathrm{poly}(n)$. In particular, the cone C' can be described by a small semidefinite program.

4.2 Preliminaries (continued)

In this section, we continue the preliminaries of Section 2 and define basic notions about Label Cover, smoothness, and SDP rounding.

Label-Cover. A Label Cover instance $\mathcal{W}$ with vertex set V and alphabets $[d \cdot R]$ and $[R]$ is specified by a distribution over triples (u, v, π) with $u, v \in V$ and $\pi \colon [d \cdot R] \to [R]$. We may assume that the instance $\mathcal{W}$ is left-regular, so that every vertex u participates in the same number of constraints of the form (u, v, π). The value of an assignment $x \in [d \cdot R]^V$ is defined as the probability

$$\mathcal{W}(x) \stackrel{\text{def}}{=} \mathop{\mathbb{P}}_{(u,v,\pi),(u',v',\pi') \sim \mathcal{W}} \{\pi(x_v) = \pi'(x_{v'}) \mid u = u'\}$$

(It is more common to define the value of $\mathcal{W}$ for pairs of assignments $x \in [d \cdot R]^V$, $y \in [R]^V$ as the probability that $y_u = \pi(x_v)$ over $(u, v, \pi) \sim \mathcal{W}$. For our purposes, the two definitions are equivalent, but the first one is more convenient.)

Smoothness. We say that a label cover instance as above is σ-smooth if for every $v_0 \in V$ and any two labels $i, j \in [d \cdot R]$, the values of a random projection on V have probability at most σ to coincide on i and j. That is, it is σ smooth if

$$\mathop{\mathbb{P}}_{(u,v,\pi),(u',v',\pi') \sim \mathcal{W}} \{\pi(i) = \pi'(j) \mid v = v' = v_0\} \leqslant \sigma$$

It was first shown in [Kho02a] that smooth label cover is NP-hard to approximate. The following theorem is a simplification of Theorem 3.5 from [GRSW12].

Theorem 4.7. *For every constants $\sigma, \varepsilon > 0$ there exist d, R such that the following holds. Given an instance of* Label Cover *over alphabets $[d \cdot R]$ and $[R]$ which is σ smooth, it is NP-hard to distinguish between the case where the value of the instance is 1 and the case where its value is at most ε.*

(We remark that the theorem remains true with parameters σ and ε slightly subconstant [MR08].)

Generic Rounding Scheme

Let $\mathfrak{I}$ be an n-variable instance of CSP(P) for some K-ary boolean predicate P. Let x be an n-dimensional Gaussian vector, e.g., as part of a feasible solution to Basic Sdp($\mathfrak{I}$)). For a one-dimensional odd[16] rounding function $\phi \colon \mathbb{R} \to [-1, 1]$, we define

$$\mathfrak{I}(x, \phi) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{x} \mathop{\mathbb{E}}_{S \sim \mathfrak{I}} P\left(\phi\left(S(x)_1\right), \ldots, \phi\left(S(x)_K\right)\right).$$

Since ϕ is odd, there always exists an integral assignment $x' \in \{\pm 1\}^n$ with value $\mathfrak{I}(x') \geqslant \mathfrak{I}(x, \phi)$. (It is equivalent to apply the rounding functions to variables as

[16] Here, odd means $\phi(-x) = -\phi(x)$.

opposed to literals.) For an L-dimensional odd rounding function $\phi\colon \mathbb{R}^L \to [-1,1]$, we define

$$\mathfrak{I}(x,\phi) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{X=(x^{(1)},\dots,x^{(L)})} \mathop{\mathbb{E}}_{S\sim\mathfrak{I}} P\left(\phi\left(S(X)_1\right),\dots,\phi\left(S(X)_K\right)\right).$$

Here, $x^{(1)},\dots,x^{(L)}$ are independent samples drawn according to the distribution of x. We think of X as a L-by-n matrix. Recall that S encodes a K-tuple of literals. We extend S to matrices in the natural way. If the k^{th} literal in S is the i^{th} variable with sign σ_k, we set $S(X)_k = \sigma_k \cdot x_i$, where x_i is the i^{th} column of X.

The idea of generic rounding schemes is to enumerate an ε-net of constant-dimensional rounding functions and show that for every CSP(P) instance, one rounding function achieves a value arbitrarily close to the integrality gap of Basic Sdp for CSP(P).

For every $\eta > 0$, there exists a $O_\eta(1)$-sized set N_η of $O_\eta(1)$-dimensional odd rounding functions with Lipschitz constant $O_\eta(1)$ such that the following theorem holds. (We can assume the rounding functions to be odd because we allow negations in instances of CSP(P). We will also assume that N_η contains the sign function for all $\eta > 0$.)

Theorem 4.8 ([Rag08], see also [RS09] for a direct proof)**.** *Suppose* Basic Sdp *achieves a (c,s)-approximation for* CSP(P) *with $0 < s < c < 1$. If $\mathfrak{I}$ is a* CSP(P) *instance and x is a Gaussian vector, that corresponds to a feasible solution for* Basic Sdp($\mathfrak{I}$) *with value at least c, then for every $\eta > 0$,*

$$\mathrm{rval}_{\mathfrak{I},\eta}(x) \stackrel{\text{def}}{=} \max_{\phi\in N_\eta} \mathfrak{I}(x,\phi) \geqslant s - \eta\,.$$

We will refer to $\mathrm{rval}_{\mathfrak{I},\eta}(x)$ as the value achieved by the generic rounding scheme for instance $\mathfrak{I}$ and SDP solution x at accuracy η. (For this work, the precise construction of the sets N_η will not be important.)

For our results on Generalized CSPs, we will be interested in Gaussian vectors that satisfy certain linear equalities and linear inequalities. Adding additional Gaussian variables p makes it easier to express these linear constraints. In our setting, the original Gaussian variables x will be linear transforms of the new variables p, so that $x = Tp$ for a linear map T. Then, we can express $\mathfrak{I}(x,\phi)$ in terms of p (considering one-dimensional rounding functions ϕ for simplicity),

$$\mathfrak{I}(x,\phi) = \mathop{\mathbb{E}}_{p} \mathop{\mathbb{E}}_{S\sim\mathfrak{I}} P\left(\phi\left(S(Tp)_1\right),\dots,\phi\left(S(Tp)_K\right)\right)$$

For every $S \in \mathfrak{I}$, we can combine S and T into a linear map A. We will denote the resulting collection of the linear maps as $\mathfrak{I}'$ and write

$$\mathfrak{I}'(p,\phi) = \mathop{\mathbb{E}}_{p} \mathop{\mathbb{E}}_{A\sim\mathfrak{I}'} P\left(\phi\circ(Ap)\right).$$

(Here, $\phi\circ(\,\cdot\,)$ means that we apply ϕ coordinate-wise.) By construction, $\mathfrak{I}'(p,\phi) = \mathfrak{I}(x,\phi)$. (Also for multi-dimensional rounding functions.) We will refer to $\mathfrak{I}'$ as a GCSP(P) instance.

Generic Dictatorship Tests for Boolean Predicates

In this section, we recall definitions and results about dictatorship tests, that we need for our reductions. In particular, we give a formal statement of Raghavendra's result that for every predicate P, there exist dictatorship tests using P with guarantees arbitrarily close to the integrality gap of Basic Sdp for CSP(P).

Let $P\colon \{\pm 1\}^K \to \{0,1\}$ be a K-ary Boolean predicate and let $\mathcal{D}$ be a mixtures of distributions D over $\{\pm 1\}^K$. (Formally, we think of $\mathcal{D}$ as a distribution over distributions). The mixture $\mathcal{D}$ specifies a family of dictatorship test gadgets.

Definition 4.9 (Dictatorship test gadget)**.** For a mixture $\mathcal{D}$ of distributions D over $\{\pm 1\}^K$, the *R-dimensional dictatorship test gadget* $\mathfrak{I}_D$ for CSP(P) is an instance of CSP(P) with variable set $[K] \times \{\pm 1\}^R$ (K disjoint copies of $\{\pm 1\}^R$). An assignment $F = (f^{(1)}, \dots, f^{(K)})$ with $f^{(k)}\colon \{\pm 1\}^R \to \{\pm 1\}$ has the following value for $\mathfrak{I}_{\mathcal{D}}$,

$$\mathfrak{I}_{\mathcal{D}}(F) = \mathop{\mathbb{E}}_{D\sim\mathcal{D}} \mathop{\mathbb{E}}_{D^R} P(f^{(1)}(x^{(1)}), \dots, f^{(K)}(x^{(K)}))\,.$$

(Here, we think of D^R as a distribution over matrices $X = (x^{(1)}, \dots, x^{(K)})$ with columns $x^{(k)} \in \{\pm 1\}^R$ and each row drawn independently from D.) We extend $\mathfrak{I}_{\mathcal{D}}$ to assignments with range $[-1,1]$ in a multilinear way.

To avoid dealing with negations or multiple predicates, we will only consider symmetric assignments $F = (f^{(1)}, \dots, f^{(K)})$ so that $f^{(k)}(-x) = -f^{(k)}(x)$. It is a standard argument that this assumption is without loss of generality. (For example, one can think of $[K] \times \{\pm 1\}^R$ as the set of literals and $[K] \times \{\pm 1\}^{R-1}$ as the set of free variables.)

For a probability space Ω over $\{\pm 1\}$ and a function $f\colon \Omega^R \to \mathbb{R}$, let $\mathrm{Inf}_{\Omega,r} f$ denote the *ε-noisy influence* of coordinate r on f,

$$\mathrm{Inf}_{\Omega,\varepsilon,r} f \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{x_{-r}\sim\Omega^{[R]\setminus\{r\}}} \mathop{\mathrm{Var}}_{x_r\sim\Omega} T_{\Omega,\varepsilon} f(x)$$

Here, $T_{\Omega,\varepsilon}$ is the ε-noise operator for functions on Ω^R, so that $T_{\Omega,\varepsilon} f(x)$ is the expected value of $f(y)$ with y obtained from x by resampling every coordinate from Ω with probability ε.

For a tuple of functions $F = (f^{(1)}, \dots, f^{(K)})$ with $f^{(k)}\colon \mathbb{R}^R \to \mathbb{R}$, we define the ε-noisy influence of coordinate r with respect to the mixture $\mathcal{D}$,

$$\mathrm{Inf}_{\mathcal{D},\varepsilon,r} F \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{D\sim\mathcal{D}} \sum_{k\in[K]} \mathrm{Inf}_{D_k,\varepsilon,r} f^{(k)}\,.$$

Here, D_k denotes the marginal of the k^{th} coordinate in the distribution D over $\{\pm 1\}^K$. We say that F is ε-quasirandom if $\mathrm{Inf}_{\mathcal{D},\varepsilon,r} F \leqslant \varepsilon$ for all $r \in [R]$. (Typically, quasirandomness is defined differently, separating the noise parameter and the influence parameter. However, this definition is qualitatively equivalent to our, because influences decrease monotonically with noise.)

Definition 4.10. We say that $\mathcal{D}$ has *quasirandom soundness* s if there exists $\varepsilon > 0$ such that $\mathfrak{I}_{\mathcal{D}}(F) \leqslant s$ for all ε-quasirandom assignments $F = (f, \dots, f)$ with $f\colon \{\pm 1\}^R \to [-1, 1]$ (and all $R \in \mathbb{N}$).

(With our definition, the set of all values $s \in (0, 1)$ such that $\mathcal{D}$ has quasirandom soundness s forms an open interval.)

A (c, s) dictatorship test. We say that the mixture $\mathcal{D}$ is a (c, s)-dictatorship test distribution if $\mathbb{E}_{\mathcal{D}} P \geqslant c$ and if it has quasirandom soundness s. The parameter c indeed corresponds to the completeness of the associated dictatorship tests, because for every $R \in \mathbb{N}$ and every dictator $\chi_{\{r\}}$, the CSP(P) instance $\mathfrak{I}_{\mathcal{D}}$ has value at least $\mathfrak{I}_{\mathcal{D}}(\chi_{\{r\}}, \dots, \chi_{\{r\}}) = \mathbb{E}_{\mathcal{D}} P \geqslant c$.

The following theorem is a special case of the main ingredient in Raghavendra's Theorem.[17] (The general version of the theorem applies also to (collections of) predicates over larger alphabets.)

Theorem 4.11 ([Rag08])**.** *For every Boolean predicate $P\colon \{\pm 1\}^K \to \{0, 1\}$ and every $c, s \in \mathbb{R}$ with $0 < s < c < 1$, either* Basic Sdp *achieves a (c, s)-approximation for* CSP(P) *or there exists a (c, s)-dictatorship test distribution for P.*

4.3 Generalized Max Cut

In this section, we describe our reduction from Label Cover to Generalized Max Cut. We show completeness and soundness of the reductions. At the end of the section, we use the reduction to prove Theorem 4.2.

Let $P\colon \{\pm 1\}^2 \to \{0, 1\}$ be the Max Cut predicate, so that $P(x, y) = 1$ if and only if $x \neq y$. (We consider here the variant of Max Cut with negations. This problem is sometimes called XOR games.) The multilinear extension of P is $P(x, y) = (1 - xy)/2$.

Let D be a distribution over $\{\pm 1\}^2$ with mean 0. (We will use this distribution as a dictatorship test distribution. For the parameter range of Max Cut which we are interested at the moment, such a distribution (as opposed to a mixture) can achieve optimal dictatorship testing parameters [KKMO04, MOO05].)

Let $\mathcal{W}$ be a Label Cover instance with vertex set V and alphabets $[d \cdot R]$ and $[R]$. Let $\mathfrak{I}_D$ be the dictatorship test gadget on $\{\pm 1\}^R \dot\cup \{\pm 1\}^R$ corresponding to the test distribution D (see Definition 4.9). The *composition* of $\mathcal{W}$ and $\mathfrak{I}_D$ is a Generalized Max Cut instance $\mathfrak{I}_{\mathcal{W},D}$ over $\mathbb{R}^{V \times [d \cdot R]}$. For a Gaussian vector p over $\mathbb{R}^{V \times [d \cdot R]}$ and a rounding function $\phi\colon \mathbb{R} \to [-1, 1]$ (e.g., $\phi = \operatorname{sign}$), the value of $\mathfrak{I}_{\mathcal{W},D}$ is defined as

$$\mathfrak{I}_{\mathcal{W},D}(p, \phi) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{p} \mathop{\mathbb{E}}_{u} \mathop{\mathbb{E}}_{(v,\pi),(v',\pi') \sim \mathcal{W}|u} \mathfrak{I}_D(\phi \circ (H_{v,\pi} p), \phi \circ (H_{v',\pi'} p)) \,.$$

[17] Raghavendra's Theorem is usually stated with allowing an arbitrarily small slack $\varepsilon > 0$ between the parameters for the Basic Sdp and the dictatorship test. We can argue that it is not necessary to allow this slack explicitly. If Basic Sdp does not achieve a (c, s)-approximation for CSP(P), there exists some $\varepsilon > 0$ such that it also does not achieve a $(c + \varepsilon, s - \varepsilon)$-approximation.

Here, $H_{v,\pi}$ with $v \in V$ and $\pi\colon [d \cdot R] \to [R]$ is a linear map from $\mathbb{R}^{V\times[d\cdot R]}$ to the set of functions $f\colon \{\pm 1\}^R \to \mathbb{R}$,

$$H_{v,\pi}(p) := \sum_{i\in[d\cdot R]} p_{v,i}\chi_{\{\pi(i)\}}\,.$$

Completeness. Suppose there exists an assignment $x \in [d \cdot R]^V$ satisfying all constraints of $\mathcal{W}$. (Hence, there exists $y \in [R]^V$ such that $y_u = \pi(x_v)$ for all $(u, v, \pi) \in \mathcal{W}$.) Let $p \in \mathbb{R}^{V\times[d\cdot R]}$ be the $\{0, 1\}$-indicator of this assignment, so that $p_{v,i} = 1$ if $x_v = i$ and $p_{v,i} = 0$ otherwise. Then, $H_{v,\pi}p = \chi_{\pi(x_v)}$ and

$$\mathfrak{I}_{\mathcal{W},D}(p, \mathrm{id}) = \mathop{\mathbb{E}}_{u\in V} \mathfrak{I}_D(\chi_{\{y_u\}}, \chi_{\{y_u\}}) = \mathop{\mathbb{E}}_{D} P\,.$$

(Since we defined $\mathfrak{I}_{\mathcal{W},D}$ only for Gaussian vectors, we can think here of p as a constant Gaussian vector.)

Soundness. Suppose p is a Gaussian vector over $\mathbb{R}^{V\times[d\cdot R]}$ with $\mathbb{E}_p\, pp^T \succcurlyeq 0$ and $\mathbb{E}_p(H_{v,\pi}p(y))^2 = 1$ for all $y \in \{\pm 1\}^R$. (The latter is the normalization condition for generalized Max Cut.)

Since $\mathfrak{I}_D$ is block-multilinear (with respect to the two copies of $\{\pm 1\}^R$),

$$\mathfrak{I}_{\mathcal{W},D}(p, \phi) = \mathop{\mathbb{E}}_{p}\mathop{\mathbb{E}}_{u} \mathfrak{I}_D(f_u, f_u)\,,$$

where $f_u = \mathbb{E}_{(v,\pi)\sim\mathcal{W}|u}\, \phi \circ (H_{v,\pi}p)$.

Suppose that D has quasirandom soundness s so that $\mathfrak{I}_D(f, f) \leqslant s$ for all ε_1-quasirandom functions $f\colon \{\pm 1\}^R \to [-1, 1]$. Also suppose that $\mathcal{W}(p, \phi) \geqslant s + \varepsilon_2$ for some $1/\varepsilon_3$-Lipschitz odd rounding function $\phi\colon \mathbb{R} \to [-1, 1]$. (The sign function is not Lipschitz, but we can approximate it by such functions without significant change of the objective value, see Section 4.8.) Let ε be the minimum of $\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$.

If f is not ε-quasirandom, then $\sum_r (\mathrm{Inf}_r\, f)^2 \geqslant \varepsilon^2$. Hence, $\mathfrak{I}_{\mathcal{W},D}(p, \phi) \geqslant s + \varepsilon$ together with the quasirandom soundness of D implies

$$\varepsilon^3 \leqslant \mathop{\mathbb{E}}_{p}\mathop{\mathbb{E}}_{u}\sum_r (\mathrm{Inf}_r\, f_u)^2 \leqslant O(1/\varepsilon^2)\mathop{\mathbb{E}}_{p}\mathop{\mathbb{E}}_{u}\sum_r \left(\mathbb{E}_{(v,\pi)\sim\mathcal{W}|u}\, \mathrm{Inf}_r\, H_{v,\pi}p\right)^2\,.$$

(The last step uses convexity of influences and the fact that ϕ is $1/\varepsilon$-Lipschitz.) In case the smoothness σ of $\mathcal{W}$ satisfies $\sigma = o(\varepsilon^5)$, Theorem 4.15 (influence decoding from smoothly folded functions) allows us to conclude that there exists an assignment for $\mathcal{W}$ with value $\mathrm{poly}(\varepsilon)$.

Since for all $\eta > 0$ it is **NP**-hard to distinguish between the case that a label cover instance $\mathcal{W}$ has value 1 and the case that it has value and smoothness at most η, we showed the following theorem.

Theorem 4.12. *If there exists a (c, s)-dictatorship test distribution for* Max Cut*, then $(c, s + \varepsilon)$-Gap Generalized* Max Cut *is **NP**-hard for all $\varepsilon > 0$.*

Proof of Theorem 4.2. Let $c, s \in \mathbb{R}$ with $0 < s < c < 1$ and $c \geqslant 0.845$. If GW rounding achieves a (c, s)-approximation, it is clear that Basic Sdp solves (c, s)-Gap Generalized Max Cut. On the other hand, if GW rounding algorithm does not achieve a (c, s)-approximation for Max Cut, then the Basic Sdp relaxation for Max Cut does not achieve a (c, s)-approximation. (Here, we use that GW rounding is optimal in our range of c.) In this case, there exists $\varepsilon > 0$ such that Basic Sdp also does not achieve a $(c, s - \varepsilon)$-approximation for Max Cut. Hence, by Theorem 4.11, there exists a $(c, s - \varepsilon)$-dictatorship test distribution for Max Cut. By the above theorem, we conclude that $(c, (s - \varepsilon) + \varepsilon)$-Gap Generalized Max Cut is **NP**-hard, as desired. □

4.4 Generalized Distribution Matching

In this section, we describe our reduction from Label Cover to Generalized distribution matching (Problem 4.3). We show completeness and soundness of the reductions. At the end of the section, we use the reduction to prove Theorem 4.4.

The reduction has the same flavor as the reduction for Generalized Max Cut. One difference is that we can avoid rounding functions.

Let D be a pairwise independent distribution over $\{\pm 1\}^K$. Let Q be a K-variate normalized multilinear polynomial. (We don't assume that Q is a predicate, i.e., its range over $\{\pm 1\}^K$ is not necessarily $\{0, 1\}$. We extend our notations for K-ary predicates to K-variate polynomials in the natural way.)

Let $\mathcal{W}$ be a Label Cover instance with vertex set V and alphabets $[d \cdot R]$ and $[R]$. Let $\mathfrak{I}_{D,Q}$ denote the dictatorship test gadget over $[K] \times \{0, 1\}^R$ for CSP(Q) corresponding to the test distribution D, see Definition 4.9. (In this section, we will be interested in the behavior of $\mathfrak{I}_{D,Q}$ for all polynomials Q as above.) Let $\mathfrak{I}_{\mathcal{W},D,Q}$ be the composition of $\mathcal{W}$ and $\mathfrak{I}_{D,Q}$, a generalized CSP(Q) instance over $\mathbb{R}^{V \times [d \cdot R]}$. For a Gaussian vector p over $\mathbb{R}^{V \times [d \cdot R]}$, define its value as

$$\mathfrak{I}_{\mathcal{W},D,Q}(p) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{p} \mathop{\mathbb{E}}_{u \in V} \mathop{\mathbb{E}}_{(v^{(1)},\pi^{(1)}),\dots,(v^{(K)},\pi^{(K)}) \sim \mathcal{W}|u} \mathfrak{I}_{D,Q}(H_{v^{(1)},\pi^{(1)}} p, \dots, H_{v^{(K)},\pi^{(K)}} p)\,.$$

As in the previous section, $H_{v,\pi}$ with $v \in V$ and $\pi \colon [d \cdot R] \to [R]$ is a linear map from $\mathbb{R}^{V \times [d \cdot R]}$ to the set of functions $f \colon \{\pm 1\}^R \to \mathbb{R}$ with $H_{v,\pi}(p) := \sum_{i \in [d \cdot R]} p_{v,i} \chi_{\{\pi(i)\}}$.

Since $\mathfrak{I}_{D,Q}$ is block-multilinear (across the K copies of $\{\pm 1\}^R$), the expression for $\mathfrak{I}_{\mathcal{W},D,Q}(p)$ simplifies to

$$\mathfrak{I}_{\mathcal{W},D,Q}(p) = \mathop{\mathbb{E}}_{p} \mathop{\mathbb{E}}_{u} \mathfrak{I}_{D,Q}(f_u, \dots, f_u)\,,$$

where $f_u = \mathbb{E}_{(v,\pi) \sim \mathcal{W}|u} H_{v,\pi} p$.

Completeness. If there exists a satisfying assignment $x \in [d \cdot R]^V$ for $\mathcal{W}$, then for the $\{0, 1\}$-indicator $p \in \mathbb{R}^{V \times [d \cdot R]}$ of x,

$$\mathfrak{I}_{\mathcal{W},D,Q}(p) = \mathop{\mathbb{E}}_{D} Q\,.$$

Since this identity holds for all Q, the promise of the **YES** case for Problem 4.3 (Generalized D-distribution Matching) is satisfied.

Soundness. Let p be a Gaussian vector p with $\mathbb{E}_p\, pp^T \succcurlyeq 0$ and $\mathbb{E}_p(H_{v,\pi}p(y))^2 \leqslant 1$ for all $y \in \{\pm 1\}^R$. (The latter corresponds up to poly(K) factors to the normalization condition of Problem 4.3.) The invariance principle (in particular, the version in Theorem 4.18) implies

$$\mathfrak{I}_{\mathcal{W},D,Q}(p) \leqslant O_K(1) \cdot \mathop{\mathbb{E}}_{p} \mathop{\mathbb{E}}_{u} (\sum_r (\mathrm{Inf}_r\, f_u)^2)^{1/4} \cdot (1 + \|f_u\| + \ldots \|f_u\|^K)\,.$$

(Recall that the f_u's are linear functions on $\{\pm 1\}^R$.) By Theorem 4.15, the expectation of $\sum_r (\mathrm{Inf}_r\, f_u)^2$ is at most $\mathrm{opt}(\mathcal{W})^{\Omega(1)}$ (assuming sufficient smoothness of $\mathcal{W}$). Hence, by Cauchy–Schwarz, $\mathfrak{I}_{\mathcal{W},D,Q}(p) \leqslant O_K(1) \cdot \mathrm{opt}(\mathcal{W})^{\Omega(1)} \cdot (\mathbb{E}_p\, \mathbb{E}_u \sum_{k=0}^{K-1} \|f_u\|^{4k})^{1/4}$. Using estimates on Gaussian moments, we can bound also the last factor by $O_K(1)$. We conclude that if $\mathcal{W}$ has optimal value and smoothness $o(1)$, then $\mathfrak{I}_{\mathcal{W},D,Q}$ satisfies the promise of the **NO** case of Problem 4.3 for all normalized multilinear polynomials Q.

Together with known **NP**-hardness results for Label Cover [MR08], our reductions shows the following hardness for Generalized D-distribution matching.

Theorem 4.13. *If D is pairwise independent, then Generalized D-distribution matching is **NP**-hard.*

Proof of Theorem 4.4. If D is not pairwise independent, there exists $\varepsilon > 0$ and a quadratic polynomial P such that $\mathbb{E}_D\, P = \varepsilon$. The Basic Sdp finds a Gaussian x vector such that the first two moments of D and the distribution $\{Ax\}_{A \leftarrow_R \mathfrak{I}}$ match. Hence, it follows that $\mathbb{E}_x\, P(Ax) = \varepsilon$, contradicting the **NO** promise.

On the other hand, if D is pairwise independent, then the theorem above shows that the problem is **NP**-hard. □

4.5 Generalized CSPs

In this section, we describe our reduction from Label Cover to Generalized CSPs (Problem 4.3). We show completeness and soundness of the reductions. At the end of the section, we use the reduction to prove Theorem 4.6.

Again the flavor of the reduction is similar to the one for Generalized Max Cut. The main difference is that we have to deal with multi-dimensional rounding functions and more general dictatorship distributions.

Let $P\colon \{\pm 1\}^K \to \{0,1\}$ be a K-ary Boolean predicate. Let $\mathcal{D}$ be a mixture of distributions D over $\{\pm 1\}$. (This mixture will serve for our dictatorship gadget.)

Let $\mathcal{W}$ be a Label Cover instance with vertex set V and alphabets $[d \cdot R]$ and $[R]$. Let $\mathfrak{I}_{\mathcal{D}}$ denote the dictatorship test gadget over $[K] \times \{0,1\}^R$ for CSP(P) corresponding to the mixture $\mathcal{D}$ (see Definition 4.9). Let $\mathfrak{I}_{\mathcal{W},\mathcal{D}}$ be the composition of $\mathcal{W}$ and $\mathfrak{I}_{D,Q}$, a generalized CSP(P) instance over $\mathbb{R}^{V \times [d \cdot R]}$. For a Gaussian

vector p over $\mathbb{R}^{V\times[d\cdot R]}$ and an L-dimensional rounding function $\phi\colon \mathbb{R}^L \to [-1,1]$, define its value as

$$\mathfrak{I}_{\mathcal{W},\mathcal{D}}(p,\phi) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{\boldsymbol{p}} \mathop{\mathbb{E}}_{u\in V} \mathop{\mathbb{E}}_{(v^{(1)},\pi^{(K)}),\dots(v^{(K)},\pi^{(K)})\sim\mathcal{W}|u} \mathfrak{I}_{\mathcal{D}}(\phi\circ(H_{v^{(1)},\pi^{(1)}}\boldsymbol{p}),\dots,\phi\circ(H_{v^{(K)},\pi^{(K)}}\boldsymbol{p}))\,.$$

Here, $\boldsymbol{p}$ denotes a L-by-$V\times[d\cdot R]$ Gaussian matrix, each row independently drawn from the distribution of the Gaussian vector p. As in the previous section $H_{v,\pi}$, with $v\in V$ and $\pi\colon[d\cdot R]\to[R]$, takes a vector in $\mathbb{R}^{V\times[d\cdot R]}$ and defines a corresponding function $f\colon\{\pm1\}^R\to\mathbb{R}$ with $H_{v,\pi}p := \sum_{i\in[d\cdot R]} p_{v,i}\chi_{\{\pi(i)\}}$. We extend $H_{v,\pi}$ to matrices, so that $H_{v,\pi}\boldsymbol{p} :: \{\pm1\}^R\to\mathbb{R}^L$ is the vector-valued function $\sum_{i\in[d\cdot R]}\boldsymbol{p}_{v,i}\cdot\chi_{\{\pi(i)\}}$. After composition with a rounding function, $\phi\circ(H_{v,\pi}\boldsymbol{p})$ is a function on $\{\pm1\}^R$ with range $[-1,1]$.

Since $\mathfrak{I}_{\mathcal{D}}$ is block-multilinear (across the K copies of $\{\pm1\}^R$), the expression for $\mathfrak{I}_{\mathcal{W},D,Q}(p)$ simplifies to

$$\mathfrak{I}_{\mathcal{W},\mathcal{D}}(p) = \mathop{\mathbb{E}}_{\boldsymbol{p}}\mathop{\mathbb{E}}_{u}\mathfrak{I}_{\mathcal{D}}(f_u,\dots,f_u)\,,$$

where $f_u = \mathbb{E}_{(v,\pi)\sim\mathcal{W}|u}\,\phi\circ(H_{v,\pi}\boldsymbol{p})$.

Completeness. If p is the $\{0,1\}$-indicator of a satisfying assignment for $\mathcal{W}$ (which we can think of as a constant Gaussian vector), then $\mathfrak{I}_{\mathcal{W},\mathcal{D}}(p,\phi) = \mathbb{E}_{\mathcal{D}}\,P$ for any rounding function ϕ with $\phi(1,\dots,1)=1$ and $\phi(-1,\dots,-1)=-1$. Hence, $\mathfrak{I}_{\mathcal{W},\mathcal{D}}$ satisfies the **YES** promise of Problem 4.5 ((c,s)-Gap Generalized CSP(P)) with $c = \mathbb{E}_{\mathcal{D}}\,P$.

Soundness. Suppose p is a Gaussian vector over $\mathbb{R}^{V\times[d\cdot R]}$ with $\mathbb{E}_p\,pp^T\succcurlyeq 0$ and $\mathbb{E}_p(H_{v,p}p(y))^2 = 1$ for all $y\in\{\pm1\}^R$. (The latter corresponds to the normalization condition for solutions to Basic Sdp($\mathfrak{I}_{\mathcal{W},\mathcal{D}}$).

Suppose that $\mathcal{D}$ has quasirandom soundness s so that $\mathfrak{I}_{\mathcal{D}}(f,\dots,f)\leqslant s$ for all ε_1-quasirandom functions $f\colon\{\pm1\}^R\to[-1,1]$ and that $\mathfrak{I}_{\mathcal{W},\mathcal{D}}(p,\phi)\geqslant s+\varepsilon_2$ for an odd L-dimensional rounding function ϕ with Lipschitz constant $1/\varepsilon_3$. Let ε be the minimum of $\{\varepsilon_1,\varepsilon_2,\varepsilon_3\}$. Using similar arguments as for Generalized Max Cut, it follows that

$$\mathop{\mathbb{E}}_{p}\mathop{\mathbb{E}}_{u}\sum_{r}\left(\mathbb{E}_{(v,\pi)\sim\mathcal{W}|u}\,\mathrm{Inf}_{\mathcal{D},r}\,H_{v,\pi}p\right)^2 \geqslant (\varepsilon/LK)^{O(1)}\,.$$

Since $H_{v,\pi}p$ is linear, it is easy to verify that $\mathrm{Inf}_{\mathcal{D},r}\,H_{v,\pi}p\leqslant K\cdot\mathrm{Inf}_r\,H_{v,\pi}p$. (The latter influence is with respect to the usual uniform distribution on $\{\pm1\}$.)

Using Theorem 4.15 (influence decoding for smoothly folded functions), we can conclude that if $\mathcal{W}$ has optimal value and smoothness $(\varepsilon/LK)^{O(1)}$, then $\mathfrak{I}_{\mathcal{W},\mathcal{D}}$ satisfies the **NO** promise of Problem 4.5 (with soundness $s+\eta+\varepsilon$).

We can conclude the following theorem.

Theorem 4.14. *If there exists a (c,s)-dictatorship test for* CSP(P)*, then $(c,s+\eta+\varepsilon)_\eta$-Gap Generalized* CSP($P$) *is* ***NP****-hard for all $\varepsilon,\eta>0$.*

Proof of Theorem 4.6. Let $c, s \in \mathbb{R}$ with $0 < c < s < 1$. If Basic Sdp achieves a (c, s)-approximation for CSP(P), Theorem 4.8 implies that Basic Sdp also solves $(c, s)_\eta$-Gap Generalized CSP(P).

Otherwise, if Basic Sdp does not achieve a (c, s)-approximation for CSP(P), it does also not achieve a $(c, s - \varepsilon_0)$-approximation for some $\varepsilon_0 > 0$. Theorem 4.11 implies that there exists a $(c, s - \varepsilon_0)$-dictatorship test distribution for CSP(P). Then, the above theorem shows that for all $\eta < \varepsilon_0$, the problem $(c, s)_\eta$-Gap Generalized CSP(P) is **NP**-hard. □

4.6 Influence Decoding from Smoothly Folded Functions

Let V be a set of vertices and let $[d \cdot R]$ and $[R]$ be alphabets. For $v \in V$ and $\pi \colon [d \cdot R] \to [R]$, let $H_{v,\pi}$ be the linear map from $\mathbb{R}^{V \times [d \cdot R]}$ to the set of functions $f \colon \{\pm 1\}^R \to \mathbb{R}$,

$$H_{v,\pi} p := \sum_{i \in [d \cdot R]} p_{v,i} \cdot \chi_{\{\pi(i)\}} .$$

Theorem 4.15. *Let $\mathcal{W}$ be a label cover instance with vertex set V, alphabets $[d \cdot R]$ and $[R]$, and smoothness σ. Let p be a Gaussian vector over $\mathbb{R}^{V \times [d \cdot R]}$ with $\mathbb{E}_p pp^T \succcurlyeq 0$. Assume that $\mathbb{E}_p (H_{v,\pi} p(y))^2 \leqslant 1$ for all $y \in \{\pm 1\}^R$ and pairs (v, π) that appear in $\mathcal{W}$. (This assumption corresponds to the normalization condition for the assignment p.)*

Suppose $\varepsilon = \mathbb{E}_p \mathbb{E}_{u \in V} \sum_{r \in [R]} (\mathbb{E}_{(v,\pi) \sim \mathcal{W}|u} \operatorname{Inf}_r H_{v,\pi} p)^2$ and $\sigma \ll \varepsilon$. Then, there exists an assignment for $\mathcal{W}$ with value $\operatorname{poly}(\varepsilon)$.

Proof. We write $p = p' + p''$ with $p'_{v,i} = p_{v,i}$ if $\|p_{v,i}\|^2 > \eta$ and $p'_{v,i} = 0$ otherwise. (The parameter η is determined later.) Our strategy is to show that the contribution of p'' to ε is negligible. The random vector p' turns out to be sparse, which makes it easy to decode an assignment for $\mathcal{W}$ from it.

Let $P_\pi \colon \mathbb{R}^{d \cdot R} \to \mathbb{R}^R$ be as in Section 4.7. Then,

$$\begin{aligned} \varepsilon &= \mathop{\mathbb{E}}_p \mathop{\mathbb{E}}_u \sum_r \left(\mathbb{E}_{(v,\pi)|u} (P_\pi p_v)_r^2 \right)^2 \\ &\leqslant O(1) \cdot \mathop{\mathbb{E}}_p \mathop{\mathbb{E}}_u \sum_r \left(\mathbb{E}_{(v,\pi)|u} (P_\pi p'_v)_r^2 \right)^2 + O(1) \cdot \mathop{\mathbb{E}}_p \mathop{\mathbb{E}}_{(u,v,\pi)} \sum_r (P_\pi p''_v)_r^4 . \end{aligned}$$

(In the last step, we use triangle inequality to separate the contributions of p and p'', and we bound the term for p'' using convexity.) Using Lemma 4.16, we can bound the expectation of the term $\sum_r (P_\pi p''_v)_r^4$ (setting $a = p''_v$ and $b = P_\pi p''_v$ in the notation of the lemma),

$$\mathop{\mathbb{E}}_p \mathop{\mathbb{E}}_{(u,v,\pi)} \sum_r (P_\pi p''_v)_r^4 \leqslant O(1) \cdot \mathop{\mathbb{E}}_v \textstyle\sum_i \|p''_{v,i}\|^4 + O(\sigma) \cdot \mathop{\mathbb{E}}_v \|\sum_i p''_{v,i}\|^4 .$$

By construction, $\|p''_{v,i}\|^2 \leqslant \eta$ for all $v \in V$ and $i \in [d \cdot R]$. Since $\mathbb{E}_p pp^T \succcurlyeq 0$ (and thus $\mathbb{E}_p (p'')(p'')^T \succcurlyeq 0$), we have $\sum_i \|p''_{v,i}\|^2 \leqslant \|\sum_i p''_{v,i}\| \leqslant \|\sum_i p_{v,i}\|$ for all $v \in V$. Using the assumption $\mathbb{E}_p (H_{v,\pi} p(y))^2 \leqslant 1$ for all $y \in \{\pm 1\}^R$ (and again $\mathbb{E}_p pp^T \succcurlyeq 0$), we can bound $\|\sum_i p_{v,i}\| \leqslant 1$. Hence, we can bound the total contribution of p'' by

$$\mathop{\mathbb{E}}_p \mathop{\mathbb{E}}_{(u,v,\pi)} \sum_r (P_\pi p''_v)_r^4 \leqslant O(\eta + \sigma) .$$

It follows that the contribution of p' is at least $\varepsilon' := \mathbb{E}_p \mathbb{E}_u \sum_r (\mathbb{E}_{(v,\pi)|u} (P_\pi p'_v)_r^2)^2 \geqslant \Omega(\varepsilon) - O(\eta + \sigma)$. Consider the assignment $x \in [d \cdot R]^V$ for $\mathcal{W}$ obtained by assigning $x_v = i$ with probability at least $\|p_{v,i}\|^2$ independently for every $v \in V$. (Since $\sum_i \|p_{v,i}\|^2 \leqslant 1$ for all $v \in V$, such a distribution over assignments x exists.) The expected value of this assignment for $\mathcal{W}$ is at least

$$\mathop{\mathbb{E}}_x \mathcal{W}(x) \geqslant \mathop{\mathbb{E}}_u \sum_r \left(\mathop{\mathbb{E}}_{(v,\pi)|u} \sum_{r \in \pi^{-1}(r)} \|p_{v,i}\|^2 \right)^2$$

Our goal is to upper bound the contribution of p' in terms of this expected value of x,

$$\begin{aligned}
\varepsilon' &= \mathop{\mathbb{E}}_p \mathop{\mathbb{E}}_u \sum_r (\mathbb{E}_{(v,\pi)|u} (P_\pi p'_v)_r^2)^2 \\
&\leqslant O(1) \cdot \mathop{\mathbb{E}}_u \sum_r \left(\mathbb{E}_{(v,\pi)|u} \left\| \textstyle\sum_{i \in \pi^{-1}(r)} p'_{v,i} \right\|^2 \right)^2 \qquad (*) \\
&\leqslant O(1/\eta^2) \cdot \mathop{\mathbb{E}}_u \sum_r \left(\mathbb{E}_{(v,\pi)|u} \textstyle\sum_{i \in \pi^{-1}(r)} \|p'_{v,i}\|^2 \right)^2 \qquad (**) \\
&\leqslant O(1/\eta^2) \mathop{\mathbb{E}}_x \mathcal{W}(x) \,.
\end{aligned}$$

Step $(*)$ uses that $\{P_{v,\pi} p_v\}$ are jointly Gaussian and thus $\mathbb{E}_p (P_{v,\pi} p_v)_r^2 (P_{v',\pi'} p_{v'})_r^2 \leqslant O(1) \cdot \mathbb{E}_p (P_{v,\pi} p_v)_r^2 \cdot \mathbb{E}_p (P_{v',\pi'} p_{v'})_r^2$. Step $(**)$ uses that every random vector p'_v has at most $1/\eta$ coordinates that are not identically zero and therefore, $\|\sum_{i \in \pi^{-1}(r)} p'_{v,i}\|^2 \leqslant 1/\eta \cdot \sum_{i \in \pi^{-1}(r)} \|p'_{v,i}\|^2$.

We conclude that there exists an assignment for $\mathcal{W}$ with value at least $\Omega(\eta^2 \cdot \varepsilon')$. (Recall $\varepsilon' = \Omega(\varepsilon) - O(\eta + \sigma)$.) Since we assumed $\sigma \ll \varepsilon$, we can choose $\eta = \Omega(\varepsilon)$ such that $\varepsilon' = \Omega(\varepsilon)$. Hence, $\mathcal{W}$ has optimal value at least $\Omega(\varepsilon^3)$. □

4.7 Smooth Distributions over Functions

For $d, R \in \mathbb{N}$, consider the alphabets $\Sigma = [R]$ and $\Sigma' = [d \cdot R]$. Let $\mathcal{D}$ be a *σ-smooth* distribution over functions $\pi \colon \Sigma' \to \Sigma$, that is, for all $i, j \in \Sigma'$, the probability of collision is at most $\mathbb{P}_{\pi \sim \mathcal{D}} \{\pi(i) = \pi(j)\} \leqslant \sigma$.

For a function $\pi \colon \Sigma' \to \Sigma$, we define H_π to be the linear operator that maps a vector $a \in \mathbb{R}^{\Sigma'}$ to the functional $H_\pi a = \sum_{i \in \Sigma'} a_i y_{\pi(i)}$. Similarly, for a function $\pi \colon \Sigma' \to \Sigma$, let $P_\pi \colon \mathbb{R}^{\Sigma'} \to \mathbb{R}^{\Sigma}$ be the linear operator with $P_\pi a = \sum_{i \in \Sigma'} a_i e_{\pi(i)}$, where $e_1, \ldots, e_R$ is the standard basis of $\mathbb{R}^{\Sigma}$.

Lemma 4.16. *Let a be an Gaussian vector over $\mathbb{R}^{\Sigma'}$ with $\mathbb{E}_a aa^T \succcurlyeq 0$. Let b be a jointly distributed random vector over $\mathbb{R}^{\Sigma}$ obtained by sampling a function $\pi \sim \mathcal{D}$ and outputting $P_\pi a$. Then,*

$$\mathop{\mathbb{E}}_b \sum_r b_r^4 \leqslant O(1) \cdot \sum_i \|a_i\|^4 + O(\sigma) \cdot \|\textstyle\sum_i a_i\|^4 \,.$$

Here, $\|X\| = (\mathbb{E}\, X^2)^{1/2}$ denotes the L_2-norm of a random variable X.

Proof. Conditioned on π, each b_r is a Gaussian variable. Hence, its fourth moment $\mathbb{E}_{b|\pi} b_r^4$ is proportional to the square of its second moment,

$$\begin{aligned}\mathop{\mathbb{E}}_{b}\sum_r b_r^4 &= O(1)\mathop{\mathbb{E}}_{\pi\sim\mathcal{D}}\sum_{r\in\Sigma}\left\|\sum_{i\in\pi^{-1}(r)} a_i\right\|^4\\ &= O(1)\sum_{i,j,k,\ell\in\Sigma'}(\mathop{\mathbb{E}}_{a} a_i a_j)(\mathop{\mathbb{E}}_{a} a_k a_\ell)\cdot\mathbb{P}_{\pi\sim\mathcal{D}}\Big\{\pi(i)=\pi(i)=\pi(j)=\pi(k)\Big\}\\ &\leqslant \sum_{i\in\Sigma'}\|a_i\|^4 + \sigma\cdot\left\|\sum_{i\in\Sigma'} a_i\right\|^4\end{aligned}$$

The last step uses that $\mathcal{D}$ is σ-smooth and that $\mathbb{E}_a aa^T \succcurlyeq 0$. □

4.8 Lipschitz Approximation of Sign

In this section, we show that for the purpose of the §4.3 it is possible to approximate the rounding function sign by a Lipschitz function $\phi\colon \mathbb{R}\to[-1,1]$. For our approximation, we will use the piece-wise linear function $\phi_\varepsilon(x) := \mathrm{sign}(x)$ for $|x|>\varepsilon$ and $\phi_\varepsilon(x) := x/\varepsilon$ for $|x|\leqslant\varepsilon$. The function ϕ_ε is $1/\varepsilon$-Lipschitz.

For convenience of the reader, we will use similar notation as in §4.3. Let $\mathfrak{I}$ be a Max Cut instance with vertex set $\{\pm1\}^R$ (our dictatorship test gadget). Let f be Gaussian function on $\{\pm1\}^R$ (that is, the values $\{f(x)\}_{x\in\{\pm1\}^R}$ have a joint Gaussian distribution). Suppose $\mathbb{E}_f f(x)^2 = 1$ for all $x\in\{\pm1\}^R$.

Lemma 4.17.

$$\mathop{\mathbb{E}}_{f}\mathfrak{I}(\mathrm{sign}\circ f) = \mathop{\mathbb{E}}_{f}\mathfrak{I}(\phi_\varepsilon\circ f)\pm O(\sqrt{\varepsilon})\,.$$

Proof. Since $\mathfrak{I}(\cdot)$ is a quadratic form with bounded eigenvalues, it is enough to show $\mathbb{E}_f\|\mathrm{sign}\circ f-\phi_\varepsilon\circ f\|^2\leqslant O(\varepsilon)$. We can verify by direct calculation,

$$\begin{aligned}\mathop{\mathbb{E}}_{f}\|\mathrm{sign}\circ f-\phi_\varepsilon\circ f\|^2 &= \mathop{\mathbb{E}}_{x\in\{\pm1\}^R}\mathop{\mathbb{E}}_{f}(\mathrm{sign}(f(x))-\phi_\varepsilon(f(x)))^2\\ &\leqslant \mathop{\mathbb{E}}_{x\in\{\pm1\}^R}\mathop{\mathbb{P}}_{f}\{|f(x)|\leqslant\varepsilon\} = O(\varepsilon)\,.\end{aligned}$$

□

4.9 Pairwise Independence and Invariance Principle

In this section, we state a special case of the invariance principle [MOO05].

Let D be a pairwise independent distribution on $\{\pm1\}^K$. For $R\in\mathbb{N}$, let Δ_D be the following multilinear form on functions $f^{(1)},\dots,f^{(K)}\colon\{\pm1\}^R\to\mathbb{R}$,

$$\Delta_D(f^{(1)},\dots,f^{(K)}) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{D^R} f^{(1)}\cdots f^{(K)}\,.$$

(Here, we think of D^R as a distribution over vectors $x^{(1)},\dots,x^{(K)}\in\{\pm1\}^R$, each coordinate drawn independently from D.)

Theorem 4.18 ([MOO05, AM09])**.** *Let $f\colon\{\pm1\}^R\to\mathbb{R}$ be a function of degree at most d and $\mathbb{E} f = 0$. Then,*

$$\Delta_D(f,\dots,f)\leqslant O_{K,d}(1)\cdot\max_r(\mathrm{Inf}_r f)^{1/2}\|f\|^{K-1}\,.$$

Acknowledgements

We thank Avi Wigderson for some useful pointers on meta algorithmic results.

References

[AAM+11] Noga Alon, Sanjeev Arora, Rajsekar Manokaran, Dana Moshkovitz, and Omri Weinstein, *Inapproximability of densest k-subgraph from average case hardness*, 2011, Manuscript. 5, 19

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer, *Subexponential algorithms for unique games and related problems*, FOCS, 2010, pp. 563–572. 3, 8

[ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson, *Public-key cryptography from different assumptions*, STOC, 2010, pp. 171–180. 22

[AH12] Per Austrin and Johan Håstad, *On the usefulness of predicates*, IEEE Conference on Computational Complexity, 2012, To appear. 14, 24

[AKK+08] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K. Vishnoi, *Unique games on expanding constraint graphs are easy*, STOC, 2008, pp. 21–28. 3, 16

[Ale03] Michael Alekhnovich, *More on average case vs approximation complexity*, FOCS, IEEE Computer Society, 2003, pp. 298–307. 19

[AM09] Per Austrin and Elchanan Mossel, *Approximation resistant predicates from pairwise independence*, Computational Complexity **18** (2009), no. 2, 249–271. 5, 13, 36

[BBH+12] Boaz Barak, Fernando G.S.L. Brandão, Aram Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou, *Hypercontractivity, sum-of-squares proofs, and their applications*, STOC, 2012, To appear. 3

[BCC+10] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan, *Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph*, STOC, 2010, pp. 201–210. 5

[BCV+12] Aditya Bhaskara, Moses Charikar, Aravindan Vijayaraghavan, Venkatesan Guruswami, and Yuan Zhou, *Polynomial integrality gaps for strong SDP relaxations of densest k-subgraph*, SODA, 2012, pp. 388–405. 21

[BFKL93] Avrim Blum, Merrick L. Furst, Michael J. Kearns, and Richard J. Lipton, *Cryptographic primitives based on hard learning problems*, CRYPTO, Lecture Notes in Computer Science, vol. 773, Springer, 1993, pp. 278–291. 5

[BGLR93] Mihir Bellare, Shafi Goldwasser, Carsten Lund, and Alexander Russell, *Efficient probabilistically checkable proofs and applications to approximations*, STOC, 1993, pp. 294–304. 5, 20

[BGMT12] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani, *SDP gaps from pairwise independence*, 2012, Manuscript. 5, 16

[BHHS11] Boaz Barak, Moritz Hardt, Thomas Holenstein, and David Steurer, *Subsampling mathematical relaxations and average-case complexity*, SODA, SIAM, 2011, pp. 512–531. 3

[BK09] Libor Barto and Marcin Kozik, *Constraint satisfaction problems of bounded width*, FOCS, IEEE Computer Society, 2009, pp. 595–603. 3

[BT06] Andrej Bogdanov and Luca Trevisan, *On worst-case to average-case reductions for np problems*, SIAM J. Comput. **36** (2006), no. 4, 1119–1159, Preliminary version in FOCS '03. 5

[Bul02] Andrei A. Bulatov, *A dichotomy theorem for constraints on a three-element set*, FOCS, IEEE Computer Society, 2002, pp. 649–658. 3

[Fei02] Uriel Feige, *Relations between average case complexity and approximation complexity*, IEEE Conference on Computational Complexity, 2002, p. 5. 5, 14, 15, 19

[FKO06] Uriel Feige, Jeong Han Kim, and Eran Ofek, *Witnesses for non-satisfiability of dense random 3CNF formulas*, FOCS, 2006, pp. 497–508. 17

[FV98] Tomás Feder and Moshe Y. Vardi, *The computational structure of monotone monadic snp and constraint satisfaction: A study through datalog and group theory*, SIAM J. Comput. **28** (1998), no. 1, 57–104. 3

[GKL88] Oded Goldreich, Hugo Krawczyk, and Michael Luby, *On the existence of pseudorandom generators*, CRYPTO, Lecture Notes in Computer Science, vol. 403, Springer, 1988, pp. 146–162. 5

[GLS81] Martin Grötschel, László Lovász, and Alexander Schrijver, *The ellipsoid method and its consequences in combinatorial optimization*, Combinatorica **1** (1981), no. 2, 169–197. 3

[GRSW12] Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu, *Bypassing UGC from some optimal geometric inapproximability results*, SODA, 2012, pp. 699–717. 7, 9, 10, 11, 24, 26

[GW95] Michel X. Goemans and David P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM **42** (1995), no. 6, 1115–1145, Preliminary version in STOC '94. 9, 22

[Hås01] Johan Håstad, *Some optimal inapproximability results*, J. ACM **48** (2001), no. 4, 798–859. 5

[HN90] Pavol Hell and Jaroslav Nesetril, *On the complexity of h-coloring*, J. Comb. Theory, Ser. B **48** (1990), no. 1, 92–110. 3

[Kho02a] Subhash Khot, *Hardness results for coloring 3 -colorable 3 -uniform hypergraphs*, FOCS, IEEE Computer Society, 2002, pp. 23–32. 26

[Kho02b] _____, *On the power of unique 2-prover 1-round games*, STOC, 2002, pp. 767–775. 3

[KKMO04] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell, *Optimal inapproximability results for Max-Cut and other 2-variable CSPs?*, FOCS, 2004, pp. 146–154. 6, 10, 29

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell, *Optimal inapproximability results for Max-Cut and other* 2*-variable CSPs?*, SIAM Journal on Computing **37** (2007), no. 1, 319–357. 9

[KMM11] Alexandra Kolla, Konstantin Makarychev, and Yury Makarychev, *How to play unique games against a semi-random adversary: Study of semi-random models of unique games*, FOCS, 2011, pp. 443–452. 3

[Lau03] Monique Laurent, *A comparison of the sherali-adams, lovász-schrijver, and lasserre relaxations for 0-1 programming*, Math. Oper. Res. **28** (2003), no. 3, 470–496. 16

[LY80] John M. Lewis and Mihalis Yannakakis, *The node-deletion problem for hereditary properties is np-complete*, J. Comput. Syst. Sci. **20** (1980), no. 2, 219–230. 3

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences invariance and optimality*, FOCS, 2005, pp. 21–30. 6, 29, 36

[MOO10] E. Mossel, R. O'Donnell, and K. Oleszkiewicz., *Noise stability of functions with low influences: invariance and optimality*, Annals of Mathematics **171** (2010), no. 1, 295–341. 9

[Mos11] Dana Moshkovitz, *The projection games conjecture and the NP-hardness of* ln *n-approximating set-cover*, Electronic Colloquium on Computational Complexity (ECCC) **18** (2011), 112. 5, 20

[MR08] Dana Moshkovitz and Ran Raz, *Two query PCP with sub-constant error*, FOCS, 2008, pp. 314–323. 26, 32

[Rag08] Prasad Raghavendra, *Optimal algorithms and inapproximability results for every CSP?*, STOC, 2008, pp. 245–254. 3, 4, 6, 7, 8, 9, 10, 22, 27, 29

[RS95] Neil Robertson and Paul D. Seymour, *Graph minors. XIII. The disjoint paths problem*, J. Comb. Theory, Ser. B **63** (1995), no. 1, 65–110. 3

[RS04] _____, *Graph minors. XX. Wagner's conjecture*, J. Comb. Theory, Ser. B **92** (2004), no. 2, 325–357. 3

[RS09] Prasad Raghavendra and David Steurer, *How to round any CSP*, FOCS, 2009, pp. 586–594. 27

[Sch78] Thomas J. Schaefer, *The complexity of satisfiability problems*, STOC, 1978, pp. 216–226. 3

[ST04] Daniel A. Spielman and Shang-Hua Teng, *Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time*, J. ACM **51** (2004), no. 3, 385–463. 19

[Tul09] Madhur Tulsiani, *CSP gaps and reductions in the Lasserre hierarchy*, STOC, 2009, pp. 303–312. 5, 16, 17, 21

[Yan78] Mihalis Yannakakis, *Node- and edge-deletion np-complete problems*, STOC, ACM, 1978, pp. 253–264. 3