

# New Algorithms and Hardness Results for Robust Satisfiability of (Promise) CSPs

## Abstract

In this paper, we continue the study of robust satisfiability of promise CSPs (PCSPs), initiated in (Brakensiek, Guruswami, Sandeep, STOC 2023), and obtain the following results:

- For the PCSP 1-IN-3-SAT vs NAE-SAT with negations, we prove that it is hard, under the Unique Games conjecture (UGC), to satisfy  $1 - \Omega(1/\log(1/\epsilon))$  constraints in a  $(1 - \epsilon)$ -satisfiable instance. This shows that the exponential loss incurred by the BGS algorithm for the case of Alternating-Threshold polymorphisms is necessary, in contrast to the polynomial loss achievable for Majority polymorphisms.
- For any Boolean PCSP that admits Majority polymorphisms, we give an algorithm satisfying  $1 - O(\sqrt{\epsilon})$  fraction of the weaker constraints when promised the existence of an assignment satisfying  $1 - \epsilon$  fraction of the stronger constraints. This significantly generalizes the Charikar–Makarychev–Makarychev algorithm for 2-SAT, and matches the optimal trade-off possible under the UGC. The algorithm also extends, with the loss of an extra  $\log(1/\epsilon)$  factor, to PCSPs on larger domains with a certain structural condition, which is implied by, e.g., a family of Plurality polymorphisms.
- We prove that assuming the UGC, robust satisfiability is preserved under the addition of equality constraints. As a consequence, we can extend the rich algebraic techniques for decision/search PCSPs to robust PCSPs. The methods involve the development of a correlated and robust version of the general SDP rounding algorithm for CSPs due to (Brown-Cohen, Raghavendra, ICALP 2016), which might be of independent interest.

# 1 Introduction

The CSP dichotomy theorem has precisely identified which problems in the rich class of constraint satisfaction problems (CSPs) are polynomial-time solvable, with the rest being NP-complete [Zhu20, Bul17]. Strikingly, this landmark result shows that simple gadget reductions from 3-SAT are the only obstructions to the existence of an efficient algorithm for a CSP, and conversely the existence of a single “non-trivial polymorphism” suffices for a polynomial time satisfiability algorithm. Informally, a polymorphism is an operator that combines multiple satisfying assignments to the predicates defining the CSP into another satisfying assignment.

On the algorithmic side, in essence there are only two broad approaches: local consistency (also captured by a few levels of the Sherali–Adams hierarchy of linear programs) [BK14] and generalizations of Gaussian elimination [BD06, IMM<sup>+</sup>10]. However, the overall algorithm in the CSP dichotomy theorem is highly non-trivial due to the intricate ways in which these two basic algorithmic paradigms might have to be combined to solve an arbitrary tractable CSP.

The picture is considerably simpler and clearer (but is still non-trivial) when focusing on polynomial time *robust satisfiability algorithms* for tractable CSPs, a concept first considered in the beautiful work of Zwick [Zwi98]. In addition to finding perfectly satisfying assignments when they exist, such algorithms are also robust in the sense that, when given as input almost-satisfiable instances (namely those that admit an assignment failing to satisfy only an  $\epsilon$  fraction of the constraints), they find an assignment satisfying all but  $g(\epsilon)$  fraction of the constraints, for some loss  $g(\epsilon)$  that vanishes as  $\epsilon \rightarrow 0$ .

For ease of terminology, let us refer to CSPs that admit such efficient robust satisfiability algorithms as *robust CSPs*. We further call them  $(\epsilon, g(\epsilon))$ -robust CSPs to indicate the loss incurred by the robust algorithm. Every robust CSP is also tractable—this follows from two works of Barto and Kozik [BK14, BK16]. However, the converse is not true and there are tractable CSPs that lack robust satisfiability algorithms. The quintessential such CSPs are defined by linear relations over an Abelian group. Satisfiability of such CSPs can be efficiently ascertained via Gaussian elimination, but by the celebrated inapproximability results of Håstad [Hås01], they are not robust.<sup>1</sup> On the other hand, tractable CSPs that are solved by local consistency (called bounded-width CSPs in the literature) are in fact robust. For Boolean CSPs, this was effectively implicit in Zwick’s original work that pioneered robust satisfiability [Zwi98]. For CSPs over any fixed finite domains, this was shown by Barto and Kozik [BK16]. An earlier breakthrough result of Barto and Kozik [BK14] had shown that any CSP that cannot express linear equations (in a certain formal sense) is solved by local consistency algorithms. Thus, we have the pleasing picture that CSPs solved by local consistency—one of the two basic algorithmic strategies—are precisely those that are robust (this statement was explicitly conjectured in [GZ12]). Further, for all such CSPs, there is a robust satisfiability algorithm based on semidefinite programming. We therefore have a single unified approach for robust satisfiability compared to the highly complex situation for exact satisfiability of CSPs. In a way, the robust CSP dichotomy offers a crisper and more comprehensible complexity criterion.

Given this backdrop concerning robust CSPs, and motivated by the quest for understanding robustness of SDP-based algorithms more broadly, Brakensiek, Guruswami, and Sandeep [BGS23b] initiated the study of robust algorithms for *promise CSPs*, which we introduce next. Promise CSPs (also PCSPs for short) are a vast generalization of CSPs that have received significant attention in

---

<sup>1</sup>In fact, even for almost-satisfiable instances, it is NP-hard to beat the approximation ratio achieved by the trivial algorithm that simply outputs a random assignment.

recent years [GS20, BWŽ21, BBB21, BG21b, BK22, BŽ22, NŽ22, ČŽ23d, ČŽ23a, KOWŽ23, BGS23a, ČŽ23c, ČŽ23b, NŽ24, FNO<sup>+</sup>24, ČŽ24, HMNŽ24, BK24, AFO<sup>+</sup>25, LŽ25, NVWŽ25, CKK<sup>+</sup>25, Mot25]. A promise CSP is defined by a fixed collection of relation pairs  $(P_i, Q_i)$  over some domain pair  $(D, E)$ , with  $P_i \subseteq Q_i$ .<sup>2</sup> Given a CSP instance based on the relations, the goal is to find an assignment satisfying the (weak) constraints given by the  $Q_i$  relations if promised that an assignment satisfying the (strong) constraints given by the  $P_i$  relations exists (but is not known). A classic example of a promise CSP is the *approximate graph coloring* problem [GJ76]: given a  $k$ -colorable graph, find an  $\ell$ -coloring of it, where  $3 \leq k \leq \ell$ . In our terminology, this is just the PCSP with a single pair of relations  $(P, Q)$ , where  $P$  is the disequality relation on a  $k$ -element set, and  $Q$  is the disequality relation on an  $\ell$ -element set.

Another, more recent, example of a PCSP is the  $(2 + \epsilon)$ -SAT problem, introduced and studied by Austrin, Guruswami, and Håstad [AGH17] (who also coined the expression promise CSP): given an instance of  $k$ -SAT with the promise that an assignment exists that satisfies at least  $g$  literals in each clause, where  $1 \leq g \leq k$ , find a standard satisfying assignment (satisfying at least 1 literal in each clause). In this case, the  $P$  relations encode Boolean clause assignments with Hamming weight at least  $g$ , where the  $Q$  relations encode Boolean clause assignments with Hamming weight at least 1. Another example is the 1-IN-3-SAT vs NAE-SAT problem, identified in the influential paper of Brakensiek and Guruswami [BG21a] that initiated a systematic study of Boolean PCSPs. Here one is given a satisfiable instance of 1-IN-3-SAT and the goal is to find an assignment that satisfies 1 or 2 variables per clause. Astoundingly, this problem is solvable in polynomial time (via an algorithm not previously considered in the context of CSPs) [BG21a]. Further, such an algorithm cannot be obtained via a reduction to (finite domain) CSPs [BBKO21]!

The study of PCSPs calls for significant new algorithmic and hardness techniques. Studying such techniques in the broader context of PCSPs has also led to new results for (standard, non-promise) CSPs, e.g., a single algorithm blending together linear programming with linear Diophantine equations [BGWŽ20] that solves all tractable Boolean CSPs. Despite a lot of attention and recent progress on PCSPs, the complexity landscape is vast and mostly not understood. In fact, the complexity of Boolean PCSPs is itself a major challenge, in contrast to the CSP world where Schaefer proved a dichotomy for Boolean CSPs already in the 1970s [Sch78]. Following a classification of Boolean symmetric PCSPs allowing negations from [BG21a], Fíček, Kozík, Olšák, and Stankiewicz obtained a classification of Boolean PCSPs with symmetric relations [FKOS19]. Moreover, Brakensiek, Guruswami, and Sandeep obtained a (conditional) classification of monotone Boolean PCSPs [BGS23a].<sup>3</sup>

Returning to robust satisfiability, given the crisp picture of robust CSPs—namely, either the natural SDP gives an efficient robust algorithm or none exists—the study of robust PCSPs is a natural goal, as proposed by Brakensiek, Guruswami, and Sandeep (BGS) [BGS23b]. A robust satisfiability algorithm for a PCSP defined by relation pairs  $(P_i, Q_i)$  means the following: given an instance such that  $(1 - \epsilon)$  fraction of the constraints are promised to be satisfiable according to the stronger relations  $P_i$ , there is an algorithm to weakly satisfy (according to the relations  $Q_i$ )  $(1 - g(\epsilon))$  fraction of the constraints, where  $g(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . As with CSPs, in this case we say that the PCSP is robust or  $(\epsilon, g(\epsilon))$ -robust.

BGS focused on Boolean PCSPs where the known satisfiability algorithms<sup>4</sup> can be attributed to

<sup>2</sup>More generally, there must be a map  $h : D \rightarrow E$  that is a homomorphism from each  $P_i$  to  $Q_i$ .

<sup>3</sup>The classification assumes the Rich 2-to-1 conjecture of Braverman, Khot, and Minzer [BKM21].

<sup>4</sup>Recall that robust PCSPs must first of all be tractable.

the existence of polymorphisms. A polymorphism for a relation pair  $(P, Q)$  is a homomorphism from a (categorical) power  $P^m$  of  $P$  to  $Q$ , with  $\text{Pol}(P, Q)$  representing the set of all polymorphisms. See Section 2.3 for a formal definition. More precisely, BGS focused on three families of polymorphisms: Majority (MAJ), Alternating Threshold (AT), and Parity [BG21a]. For any odd  $L \in \mathbb{N}$ , we let  $\text{MAJ}_L, \text{AT}_L, \text{PAR}_L : \{-1, 1\}^L \rightarrow \{-1, 1\}$  be defined as

$$\begin{aligned} \text{MAJ}_L(x_1, \dots, x_L) &:= \mathbf{1} \left[ \sum_{i=1}^L x_i \geq 0 \right], \\ \text{AT}_L(x_1, \dots, x_L) &:= \mathbf{1} \left[ \sum_{i=1}^L (-1)^{i-1} x_i \geq 0 \right], \\ \text{PAR}_L(x_1, \dots, x_L) &:= \mathbf{1} \left[ \sum_{i=1}^L x_i \equiv L \pmod{4} \right]. \end{aligned}$$

We let  $\text{MAJ} := \{\text{MAJ}_L : L \in \mathbb{N} \text{ odd}\}$ ,  $\text{AT} := \{\text{AT}_L : L \in \mathbb{N} \text{ odd}\}$ ,  $\text{PAR} := \{\text{PAR}_L : L \in \mathbb{N} \text{ odd}\}$  be the respective sets of polymorphisms. In two of these cases, Majority and AT, BGS showed that the associated PCSPs are robust via an algorithm based on semidefinite programming [BGS23b]. They also showed that a partial converse holds: for PCSPs defined by a single pair  $(P, Q)$  of symmetric Boolean relations (plus allowing negations), if the relation pair  $(P, Q)$  lacks some odd-arity MAJ and some odd-arity AT as a polymorphism, then the PCSP is not robust (assuming the Unique Games conjecture).

The quantitative aspects of the robust satisfaction algorithms in [BGS23b] for the two cases, Majority and AT, however, diverged significantly. For Majority, the BGS algorithm, which is really the same as the Charikar–Makarychev–Makarychev algorithm for 2-SAT but analyzed in greater generality assuming only a Majority polymorphism, guaranteed that at most  $\tilde{O}(\epsilon^{1/3})$  fraction of the constraints are violated. This is weaker than the  $(\epsilon, O(\sqrt{\epsilon}))$  robustness guarantee for 2-SAT [CMM09]—which is tight [KKMO07, MOO10] under the Unique Games conjecture [Kho02, KKMO07, MOO10].<sup>5</sup> A natural question then is whether the BGS loss guarantee for Majority polymorphisms can be improved to  $O(\sqrt{\epsilon})$ , which would then give the right polymorphic generalization of the CMM robust algorithm for 2-SAT.

The situation for robust algorithms for Boolean PCSPs with AT polymorphisms is worse, as the BGS algorithm only showed  $(\epsilon, O(1/\log(1/\epsilon)))$ -robustness. A natural question then is whether this exponential loss is necessary, or whether one can achieve polynomial loss also for the AT case similar to the Majority case. We address and resolve both of these questions in this work.

## 1.1 Our Results

Our contributions in this paper fall into three parts. First, we show that the robust algorithm for Alternating Threshold due to Brakensiek–Guruswami–Sandeep [BGS23b] has a near-matching hardness result under the UGC. This is based on a novel integrality gap for (1-IN-3-SAT, NAE-SAT). Second, we show that any promise template with the Majority polymorphism has a robust algorithm with loss  $g(\epsilon) = O(\sqrt{\epsilon})$ , which is asymptotically tight, and improves over BGS’s analysis of

<sup>5</sup>Historically, showing evidence for the near-optimality of the Goemans–Williamson robust algorithm for Max-Cut, which also achieves  $O(\sqrt{\epsilon})$  loss, was the original motivation for the formulation of the Unique Games conjecture in [Kho02].

$O(\epsilon^{1/3})$ . We further extend this analysis to show that similar algorithms achieve a robustness of  $O(\sqrt{\epsilon} \log(1/\epsilon))$  for Plurality and related polymorphisms. Finally, we show that the robustness of PCSPs is (approximately) preserved under a large family of gadget reductions under the UGC. We do this by solving a seemingly elementary but technically complex problem: given a promise template with a robust algorithm, show that adding the equality relation to the template (approximately) preserves the robustness of the problem.

**Hardness for Alternating Threshold.** A key result of Brakensiek–Guruswami–Sandeep [BGS23b] is that for any promise template  $(P, Q)$  with  $\text{AT} \subseteq \text{Pol}(P, Q)$ ,  $\text{PCSP}(P, Q)$  is robust with  $g(\epsilon) = O(\frac{\log \log(1/\epsilon)}{\log(1/\epsilon)})$ . Interestingly, for CSPs, similar asymptotics appear with the OR and AND families<sup>6</sup> of polymorphisms, and these are known to be tight [Zwi98, GZ12]. We show that AT exhibits a similar behavior by proving UGC hardness.

**Theorem 1.1** (AT hardness, informal). *Assuming UGC,  $\text{fiPCSP}(1\text{-IN-3-SAT}, \text{NAE-SAT})$  is not  $(\epsilon, \Omega(1/\log(1/\epsilon)))$ -robust.*

Here  $\text{fiPCSP}$  refers to PCSPs that allow for variables to be negated and set as constants (see [BGS23b]). The use of negations is necessary, as Brakensiek–Guruswami–Sandeep [BGS23b] observed that  $(1\text{-IN-3-SAT}, \text{NAE-SAT})$  without negations is robust with polynomial loss. Theorem 1.1 is proved in Section 3. Using Raghavendra’s theorem [Rag08], we prove Theorem 1.1 by constructing an explicit integrality gap—the same high-level strategy as used by Guruswami–Zhou [GZ12] for HORN-3-SAT, although the execution and analysis in our setting are significantly more complex. We explain further details in Section 1.2.

**Improved Analysis for Majority and Beyond.** Our next result is a robust algorithm for PCSPs with Majority polymorphisms with an improved loss function.

**Theorem 1.2** (MAJ robustness, informal). *For any promise template  $(P, Q)$  with  $\text{MAJ} \subseteq \text{Pol}(P, Q)$ ,  $\text{PCSP}(P, Q)$  is  $(\epsilon, O(\epsilon^{1/2}))$ -robust.*

This result is proved in Section 4. We note that the algorithm used in Theorem 1.2 is identical to the one used in [BGS23b]. The main improvement in the analysis comes from a more refined analysis of multivariate normal distributions. See Section 1.2 for further details.

Note that Theorem 1.2 only applies to Boolean PCSPs. A commonly studied non-Boolean variant of the Majority polymorphisms is the *Plurality* polymorphisms. Typical examples of (P)CSPs admitting such polymorphisms are *unique games* [Kho02] as well as the so-called SetSAT problem—a non-Boolean generalization of  $(2 + \epsilon)$ -SAT [AGH17] introduced in [BWŻ21].

**Theorem 1.3** (PLUR robustness, informal). *For any promise template  $(P, Q)$  with  $\text{PLUR} \subseteq \text{Pol}(P, Q)$ ,  $\text{PCSP}(P, Q)$  is  $(\epsilon, O(\epsilon^{1/2} \log(1/\epsilon)))$ -robust.*

This result is proved in Section 5. We note that Theorem 1.3 is merely a special case of our main result in Section 5, which applies to any *separable* PCSP, see Theorem 5.3. We describe this broader family more precisely in Section 1.2.

**Robust Gadget Reductions (adding EQUALITY).** In virtually all classifications of (variants of) CSPs, an essential tool is *gadget reductions* between CSP templates. For example, in the CSP

<sup>6</sup>For any  $L \in \mathbb{N}$ ,  $\text{OR}_L(x_1, \dots, x_L) = 1$  if  $x_i = 1$  for some  $i \in [L]$  and  $\text{AND}_L(x_1, \dots, x_L) = 1$  if  $x_i = 1$  for all  $i \in [L]$ .

dichotomy, the hardness side of the CSP is done using gadget reductions from 3-SAT [BJK05]. Robust (P)CSPs are no exception, and gadget reductions are frequently used to study the relationship between templates [DK13, BGS23b]. However, there is a significant distinction between the ordinary CSP dichotomy [Bul17, Zhu20] and the one for robust (P)CSPs [BK16]: the allowance of *equality constraints*, which we denote by EQ. For exact satisfiability of (P)CSPs, if we specify that some variables are to be equal, we can efficiently compute the connected components of the equality relation and distill the problem down to a smaller number of variables (and without any equality constraints). However, for robust (P)CSPs, equality is a rather subtle concept. If only  $1 - \epsilon$  constraints are satisfied in the optimal assignment, we do not know whether each equality constraint should be trusted or ignored. That said, it still seems quite reasonable to assume that adding equality constraints should only mildly change the robustness of the resulting PCSP.

**Question 1.4** (Barto–Kozik [BK16]). *Let  $\Gamma$  be a CSP template, i.e., a set of relations. Assume that  $\text{CSP}(\Gamma)$  is  $(\epsilon, f(\epsilon))$ -robust. Is  $\text{CSP}(\Gamma \cup \{\text{EQ}\})$  then  $(\epsilon, O(f(\epsilon)))$ -robust?*

Despite Barto–Kozik [BK16] giving a complete classification of all robust CSPs,<sup>7</sup> they do not answer Question 1.4 except in the very weak sense that  $\text{CSP}(\Gamma \cup \{\text{EQ}\})$  is  $(\epsilon, O(\log \log(1/\epsilon)/\log(1/\epsilon)))$ -robust, independent of  $f$ . In this paper, we establish Question 1.4 is *nearly* true for both CSPs and PCSPs:

**Theorem 1.5** (Robustness of Equality, informal). *Assume UGC. For any promise template  $\Gamma$ , if  $\text{PCSP}(\Gamma)$  is  $(\epsilon, f(\epsilon))$ -robust, then  $\text{PCSP}(\Gamma \cup \{\text{EQ}\})$  is  $(\epsilon, O(f(\epsilon^{1/6})))$ -robust.*

As an immediate corollary, we can now use the most general gadget reductions available for studying (P)CSPs to study robust (P)CSPs, modulo a polynomial loss in robustness.

**Theorem 1.6** (Gadget Reductions, informal). *Assume UGC. Let  $\Gamma$  and  $\Gamma'$  be promise templates such that there is a gadget reduction<sup>8</sup> from  $\text{PCSP}(\Gamma')$  to  $\text{PCSP}(\Gamma)$ . If  $\text{PCSP}(\Gamma)$  is  $(\epsilon, f(\epsilon))$ -robust, then  $\text{PCSP}(\Gamma')$  is  $(\epsilon, O(f(\epsilon^{1/6})))$ -robust.*

These results are proved in Section 6. In the technical overview (Section 1.2), we describe the techniques we use to establish our results, including an adaptation of the algorithm of Brown-Cohen and Raghavendra [BR16] for approximate (P)CSPs.

## 1.2 Technical Overview

**Hardness for Alternating Threshold.** Our integrality gap instance for 1-IN-3-SAT vs NAE-SAT has a similar high-level idea to the following LP integrality gap instance for Horn-SAT [GZ12]:

1. We have variables  $\{x_{j1}, x_{j2} : j \in [k]\}$  where  $k = \lceil 1/\log_2(\epsilon) \rceil + 1$ .
2. We have the unary constraints  $x_{11}$ ,  $x_{12}$ , and  $\neg x_{k1}$  where 1 is True and  $-1$  is False.
3. For all  $j \in [k - 1]$ , we have the constraints  $\neg x_{j1} \vee \neg x_{j2} \vee x_{(j+1)1}$  and  $\neg x_{j1} \vee \neg x_{j2} \vee x_{(j+1)2}$ .

---

<sup>7</sup>We observe that Barto–Kozik were able to establish this classification by using gadget reductions that do *not* allow equality. Using Theorem 1.6, one could (in theory) simplify parts of Barto–Kozik’s proof, although the proof of Theorem 1.6 is much more complicated than the workarounds needed by Barto–Kozik.

<sup>8</sup>More precisely, there is a *minion homomorphism* from  $\text{Pol}(\Gamma)$  to  $\text{Pol}(\Gamma')$ .

Clearly this instance is unsatisfiable, as we insist that  $x_{11}, x_{12}$  are True, and then  $x_{k1}, x_{k2}$  should also be True due to the chain of implication constraints, but we insist that  $x_{k1}$  is False. As there are only  $O(k) = O(\log(1/\epsilon))$  constraints, the integral value of this instance is at most  $1 - \Omega(1/\log(1/\epsilon))$ .

On the other hand, the following LP solution gives value at least  $1 - \epsilon$  to all of the constraints:

1. For each  $j \in [k]$ , we give  $x_{j1}$  and  $x_{j2}$  bias  $1 - 2^{j+1-k}$ .
2. For the constraint  $x_{11}$ , we can set  $x_{11} = 1$  with probability  $1 - 2^{1-k}$  and  $-1$  with probability  $2^{1-k}$  and we will have that  $\mathbb{E}[x_{11}] = 1 - 2^{1-k} - 2^{1-k} = 1 - 2^{2-k}$ . Thus, the LP gives a value of  $1 - 2^{1-k} \geq 1 - \epsilon$  for this constraint. By symmetry, the LP also gives a value of  $1 - 2^{1-k} \geq 1 - \epsilon$  for the constraint  $x_{12}$ .
3. For each  $j \in [k-1]$ , for the constraint  $\neg x_{j1} \vee \neg x_{j2} \vee x_{(j+1)1}$ , we can take the distribution where
  - a. With probability  $1 - 2^{j+1-k}$ , we set  $x_{j1} = x_{j2} = x_{(j+1)1} = 1$ .
  - b. With probability  $2^{j-k}$ , we set  $x_{j1} = 1$ ,  $x_{j2} = -1$ , and  $x_{(j+1)1} = 1$ .
  - c. With probability  $2^{j-k}$ , we set  $x_{j1} = -1$ ,  $x_{j2} = 1$ , and  $x_{(j+1)1} = 1$ .

With this distribution,  $\mathbb{E}[x_{j1}] = \mathbb{E}[x_{j2}] = (1 - 2^{j-k}) - 2^{j-k} = 1 - 2^{j+1-k}$  and  $\mathbb{E}[x_{(j+1)1}] = (1 - 2^{j+1-k}) - 2^{j+1-k} = 1 - 2^{j+2-k}$ . Thus, the LP gives this constraint a value of 1. By symmetry, the LP also gives a value of 1 to the constraint  $\neg x_{j1} \vee \neg x_{j2} \vee x_{(j+1)2}$ .
4. The bias for  $x_{k1}$  is  $1 - 2 = -1$  so  $x_{k1}$  is always set to  $-1$  which satisfies the constraint  $\neg x_{k1}$ .

One way to think about this integrality gap instance is as follows. In order to avoid violating a constraint, the following must hold.

1. Variables with bias  $1 - 2^{j+1-k}$  must be rounded to 1.
2. For all  $j \in [k-1]$ , if the variables with bias  $1 - 2^{j+1-k}$  are rounded to 1 then the variables with bias  $1 - 2^{j+2-k}$  are rounded to 1.
3. Variables with bias  $-1$  must be rounded to  $-1$ .

Since we can obtain a contradiction in  $O(\log(1/\epsilon))$  steps, at least  $\Omega(1/\log(1/\epsilon))$  of the constraints must be violated.

We will use a similar idea for our integrality gap instance. We will construct our instance so that while the SDP value is at least  $1 - \epsilon$ , if we want to avoid violating a significant number of constraints,

1. Almost all of the variables with bias  $1 - 2^{-k}$  must be set to 1 and almost all of the variables with bias  $2^{-k} - 1$  must be set to  $-1$ .
2. For all  $j \in [k]$ , if almost all of the variables with bias  $1 - 2^{-j}$  are set to 1 then almost all of the variables with bias  $2^{1-j} - 1$  are set to  $-1$ . Similarly, for all  $j \in [k]$ , if almost all of the variables with bias  $2^{-j} - 1$  are set to  $-1$  then almost all of the variables with bias  $1 - 2^{1-j}$  are set to 1.

We then observe that these conditions imply that almost all of the variables with bias 0 are set to 1 and almost all of the variables with bias 0 are set to  $-1$ , which is impossible. Since we can obtain a contradiction in  $O(\log(1/\epsilon))$  steps, at least  $\Omega(1/\log(1/\epsilon))$  of the constraints must be violated.

In order to have 1-IN-3-SAT constraints, it turns out that we need general vectors of the form  $\{x\mathbf{v}_0 + \sqrt{1-x^2}\mathbf{w}\}$  where  $\mathbf{w}$  is orthogonal to  $\mathbf{v}_0$ . A natural choice for this is to use all  $\mathbf{w} \in S^{d-1}$  for some large  $d$  which depends on  $\epsilon$ . While this gives an integrality gap instance, it has infinite size and is not that easy to analyze since it involves functions on the sphere rather than functions with multivariate Gaussian inputs. Thus, we modify this integrality gap instance as follows:

1. Instead of using vectors  $\mathbf{w} \in S^{d-1}$ , we will use vectors  $\mathbf{w} \sim \mathcal{N}(0, 1/d)^d$ .
2. For all but a negligible portion of the constraints, our vectors  $\mathbf{w} \sim \mathcal{N}(0, 1/d)^d$  are very close to unit vectors so we can discard the negligible number of constraints where the vectors are badly behaved.
3. We discretize our instance by splitting our space into regions and mapping all vectors in each region to a representative vector in that region.

Through a careful analysis, we show that for this modified instance,  $\Omega(1/\log(1/\epsilon))$  fraction of the constraints must be violated and even after these modifications, the SDP value for our instance is at least  $1 - \epsilon$ .

**Improved Analysis for Majority.** We now switch to designing robust algorithms for families of PCSPs. To begin, we discuss the algorithm used by Brakensiek, Guruswami, Sandeep [BGS23b] for promise templates with Majority polymorphisms. This algorithm was inspired by the algorithm used by Charikar, Makarychev, Makarychev [CMM09] for robust MAX 2-SAT.

For convenience, we relabel the Boolean domain as  $\{-1, +1\}$ . Fix a template  $(P, Q)$  with  $\text{MAJ} \subseteq \text{Pol}(P, Q)$ . Consider an instance of PCSP $(P, Q)$  on variables  $x_1, \dots, x_n$  and clauses  $C_1, \dots, C_m$ . The algorithm begins by solving the Basic SDP for this instance by finding unit vectors  $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^{n+1}$  (with  $\mathbf{v}_0$  representing the “truth” vector) such that the average value of the vector assignment to the clauses is  $1 - \epsilon$ , where  $\epsilon > 0$  is the specified robustness parameter. Next, we sample a random multivariate normal vector  $\mathbf{r} \in \mathcal{N}(0^{n+1}, I_{n+1})$ . Then, for all  $i \in [n]$ , we round  $x_i$  to  $+1$  if  $\langle \mathbf{v}_i, \mathbf{v}_0 + \mathbf{r} \cdot \epsilon^{2/3} \rangle \geq 0$  and  $-1$  otherwise. Here, we improve  $\epsilon^{2/3}$  to  $\sqrt{\epsilon}$  via a new analysis.

We briefly explain the key ideas in Brakensiek, Guruswami, Sandeep [BGS23b] in the analysis of this algorithm. Using a reduction in their paper, we may also assume without loss of generality that  $Q = \{-1, +1\}^k \setminus \{(-1)^k\}$ . In other words, assume that the majority vote of any list of assignments to  $P$  is never all  $-1$ 's. For simplicity, fix a clause  $C_i$  on variables  $x_1, \dots, x_k$  such that the SDP vectors  $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_k$  give a value of  $1 - \epsilon$ . It may be the case that for all  $i \in [n]$ ,  $\langle \mathbf{v}_i, \mathbf{v}_0 \rangle \approx -\Theta(\epsilon)$ , so purely rounding  $\langle \mathbf{v}_i, \mathbf{v}_0 \rangle$  will fail to satisfy any of the clauses. A key observation by BGS is that as long as the vectors have completeness  $1 - \epsilon$ , there exists a probability distribution  $(w_1, \dots, w_k)$  such that  $\sum_{i=1}^k w_i \langle \mathbf{v}_i, \mathbf{v}_0 \rangle \geq -\epsilon$  (see Lemma 4.3). Let  $\mathbf{u} := \sum_{i=1}^k w_i \mathbf{v}_i$ . By concentration, we can assume with probability  $1 - \epsilon^{O(1)}$  that  $|\langle \mathbf{u}, \mathbf{r} \rangle| = O(\log(1/\epsilon))$ . As a key observation, note that since  $\langle \mathbf{u}, \mathbf{v}_0 + \mathbf{r} \cdot \epsilon^{2/3} \rangle = \sum_{i=1}^k w_i \langle \mathbf{v}_i, \mathbf{v}_0 + \mathbf{r} \cdot \epsilon^{2/3} \rangle$ , if all  $k$  variables round to  $-1$ , then  $\langle \mathbf{u}, \mathbf{v}_0 + \mathbf{r} \cdot \epsilon^{2/3} \rangle$  is negative. However,  $\langle \mathbf{u}, \mathbf{v}_0 \rangle \geq -\epsilon$  and the standard deviation of  $\langle \mathbf{u}, \mathbf{r} \cdot \epsilon^{2/3} \rangle$  is at most  $\epsilon^{1/3}$ . Thus, if  $\langle \mathbf{u}, \mathbf{v}_0 + \mathbf{r} \cdot \epsilon^{2/3} \rangle$  is negative, it is barely negative. Hence, this tightly constrains the value of each  $\langle \mathbf{v}_i, \mathbf{v}_0 + \mathbf{r} \cdot \epsilon^{2/3} \rangle$ , which is unlikely due to anti-concentration of the normal distribution.

With a more careful analysis of these rounding probabilities, we can change the rounding threshold to  $\langle \mathbf{v}_i, \mathbf{v}_0 + \mathbf{r} \cdot \sqrt{\epsilon} \rangle$  and get a  $1 - O(\sqrt{\epsilon})$  success probability (soundness). The key idea is, instead of directly comparing each  $\langle \mathbf{u}, \mathbf{r} \rangle$  to the individual distributions  $\langle \mathbf{v}_i, \mathbf{r} \rangle$ , we use a more

careful decomposition of the vectors. In particular, define  $\mathbf{v}_i^\parallel$  to be the component of  $\mathbf{v}_i$  parallel to  $\mathbf{u}$  and let  $\mathbf{v}_i^\perp$  be the component of  $\mathbf{v}_i$  perpendicular to  $\mathbf{u}$ . Since  $\mathbf{v}_i^\parallel$  and  $\mathbf{u}$  are related by a scalar,  $\langle \mathbf{v}_i^\parallel, \mathbf{r} \rangle$  and  $\langle \mathbf{u}, \mathbf{r} \rangle$  are also related by a scalar. However,  $\langle \mathbf{v}_i^\perp, \mathbf{r} \rangle$  is independent of  $\langle \mathbf{u}, \mathbf{r} \rangle$ . Using this observation, we can split our argument into three high-level cases.

First, if  $\|\mathbf{u}\|_2^2 = \Omega(\epsilon \log(1/\epsilon))$ , then  $\langle \mathbf{u}, \mathbf{v}_0 \rangle$  will dominate  $\langle \mathbf{u}, \mathbf{r}\sqrt{\epsilon} \rangle$ , so the chances that  $\langle \mathbf{u}, \mathbf{v}_0 + \mathbf{r}\sqrt{\epsilon} \rangle \leq 0$  are quite small.

Second, if a perpendicular component is large, that is  $\|w_i \mathbf{v}_i^\perp\|_2 = \Omega(1)$  for some  $i \in [k]$ , then even if we condition on  $\langle \mathbf{u}, \mathbf{r} \rangle$ , the value of  $\langle w_i \mathbf{v}_i, \mathbf{v}_0 + \sqrt{\epsilon} \mathbf{r} \rangle$  still has considerable variance. In particular, most likely  $\langle w_i \mathbf{v}_i, \mathbf{v}_0 + \sqrt{\epsilon} \mathbf{r} \rangle$  will either be (1) too positive, in which case  $i$  is rounded correctly, or (2) too negative, in which case the average of  $\sum_{j \neq i} w_j \langle \mathbf{v}_j, \mathbf{v}_0 + \sqrt{\epsilon} \mathbf{r} \rangle$  is positive, so some other  $j$  is rounded correctly.

These two cases themselves are enough to get a  $\sqrt{\epsilon} \log(1/\epsilon)$  loss. To shave the log, in the third and final case, we finely partition the space of potential “bad” outcomes and show that these in total contribute at most  $O(\sqrt{\epsilon})$  loss to the rounding. This is the most technical part of the argument.

**New algorithms for Plurality and Separable Families.** We extend these rounding techniques for Majority to non-Boolean domains and more general rounding functions. To do this, we abstract out the essential feature of the analysis of Majority: the existence of a hyperplane separation between the strong form of the constraint  $P$  and SDP-configurations whose rounding lies outside of the weak form  $Q$ . We define templates  $(P, Q)$  with a generalization of this property that we call *separable families*. The precise description is given in Definition 5.1 but, at a high level, here is the idea. In the Boolean Majority case, our analysis relies upon a linear function separating  $\langle \mathbf{v}_i, \mathbf{v}_0 \rangle$  from an absent tuple of  $Q$ , which in turn can be expressed via the inner product with a weight vector  $w$ . In non-Boolean domains  $D$ , we have a separate vector  $\mathbf{v}_{i,d}$  for each variable  $x_i$  and each domain element  $d \in D$ . Hence, we encode both  $P$  and the SDP-configurations whose rounding lies outside of  $Q$  as two convex bodies living in the matrix space  $\mathbb{R}^{k \times D}$ , where  $k$  is the arity of the constraint. The first is the convex hull of the *one-hot* encodings  $\Pi_{\mathbf{p}}$  of tuples in  $\mathbf{p} \in P$  (where the  $(i, d)$ -th entry equals 1 if  $p_i = d$  and 0 otherwise). The second is the preimage under the given rounding function  $\rho$  of “bad” tuples—those lying outside of  $Q$ . If the polymorphisms of  $(P, Q)$  are rich enough, one can show that these two convex sets are disjoint—in which case, they must admit a hyperplane separation. The latter is naturally expressed via a linear functional over  $\mathbb{R}^{k \times D}$  determined by the Frobenius inner product times a suitable weight matrix  $W$ —the non-Boolean analogue of the weight vector  $w$ . If, for some rounding function  $\rho$ , the template  $(P, Q)$  admits such a separation, we call it  $\rho$ -*separable*.<sup>9</sup> Provided that  $\rho$  satisfies an extra *conservativity* condition (roughly speaking, the winning element of a given distribution must have weight bounded away from zero), we are able to use the hyperplane separation property to show that  $(P, Q)$  is robustly solved by SDP with loss  $O(\sqrt{\epsilon} \log(1/\epsilon))$ , by generalising part of the analysis performed in the Boolean Majority case. Note that the loss we achieve in this setting is slightly worse than the one for Majority, by a  $\log(1/\epsilon)$  factor (although, even for Majority, better than the loss achieved in [BGS23b]). This is due to the fact that the trick of splitting the SDP vectors into parallel and orthogonal components does not carry over in the non-Boolean domain.

A notable example of a separable family is the Unique Games problem. Unique Games has been known to have a robust algorithm for a long time due to the algorithm of Charikar, Makarychev,

<sup>9</sup>Such predicates have some resemblance to the “regional polymorphisms” defined by Brakensiek–Guruswami [BG19].

and Makarychev [CMM06] (although this algorithm is rather different from the one used by the same authors for 2-SAT). The fact that their 2-SAT algorithm can be extended to Unique Games may be of independent interest. The underlying polymorphism driving this is *Plurality*, which takes the most commonly occurring element in a list of domain elements, even if its frequency is much less than  $1/2$ . More interestingly, our polymorphism-based result captures *all* PCSPs admitting Plurality—in particular, the family of so-called SetSAT PCSPs identified in [BWŻ21] as a natural non-Boolean generalization of  $(2 + \epsilon)$ -SAT [AGH17]. Unlike Unique Games, these problems were previously not known to be robustly solvable with any loss.

**Robust Gadget Reductions (Adding Equality).** We now discuss the proofs of Theorem 1.5 and Theorem 1.6. Assuming Theorem 1.5, Theorem 1.6 is straightforward to establish by combining existing gadget reductions for robust (P)CSPs [DK13, BK16, BGS23b] with state-of-the-art gadget reductions for (P)CSPs [BBKO21]. The precise details are worked out in Corollary 6.2.

As such, we focus on sketching the proof of Theorem 1.5. We crucially build off the algorithm of Brown-Cohen and Raghavendra [BR16] (“BCR algorithm”) for solving approximate MAX (P)CSPs.<sup>10</sup> For the purposes of this high-level overview, we assume that the our promise template is a single pair of Boolean relations  $(P, Q)$  with  $P \subseteq Q \subseteq \{-1, 1\}^k$ .

We first describe the essential features of the BCR algorithm. For an instance of PCSP $(P, Q)$  on variable set  $x_1, \dots, x_n$  and clauses  $C_1, \dots, C_m$ , we can think of an SDP solution as a collection of unit vectors  $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^{n+1}$ . After finding an SDP solution with near-optimal value, the BCR algorithm proceeds by sampling two random objects: (i) a list of random vectors  $\mathbf{r}_1, \dots, \mathbf{r}_D \in \mathbb{R}^{n+1}$  sampled from a multivariate normal distribution, and (ii) a rounding function<sup>11</sup>  $H : \mathbb{R}^D \rightarrow [-1, 1]$ . For each  $i \in [n]$ , we then define a fractional assignment  $z_i \in [-1, 1]$  via

$$z_i := H(\mathbf{v}_0 \cdot \mathbf{v}_1 + \mathbf{v}_i^\perp \cdot \mathbf{r}_1, \dots, \mathbf{v}_i \cdot \mathbf{v}_0 + \mathbf{v}_i^\perp \cdot \mathbf{r}_D),$$

where we set  $\mathbf{v}_i^\perp = \mathbf{v}_i - (\mathbf{v}_0 \cdot \mathbf{v}_i)\mathbf{v}_0$ . We then get an integral solution to the PCSP by independently rounding  $x_i$  to  $+1$  with probability  $\frac{1+z_i}{2}$  and  $-1$  otherwise.

Let  $\mathcal{R}$  be the probability distribution over the choices of  $(\mathbf{r}_1, \dots, \mathbf{r}_D)$  and  $H$ . We can thus think of the BCR rounding scheme as a map  $\text{BCR} : \mathbb{R}^{n+1} \times \mathcal{R} \rightarrow [-1, 1]$  for which *global shared randomness*  $R \in \mathcal{R}$  is picked at the start of the algorithm, and then for each  $i \in [n]$ , we set  $z_i := \text{BCR}(\mathbf{v}_i, R)$ . The assumed robustness of the algorithm then translates into the following guarantee.

**Key Property.** For *every* SDP solution  $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n$  with value at least  $1 - \epsilon$ , the rounded assignment  $\text{BCR}(\mathbf{v}_1, R), \dots, \text{BCR}(\mathbf{v}_n, R)$  will satisfy  $1 - f(\epsilon)$  of the constraints of our instance in expectation over the choice of  $R \in \mathcal{R}$ .

As is, the existing scheme may not be robust for PCSP(EQ) for the following reason: given two vectors  $\mathbf{v}_i$  and  $\mathbf{v}_j$  that are  $\delta$  apart in Euclidean distance, the corresponding  $\text{BCR}(\mathbf{v}_i, R)$  and  $\text{BCR}(\mathbf{v}_j, R)$  might be very different for a typical  $R \sim \mathcal{R}$ .<sup>12</sup> In order to make the BCR rounding

<sup>10</sup>Technically, Raghavendra’s theorem and the result of Brown-Cohen–Raghavendra are only stated for CSPs, but as noted by Brakensiek, Guruswami, Sandeep [BGS23b], their arguments extend to PCSPs with minimal modification.

<sup>11</sup>The choice of rounding function is based on the existence of certain *approximate polymorphisms*, see Section 6.1 for a precise definition.

<sup>12</sup>In general  $H$  is very slightly smooth, so one can directly combine the BCR algorithm with a correlated rounding trick to get a robust algorithm for equality. The “catch” is that soundness of the robust algorithm depends on the arity of the approximate polymorphisms considered by BCR. However, no effective bound is given by BCR on the size of these approximate polymorphisms, resulting in guarantees much worse than Theorem 1.5.

scheme also robust for equality, we exploit the Key Property in the following way. Consider a map  $M_\delta : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$  (not necessarily linear) such that, for every unit vector  $\mathbf{v}$ , the distance between  $\mathbf{v}$  and  $M_\delta(\mathbf{v})$  is at most  $\delta$ . We can then consider the following scheme:  $S_\delta(\mathbf{v}, R) := \text{BCR}(M_\delta(\mathbf{v}), R)$ . A key observation is that  $S_\delta$  is still a robust rounding scheme with slightly worse parameters.

To see why, using  $S_\delta$  to round a solution  $\mathbf{v}_1, \dots, \mathbf{v}_n$  is effectively the same as using BCR to round  $M_\delta(\mathbf{v}_1), \dots, M_\delta(\mathbf{v}_n)$ . Since each  $\mathbf{v}_i$  is close in Euclidean distance to  $M_\delta(\mathbf{v}_i)$ , for  $\delta$  sufficiently small, the SDP value of the solution<sup>13</sup>  $M_\delta(\mathbf{v}_1), \dots, M_\delta(\mathbf{v}_n)$  is still approximately  $1 - \epsilon$ , and as such we still satisfy roughly  $1 - f(\epsilon)$  constraints on average.

More generally,  $M_\delta : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$  does not need to be deterministic, rather it can be any randomized map such that the input vector can never be more than  $\delta$  far from the output vector. We call such a randomized map a  $\delta$ -spread if for each unit vector  $\mathbf{v}$ , if the probability distribution  $M_\delta(\mathbf{v})$  is supported within the ball  $B(\mathbf{v}, \delta)$  and the probability degrades smoothly with distance. See Definition 6.27 for a precise definition.

For a given  $\delta$ -spread (a distribution of  $M_\delta$ 's), the corresponding rounding scheme  $S_\delta(\mathbf{v}, R) := \mathbb{E}_{M_\delta}[\text{BCR}(M_\delta(\mathbf{v}), R)]$  is called a  $\delta$ -smoothing of BCR. By the aforementioned logic, any  $\delta$ -smoothing of BCR is still approximately  $(\epsilon, f(\epsilon))$ -robust. In particular, we now have a large collection of rounding schemes that are all robust for PCSP( $P, Q$ ).

Our next step is to pick one of these  $\delta$ -smoothings that is also robust for equality. For each unit vector  $\mathbf{v} \in \mathbb{R}^{n+1}$ , we look at the following  $L^2$  norm:

$$\|S_\delta(\mathbf{v})\|_2 := \sqrt{\mathbb{E}_{R \sim \mathcal{R}}[S_\delta(\mathbf{v}, R)^2]}.$$

Recall that the range of  $S_\delta$  is  $[-1, 1]$ , so  $S_\delta(\mathbf{v}, R)^2$  roughly measures the certainty the rounding scheme has for this value. Rather unintuitively, we select the  $\delta$ -smoothing of BCR such that  $\|S_\delta(\mathbf{v})\|_2$  is *minimized* for all unit vectors  $\mathbf{v} \in \mathbb{R}^{n+1}$ . We call this scheme  $\text{REQ}_\delta$ , as we shall soon see it is Robust for Equality. Roughly speaking,  $\text{REQ}_\delta$  is the  $\delta$ -smoothing of BCR with maximal entropy.

The key lemma we seek to show is that there are (small) constants  $c_1, c_2 \geq 1$  such that if two vectors  $\mathbf{v}$  and  $\mathbf{w}$  are within distance  $\delta^{c_1}$  of each other, then  $\sqrt{\mathbb{E}_{R \sim \mathcal{R}}[(\text{REQ}_\delta(\mathbf{v}, R) - \text{REQ}_\delta(\mathbf{w}, R))^2]} \leq \delta^{c_2}$ . See Lemma 6.31 for a precise statement. The proof of this lemma uses the fact that the possible  $\delta$ -smoothings around  $\mathbf{v}$  are quite similar to the  $\delta$ -smoothings around  $\mathbf{w}$  (see Proposition 6.32). If we think of these spaces of  $\delta$ -smoothings as convex bodies, the  $L^2$  minimizer of one of these convex bodies must then be in close proximity to the  $L^2$  minimizer of the other convex body. This is enough to prove Lemma 6.31.

However,  $\text{REQ}_\delta$  by itself is not a robust rounding scheme for  $(P, Q)$  with equality. The reason why is that so far we have only established that if  $\mathbf{v}_i$  and  $\mathbf{v}_j$  are close, then on average over the choice of  $R \in \mathcal{R}$ , the outputs  $z_i = \text{REQ}_\delta(\mathbf{v}, R)$  and  $z_j = \text{REQ}_\delta(\mathbf{w}, R)$  are close. However, remember that the  $z_i$ 's are rounded into an integral assignment with *independent* coin flips for each  $i \in [n]$ . In particular, if  $z_i = z_j = 0$ , then the equality constraint is satisfied only 1/2 of the time.

To correct this issue, as our final step we modify the independent rounding into *correlated* rounding. In the Boolean setting, this involves picking a random *global* threshold  $t \in [-1/2, 1/2]$  and rounding each  $x_i$  to 1 if  $z_i > t$  and  $-1$  otherwise. Now, if  $z_i \approx z_j$ , the corresponding equality constraint will almost always be satisfied. The tradeoff is that the robustness  $f(\epsilon)$  for PCSP( $P, Q$ )

<sup>13</sup>It may be the case that  $M_\delta(\mathbf{v}_1), \dots, M_\delta(\mathbf{v}_n)$  is no longer a valid SDP solution due to violating triangle inequalities. Circumventing this issue is highly technical and involves adapting a smoothing trick due to Raghavendra and Steurer [RS09]. We ignore this important issue for the purposes of this overview.

gets worse by a constant factor. See Section 6.6 for precise details.

Altogether, our rounding scheme is  $\text{REQ}_\delta$  for  $\delta = \epsilon^{O(1)}$  with correlated rounding. Due to various polynomial losses in the course of the proof, the new scheme is (approximately)  $(\epsilon, f(\epsilon^{1/6}))$ -robust.

### 1.3 Paper Outline

In Section 2, we give necessary background material and notation for understanding our main results. Section 3 gives an integrality gap for robustly solving  $\text{fiPCSP}(1\text{-IN-3-SAT}, \text{NAE-SAT})$ , proving Theorem 1.1. Section 4 gives an asymptotically optimal analysis of Brakensiek–Guruswami–Sandeep’s algorithm for the Majority polymorphism, proving Theorem 1.2. Section 5 generalizes Theorem 1.1 to separable PCSPs, proving in particular Theorem 1.3. Section 6 proves Theorem 1.5 and Theorem 1.6 by adapting the rounding scheme of Brown-Cohen and Raghavendra [BR16]. In Section 7, we give some concluding thoughts and list some open questions.

## 2 Notation and Preliminaries

### 2.1 Promise CSPs

Let  $D, E$  be finite domains. A PCSP *template* is a collection  $\Gamma = \{(P_1, Q_1), \dots, (P_t, Q_t)\}$  of pairs of relations of arities  $k_1, \dots, k_t \in \mathbb{N}$ , where for all  $i \in [t]$   $P_i \subseteq D^{k_i}$  and  $Q_i \subseteq E^{k_i}$ . We further require that there exists a *homomorphism*  $h$  that, for all  $i \in [t]$ , sends tuples in  $P_i$  to tuples in  $Q_i$ ; i.e., a map  $h : D \rightarrow E$  such that  $h(\mathbf{x}) \in Q_i$  for each  $i \in [t]$  and each  $\mathbf{x} \in P_i$ .<sup>14</sup> When  $t = 1$  (i.e., when  $\Gamma$  contains a unique promise relation  $(P, Q)$ ), we shall also write  $\text{PCSP}(P, Q)$  for  $\text{PCSP}(\Gamma)$ . Also, when  $P_i = Q_i$  for all  $i \in [t]$ , we say that  $\Gamma$  is a CSP *template*.

Let  $\Gamma$  be a PCSP template. An *instance* of  $\text{PCSP}(\Gamma)$  consists of a set of variables  $V = \{x_1, \dots, x_n\}$  and a set of constraints  $\mathcal{C} = \{C_1, \dots, C_m\}$ , where each constraint  $C_i$  is described by a tuple of variables  $(x_{i,1}, \dots, x_{i,k_i})$  and a relation pair  $(P_i, Q_i) \in \Gamma$  of arity  $k_i$ . We say that a map  $g : V \rightarrow D$  *strongly* satisfies  $C_i$  if  $(g(x_{i,1}), \dots, g(x_{i,k_i})) \in P_i$ , while we say that a map  $g' : V \rightarrow E$  *weakly* satisfies  $C_i$  if  $(g'(x_{i,1}), \dots, g'(x_{i,k_i})) \in Q_i$ . Observe that composing any map  $g : V \rightarrow D$  with the required homomorphism  $h$  between the  $P_i$ ’s and the  $Q_i$ ’s results in a map  $g' : V \rightarrow E$  that weakly satisfies all constraints that are strongly satisfied by  $g$ .

Consider a function  $f : (0, 1) \rightarrow (0, 1)$  with  $f(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . We say that  $\text{PCSP}(\Gamma)$  is *robustly solvable with loss  $f$*  (in short,  *$f$ -robust*) if there exists an algorithm Alg with the following properties:

- The inputs to Alg are an instance of  $\text{PCSP}(\Gamma)$  and a threshold parameter  $\epsilon \in (0, 1)$ , with the promise that there is a map  $g : V \rightarrow D$  that strongly satisfies at least  $1 - \epsilon$  fraction of the constraints;
- Alg outputs the description of a map  $g' : V \rightarrow E$  that weakly satisfies at least  $1 - f(\epsilon)$  fraction of the constraints;
- Alg runs in polynomial time on the size of the instance.

<sup>14</sup>At times, it will be more convenient to denote a PCSP template  $\Gamma$  as a pair  $(\mathbf{A}, \mathbf{B})$  of homomorphic *relational structures*, where  $\mathbf{A}$  has domain  $A$  and relations  $P_1, \dots, P_t$  over  $A$ , and  $\mathbf{B}$  has domain  $B$  and relations  $Q_1, \dots, Q_t$  over  $B$ . We shall adopt both notations interchangeably.

We shall sometimes consider *weighted* instances of  $\text{PCSP}(\Gamma)$ , where the constraint set  $\mathcal{C}$  is equipped with a probability distribution  $w : \mathcal{C} \rightarrow \mathbb{R}^+$ . In this case, the definition of robust solvability of  $\text{PCSP}(\Gamma)$  is entirely analogous, except that it considers the total weight—rather than the fraction—of satisfied constraints.

## 2.2 Basic SDP

Let  $D$  be a finite domain, let  $P \subseteq D^k$ , and let  $\epsilon \in [0, 1]$ . Let  $N \in \mathbb{N}$  be a positive integer. We say that a tuple of  $|D|k + 1$  vectors in  $\mathbb{R}^N$ ,  $(\mathbf{v}_0, \{\mathbf{v}_{1,d} : d \in D\}, \dots, \{\mathbf{v}_{k,d} : d \in D\})$  is a  $1 - \epsilon$  *approximate vector assignment* to  $P$  if there exists a probability distribution  $\mu$  on  $D^k$  with the following properties.

$$\mu(P) \geq 1 - \epsilon, \tag{1}$$

$$\forall i \in [k], d \in D, \langle \mathbf{v}_0, \mathbf{v}_{i,d} \rangle = \sum_{\mathbf{a} \in D^k} \mathbf{1}[a_i = d] \mu(\mathbf{a}), \tag{2}$$

$$\forall i, i' \in [k], d, d' \in D, \langle \mathbf{v}_{i,d}, \mathbf{v}_{i',d'} \rangle = \sum_{\mathbf{a} \in D^k} \mathbf{1}[a_i = d] \mathbf{1}[a_{i'} = d'] \mu(\mathbf{a}). \tag{3}$$

If  $D = \{-1, 1\}$ , this is equivalent to looking at  $\mathbf{v}_i := \mathbf{v}_{i,1} - \mathbf{v}_{i,-1}$  subject to the following constraints (see, e.g., [BGS23b] for a justification)

$$\mu(P) \geq 1 - \epsilon, \tag{4}$$

$$\forall i \in [k], \langle \mathbf{v}_0, \mathbf{v}_i \rangle = \sum_{\mathbf{a} \in D^k} a_i \mu(\mathbf{a}), \tag{5}$$

$$\forall i, i' \in [k], \langle \mathbf{v}_i, \mathbf{v}_{i'} \rangle = \sum_{\mathbf{a} \in D^k} a_i a_{i'} \mu(\mathbf{a}). \tag{6}$$

Given an instance of  $\text{CSP}(P)$  on variables  $x_1, \dots, x_n$  and clauses  $C_1, \dots, C_m$ , we say that a tuple of  $|D|n + 1$  vectors  $(\mathbf{v}_0, \{\mathbf{v}_{1,d} : d \in D\}, \dots, \{\mathbf{v}_{n,d} : d \in D\})$  is a basic SDP solution if for every  $j \in [m]$ , there is a probability distribution  $\mu_j$  over  $D^k$  such that (2) and (3) (or equivalently, (5) and (6) in the Boolean setting) hold for the  $|D|k + 1$  tuples of vectors corresponding to the variables  $x_{j,1}, \dots, x_{j,k_j}$  in the clause  $C_j$ . The value (or completeness) of the SDP solution is equal to  $\frac{1}{m} \sum_{j=1}^m \mu_j(P)$ . Note that for any  $\epsilon > 0$ , one can compute the optimal value of the SDP up to an additive  $\epsilon$  in  $(n/\epsilon)^{O(1)}$  time [GLS93, GM12].<sup>15</sup>

When we refer to the SDP relaxation of an instance of  $\text{PCSP}(P, Q)$ , we mean the SDP relaxation of the instance as an instance of  $\text{CSP}(P)$ . The predicate  $Q$  only appears in the *rounding* of the SDP relaxation, where the continuous vectors are converted into a discrete assignment.

Given a  $\text{PCSP}$  template  $\Gamma$  and a weighted instance of  $\text{PCSP}(\Gamma)$ , we define the SDP relaxation of an instance of  $\text{PCSP}(\Gamma)$  analogously, where each constraint  $C_j$  has the constraints (2) and (3) corresponding to  $P_j$  (and the variables of  $C_j$ ). Furthermore, the value of the CSP is  $\sum_{j=1}^m w(C_j) \mu_j(P_j)$ .

<sup>15</sup>Brakensiek, Guruswami, and Sandeep [BGS23b] note that finding such a solution is somewhat subtle as the constraints (5) and (6) need to be exactly satisfied. See Appendix A of the full version of their paper for a more detailed discussion.

### 2.3 Polymorphisms

Given a pair of predicates  $P \subseteq Q \subseteq D^k$ , we say that an operator  $f : D^L \rightarrow D$  is a *polymorphism* of  $(P, Q)$  if for any  $t^1, \dots, t^L \in P$ , we have that  $t \in Q$ , where  $t_i = f(t_i^1, \dots, t_i^L)$  for all  $i \in [k]$ . We let  $\text{Pol}(P, Q)$  denote the set of all polymorphisms of  $(P, Q)$ , with  $\text{Pol}^L(P, Q)$  being specifically the set of polymorphisms of arity  $L$ .

In this paper, we focus on a couple commonly-studied families of polymorphisms. For any odd  $L \in \mathbb{N}$ , let  $\text{MAJ}_L, \text{AT}_L : \{-1, 1\}^L \rightarrow \{-1, 1\}$  be defined as

$$\begin{aligned} \text{MAJ}_L(x_1, \dots, x_L) &:= \mathbf{1} \left[ \sum_{i=1}^L x_i \geq 0 \right], \\ \text{AT}_L(x_1, \dots, x_L) &:= \mathbf{1} \left[ \sum_{i=1}^L (-1)^{i-1} x_i \geq 0 \right]. \end{aligned}$$

We let  $\text{MAJ} := \{\text{MAJ}_L : L \in \mathbb{N} \text{ odd}\}$  and  $\text{AT} := \{\text{AT}_L : L \in \mathbb{N} \text{ odd}\}$ . For example,  $\text{MAJ} \subseteq \text{Pol}(P, Q)$  means that  $\text{Pol}(P, Q)$  has majority polymorphisms of all odd arities and in fact has all folded (i.e., unbiased) weighted threshold functions as polymorphisms [BGS23a].

### 2.4 Analytical facts

We let  $\mathcal{N}(a, b)$  denote the normal distribution with mean  $a$  and variance of  $b$  (i.e., standard deviation of  $\sqrt{b}$ ). Likewise, given  $\boldsymbol{\mu} \in \mathbb{R}^n$  and  $\Sigma \in \mathbb{R}^{n \times n}$ , we let  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$  denote the  $n$ -variate normal distribution with mean  $\boldsymbol{\mu}$  and covariance  $\Sigma$ . We cite the following standard facts about the normal distribution.

**Proposition 2.1** (e.g., [Ver18], [BGS23b]). *Let  $X$  be distributed as  $\mathcal{N}(\mu, \sigma^2)$ . Then,*

- (a) *For any interval  $[a, b] \subseteq \mathbb{R}$ ,  $\Pr[X \in [a, b]] \leq \frac{b-a}{2\sigma}$ .*
- (b) *For any  $t \geq 0$ ,  $\Pr[X \geq t + \mu] \leq e^{-\frac{t^2}{2\sigma^2}}$ .*

**Proposition 2.2** (e.g., [Ver18], [BGS23b]). *Let  $x$  be a Gaussian variable with mean zero and variance one. Then for all  $t \geq 0$*

1.  $P(x \geq t) \leq \frac{1}{2}e^{-\frac{t^2}{2}}$ .
2.  $P(t \leq x \leq 2t) \leq \frac{t}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$ .
3. For all  $a \in \mathbb{R}$ ,  $P(|x - a| \leq t) \leq \frac{2t}{\sqrt{2\pi}} \leq t$ .
4. Given  $I \subseteq [t, \infty]$ ,  $P(x \in I) \leq |I|e^{-\frac{t^2}{2}}$ .

*Proof.* Observe that  $P(x \geq t) = \int_t^\infty \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \int_0^\infty \frac{e^{-\frac{(y+t)^2}{2}}}{\sqrt{2\pi}} dy \leq e^{-\frac{t^2}{2}} \int_0^\infty \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy = \frac{1}{2}e^{-\frac{t^2}{2}}$ ,  $P(t \leq x \leq 2t) = \int_t^{2t} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \leq \frac{t}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$ , and  $P(|x - a| \leq t) = \int_{a-t}^{a+t} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \leq \frac{2t}{\sqrt{2\pi}}$ .  $\square$

We shall also use the following concentration bound for the squared norm of matrices  $X \sim \mathcal{N}(\mathbf{0}, I_n)$ , which follows from Laurent-Massart bounds on the concentration of the chi-squared distribution.

**Proposition 2.3** ([LM00]). *Let  $X$  be distributed as  $X \sim \mathcal{N}(0_n, I_n)$ . Then, for each  $\alpha \in \mathbb{R}$ , it holds that*

$$\Pr(\|X\|^2 \in [n - 2n^{1/2+\alpha}, n + 2n^{1/2+\alpha} + 2n^\alpha]) \geq 1 - e^{-n^\alpha}.$$

### 3 Integrality Gap for AT Polymorphism

#### 3.1 Attempted Integrality Gap Instance

Before giving our full analysis, we first describe a simple attempted integrality gap instance which fails but captures the intuition for our integrality gap instance.

Let  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$  be three unit vectors that are orthogonal to  $\mathbf{v}_0$  such that  $\mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3 = 0$ . For example, we could have  $\mathbf{v}_0 = (1, 0, 0)$ ,  $\mathbf{w}_1 = (0, 1, 0)$ ,  $\mathbf{w}_2 = (0, -\frac{1}{2}, \frac{\sqrt{3}}{2})$ , and  $\mathbf{w}_3 = (0, -\frac{1}{2}, -\frac{\sqrt{3}}{2})$ .

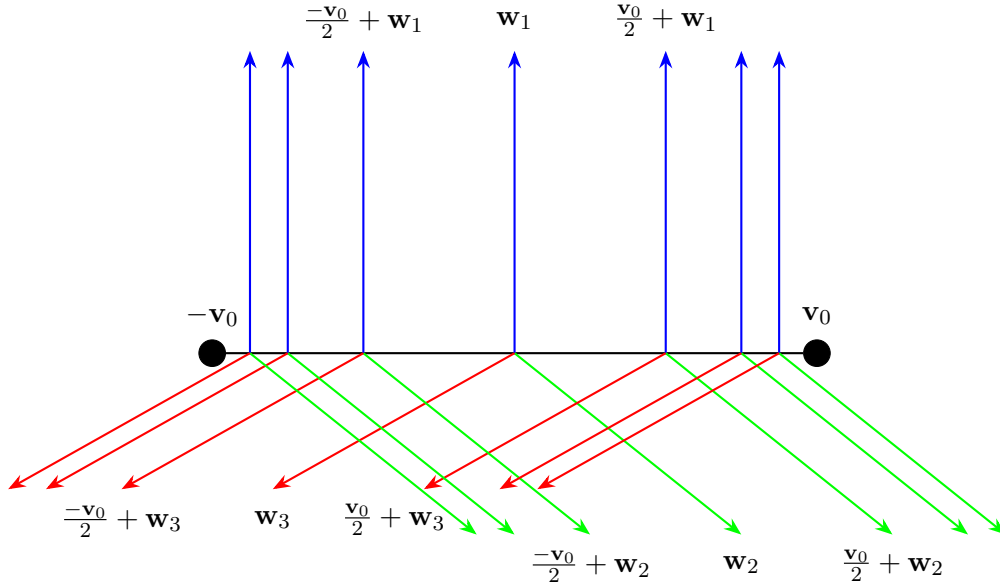
Let  $k_0 = \lceil \log_2(n) \rceil$ . Let us pretend that we have the vectors

$$\{(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_i : k \in \{0, 1, \dots, k_0\}, i \in [3]\} \cup \{-(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_i : k \in \{1, 2, \dots, k_0\}, i \in [3]\}$$

and moreover, one of the following two cases holds

1.  $(1 - 2^{-k_0})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to 1 for all  $i \in [3]$  and  $-(1 - 2^{-k_0})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to  $-1$  for all  $i \in [3]$ .
2.  $(1 - 2^{-k_0})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to  $-1$  for all  $i \in [3]$  and  $-(1 - 2^{-k_0})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to 1 for all  $i \in [3]$ .

See the figure below for an illustration of this set of vectors when  $k_0 = 3$ .



If so, then consider the triples

$$\{(b(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_i, b(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_{i'}, -b(1 - 2^{1-k})\mathbf{v}_0 + \mathbf{w}_{i''})\}$$

where  $k \in [k_0]$ ,  $b \in \{-1, 1\}$  and  $i, i', i''$  are distinct elements of  $[3]$ . The vectors in each such triple sum to either  $\mathbf{v}_0$  or  $-\mathbf{v}_0$  so for each such triple, we can add a 1-IN-3-SAT constraint on either the vectors or their negations. However, it is not hard to show that at least one of the resulting NAE-SAT constraints must be violated.

To see this, assume that all of the NAE constraints are satisfied. We can assume without loss of generality that  $(1 - 2^{-k_0})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to 1 for all  $i \in [3]$  and  $-(1 - 2^{-k_0})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to  $-1$  for all  $i \in [3]$ . We now make the following observations:

1.  $(1 - 2^{-k_0})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to 1 for all  $i \in [3]$ .

2. Since we have NAE-SAT constraints on the triples

$$((-1)^{k_0-k}(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_i, (-1)^{k_0-k}(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_{i'}, (-1)^{k_0-k+1}(1 - 2^{1-k})\mathbf{v}_0 + \mathbf{w}_{i''})$$

such that  $i, i', i''$  are distinct elements of  $[3]$ , if  $(-1)^{k_0-k}(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to  $(-1)^{k_0-k}$  for all  $i \in [3]$  then  $(-1)^{k_0-k+1}(1 - 2^{1-k})\mathbf{v}_0 + \mathbf{w}_i$  must be rounded to  $(-1)^{k_0-k+1}$  for all  $i \in [3]$ .

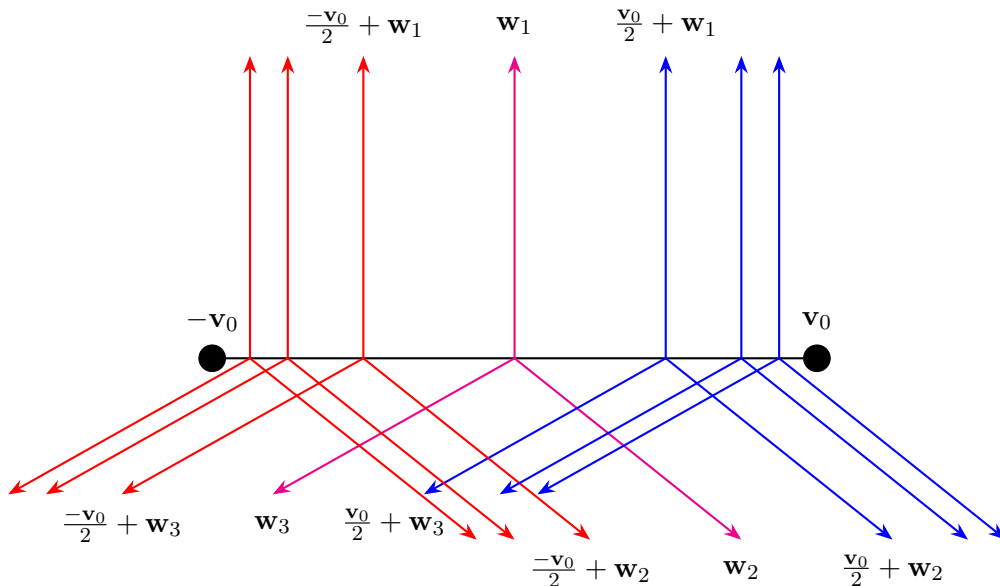
3.  $-(1 - 2^{-k_0})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to  $-1$  for all  $i \in [3]$ .

4. Since we have NAE-SAT constraints on the triples

$$((-1)^{k_0-k+1}(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_i, (-1)^{k_0-k+1}(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_{i'}, (-1)^{k_0-k}(1 - 2^{1-k})\mathbf{v}_0 + \mathbf{w}_{i''})$$

such that  $i, i', i''$  are distinct elements of  $[3]$ , if  $(-1)^{k_0-k+1}(1 - 2^{-k})\mathbf{v}_0 + \mathbf{w}_i$  is rounded to  $(-1)^{k_0-k+1}$  for all  $i \in [3]$  then  $(-1)^{k_0-k}(1 - 2^{1-k})\mathbf{v}_0 + \mathbf{w}_i$  must be rounded to  $(-1)^{k_0-k}$  for all  $i \in [3]$ .

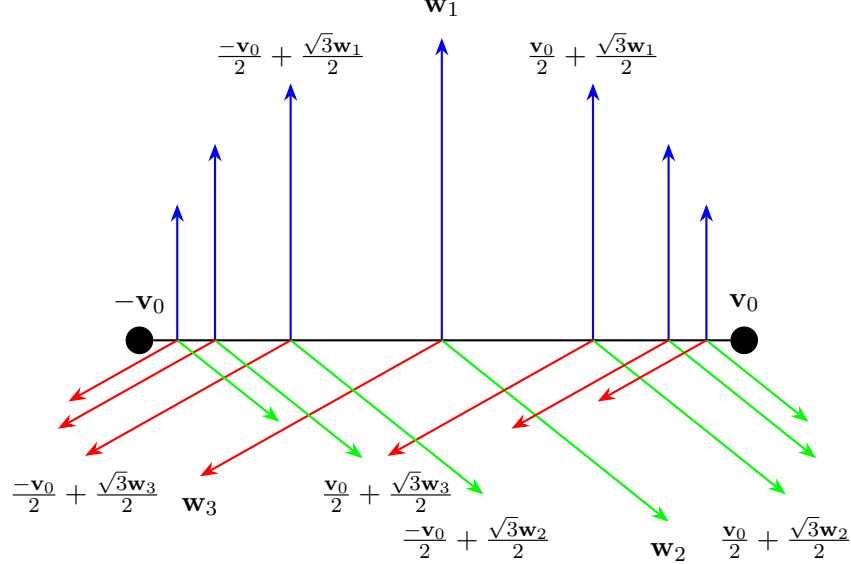
Putting these observations together, we have that for all  $k \in [k_0 + 1]$ ,  $(-1)^{k_0-k+1}(1 - 2^{1-k})\mathbf{v}_0 + \mathbf{w}_i$  must be rounded to  $(-1)^{k_0-k+1}$  for all  $i \in [3]$  and  $(-1)^{k_0-k}(1 - 2^{1-k})\mathbf{v}_0 + \mathbf{w}_i$  must be rounded to  $(-1)^{k_0-k}$  for all  $i \in [3]$ . Plugging in  $k = 1$ , we have that  $\mathbf{w}_1, \mathbf{w}_2$ , and  $\mathbf{w}_3$  must all be rounded to both 1 and  $-1$  which gives a contradiction. This argument is illustrated in the figure below where blue represents vectors which should be rounded to 1 and red represents vectors which should be rounded to  $-1$ .



The problem with this attempted integrality gap instance is that these vectors are not unit vectors. In order to have unit vectors, we would need to have the vectors

$$\{b(1 - 2^{-k})\mathbf{v}_0 + \sqrt{2^{1-k} - 2^{-2k}}\mathbf{w}_i : k \in \{0, 1, \dots, k_0\}, b \in \{-1, 1\}, i \in [3]\},$$

which are illustrated in the figure below. However, we can no longer have 1-IN-3-SAT constraints on these vectors. To handle this, we will consider vectors of the form  $b(1 - 2^{-k})\mathbf{v}_0 + \sqrt{2^{1-k} - 2^{-2k}}\mathbf{w}$  where  $b \in \{-1, 1\}$  and  $\|\mathbf{w}\| \approx 1$  and show that we can use a similar argument.



### 3.2 Continuous Integrality Gap Instance

We now describe a natural modification of the failed integrality gap instance described above which gives a candidate integrality gaps instance with infinitely many variables and constraints. We will then further modify this integrality gap instance to discretize it and shift to multivariate Gaussian vectors rather than unit vectors as multivariate Gaussian vectors are easier to analyze and (with the appropriate scaling) are close to unit vectors when the dimension is sufficiently high.

Let  $k_0 = \lceil \log(\frac{1}{\epsilon}) \rceil + 20$ . We will work in  $\mathbb{R}^{d+1}$  where  $\mathbf{v}_0 = \mathbf{e}_{d+1}$  is orthogonal to  $\mathbb{R}^d$  and  $d$  will be chosen later.

**Definition 3.1.** Let  $A_x$  be the set of points  $\{x\mathbf{v}_0 + \sqrt{1 - x^2}\mathbf{w} : \mathbf{w} \in S^{d-1}\}$ .

**Definition 3.2.** For each  $x$  of the form  $1 - 2^{-k}$  where  $k \in [k_0]$ ,

1. We define  $T_x$  to be the set of triples of points  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$  such that  $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = 0$ ,  $\mathbf{p}_1, \mathbf{p}_2 \in A_x$ , and  $\mathbf{p}_3 \in A_{1-2x}$ .
2. Similarly, we define  $T_{-x}$  to be the set of triples of points  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$  such that  $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = 0$ ,  $\mathbf{p}_1, \mathbf{p}_2 \in A_{-x}$ , and  $\mathbf{p}_3 \in A_{2x-1}$ .

Our candidate integrality gap instance is as follows:

1. For each  $k \in [k_0]$ , letting  $x = 1 - 2^{-k}$ , we choose a random triple from  $T_x$  with probability  $\frac{1}{2^{(k_0+1)}}$  and we choose a random triple from  $T_{-x}$  with probability  $\frac{1}{2^{(k_0+1)}}$ .
2. Let  $x_0 = 1 - 2^{-k_0}$ . With probability  $\frac{1}{2^{(k_0+1)}}$ , we choose a random  $\mathbf{p} \in A_{x_0}$  and take the triple  $(-v_0, -v_0, \mathbf{p})$ . Similarly, with probability  $\frac{1}{2^{(k_0+1)}}$ , we choose a random  $\mathbf{p} \in A_{-x_0}$  and take the triple  $(v_0, v_0, \mathbf{p})$ .

Intuitively, this is a gap instance because the SDP thinks that almost all of the constraints can be satisfied but in any actual assignment, one of the following must occur which leads to at least  $\frac{\delta}{16^{(k_0+1)}}$  of the NAE-SAT constraints being violated where  $\delta > 0$  is an absolute constant which we will choose later. Without loss of generality, we can assume that  $\mathbf{v}_0$  is rounded to 1 and  $-\mathbf{v}_0$  is rounded to  $-1$ .

1. A  $\delta$  proportion of  $\mathbf{p} \in A_{x_0}$  are assigned  $-1$  which means that  $\delta$  of the triples  $(-v_0, -v_0, \mathbf{p})$  are all  $-1$ .
2. A  $\delta$  proportion of  $\mathbf{p} \in A_{-x_0}$  are assigned 1 which means that  $\delta$  of the triples  $(v_0, v_0, \mathbf{p})$  are all 1.
3. For some  $k \in [k_0]$ , letting  $x = 1 - 2^{-k}$ , less than  $\delta$  of the vectors in  $A_x$  are assigned  $-1$  and at least  $\delta$  of the vectors in  $A_{1-2x}$  are assigned 1 which implies that at least  $\frac{\delta}{8}$  of the triples  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3) \in T_x$  are all 1.
4. For some  $k \in [k_0]$ , letting  $x = 1 - 2^{-k}$ , less than  $\delta$  of the vectors in  $A_{-x}$  are assigned 1 and at least  $\delta$  of the vectors in  $A_{1-2x}$  are assigned  $-1$  which implies that at least  $\frac{\delta}{8}$  of the triples  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3) \in T_{-x}$  are all  $-1$ .

Before we modify our integrality gap instance and make this intuition rigorous, we give a more precise description of the triples of points in  $T_x$ .

**Proposition 3.3.** *If  $\mathbf{p}_1, \mathbf{p}_2 \in A_x$  and  $\mathbf{p}_3 \in A_{1-2x}$  then  $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = \mathbf{v}_0$  if and only if*  

$$\mathbf{p}_3 = (1 - 2x)\mathbf{v}_0 + \sqrt{1 - (2|x| - 1)^2}\mathbf{w}, \mathbf{p}_1 = x\mathbf{v}_0 + \sqrt{1 - x^2} \left( -\frac{\sqrt{1 - (2|x| - 1)^2}}{2\sqrt{1 - x^2}}\mathbf{w} + \sqrt{1 - \frac{1 - (2|x| - 1)^2}{4(1 - x^2)}}\mathbf{u} \right),$$
*and  $\mathbf{p}_2 = x\mathbf{v}_0 + \sqrt{1 - x^2} \left( -\frac{\sqrt{1 - (2|x| - 1)^2}}{2\sqrt{1 - x^2}}\mathbf{w} - \sqrt{1 - \frac{1 - (2|x| - 1)^2}{4(1 - x^2)}}\mathbf{u} \right)$  for some  $\mathbf{w}, \mathbf{u} \in S^{d-1}$  such that  $\mathbf{w} \cdot \mathbf{u} = 0$ .*

*Proof.* Assume that  $\mathbf{p}_1, \mathbf{p}_2 \in A_x$ ,  $\mathbf{p}_3 \in A_{1-2x}$ , and  $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = \mathbf{v}_0$ . Since  $\mathbf{p}_3 \in A_{1-2x}$ ,  $\mathbf{p}_3 = (1 - 2x)\mathbf{v}_0 + \sqrt{1 - (2|x| - 1)^2}\mathbf{w}$  for some  $\mathbf{w} \in S^{d-1}$ . Now write  $\mathbf{p}_1 = x\mathbf{v}_0 + a\mathbf{w} + b\mathbf{u}$  where  $\mathbf{u} \in S^{d-1}$ ,  $\mathbf{w} \cdot \mathbf{u} = 0$ , and  $b \geq 0$ . We now make the following observations:

1. Since  $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = \mathbf{v}_0$ ,  $\mathbf{p}_2 = x\mathbf{v}_0 - \left( a + \sqrt{1 - (2|x| - 1)^2} \right) \mathbf{w} - b\mathbf{u}$ .
2. Since  $\|\mathbf{p}_1\|^2 = \|\mathbf{p}_2\|^2$ ,  $a^2 = \left( a + \sqrt{1 - (2|x| - 1)^2} \right)^2$  which implies that  $a = -\frac{1}{2}\sqrt{1 - (2|x| - 1)^2}$ .
3. Since  $\|\mathbf{p}_1\|^2 = 1$ ,  $b = \sqrt{1 - x^2 - a^2} = \sqrt{1 - x^2 - \frac{1 - (2|x| - 1)^2}{4}}$ .

The other direction is easy to verify directly. □

Based on this, we define the following noise coefficient  $\rho_x$ .

**Definition 3.4.** We define  $\rho_x = \frac{\sqrt{1-(2|x|-1)^2}}{2\sqrt{1-x^2}}$ .

Observe that if  $|x| = 1 - 2^{-k}$  and  $k$  is large then  $\rho_x = \frac{\sqrt{1-(2|x|-1)^2}}{2\sqrt{1-x^2}} \approx \frac{\sqrt{2^{2-k}}}{2\sqrt{2^{1-k}}} = \frac{1}{\sqrt{2}}$ .

### 3.3 Discretized Integrality Gap Instance

We now modify our candidate integrality gap instance by having it use multivariate Gaussian vectors rather than unit vectors and discretizing it.

**Definition 3.5.** Given  $x \in [-1, 1]$ , we define  $A'_x$  to be the distribution of vectors obtained by sampling  $\mathbf{w} \sim \mathcal{N}(0, 1/d)^d$  and taking the vector  $x\mathbf{v}_0 + \sqrt{1-x^2}\mathbf{w}$ .

**Definition 3.6.** Given  $x \in [-1, 1]$ , we define  $T'_x$  to be the distribution of triples of points  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$  obtained by sampling  $\mathbf{u}, \mathbf{w} \sim \mathcal{N}(0, 1/d)^d$  and then taking  $\mathbf{p}_3 = (1-2x)\mathbf{v}_0 + \sqrt{1-(2|x|-1)^2}\mathbf{w}$ ,

$$\mathbf{p}_1 = x\mathbf{v}_0 + \sqrt{1-x^2} \left( -\frac{\sqrt{1-(2|x|-1)^2}}{2\sqrt{1-x^2}}\mathbf{w} + \sqrt{1-\frac{1-(2|x|-1)^2}{4(1-x^2)}}\mathbf{u} \right), \text{ and}$$

$$\mathbf{p}_2 = x\mathbf{v}_0 + \sqrt{1-x^2} \left( -\frac{\sqrt{1-(2|x|-1)^2}}{2\sqrt{1-x^2}}\mathbf{w} - \sqrt{1-\frac{1-(2|x|-1)^2}{4(1-x^2)}}\mathbf{u} \right).$$

Similarly, we define  $T'_{-x}$  to be the distribution of triples of points  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$  obtained by sampling  $\mathbf{u}, \mathbf{w} \sim \mathcal{N}(0, 1/d)^d$  and then taking  $\mathbf{p}_3 = (2x-1)\mathbf{v}_0 + \sqrt{1-(2|x|-1)^2}\mathbf{w}$ ,

$$\mathbf{p}_1 = -x\mathbf{v}_0 + \sqrt{1-x^2} \left( -\frac{\sqrt{1-(2|x|-1)^2}}{2\sqrt{1-x^2}}\mathbf{w} + \sqrt{1-\frac{1-(2|x|-1)^2}{4(1-x^2)}}\mathbf{u} \right), \text{ and}$$

$$\mathbf{p}_2 = -x\mathbf{v}_0 + \sqrt{1-x^2} \left( -\frac{\sqrt{1-(2|x|-1)^2}}{2\sqrt{1-x^2}}\mathbf{w} - \sqrt{1-\frac{1-(2|x|-1)^2}{4(1-x^2)}}\mathbf{u} \right).$$

Our modified integrality gap instance is as follows:

1. For each  $k \in [k_0]$ , letting  $x = 1 - 2^{-k}$ , we sample a triple from  $T'_x$  with probability  $\frac{1}{2(k_0+1)}$  and we sample a triple from  $T'_{-x}$  with probability  $\frac{1}{2(k_0+1)}$ .
2. Let  $x_0 = 1 - 2^{-k_0}$ . With probability  $\frac{1}{2(k_0+1)}$ , we sample a vector  $\mathbf{p}$  from  $A'_{x_0}$  and take the triple  $(-v_0, -v_0, \mathbf{p})$ . Similarly, with probability  $\frac{1}{2(k_0+1)}$ , we sample a vector  $\mathbf{p}$  from  $A'_{-x_0}$  and take the triple  $(v_0, v_0, \mathbf{p})$ .
3. We discard any triples such that one of the following is true:
  - (a) When we sampled  $\mathbf{w}$  from  $\mathcal{N}(0, 1/d)^d$ ,  $|(\|\mathbf{w}\|^2 - 1)| > \frac{\epsilon}{1000}$ .
  - (b) When we sampled  $\mathbf{u}$  from  $\mathcal{N}(0, 1/d)^d$ , either  $|(\|\mathbf{u}\|^2 - 1)| > \frac{\epsilon}{1000}$  or  $|\mathbf{w} \cdot \mathbf{u}| > \frac{\epsilon}{1000}$ .
4. We discretize the instance by partitioning the set of vectors in  $\mathbb{R}^{d+1}$  of length at most 2 into regions of radius at most  $\frac{\epsilon}{1000}$  and choosing a center for each region which is within  $\frac{\epsilon}{1000}$  of all points in the region. We also choose a center for the region with vectors of length greater than 2 (we can choose any point for this center as only discarded triples of points will have a vector in this region). We then replace all of the vectors with the center of the region they are contained in.

**Theorem 3.7.** For all  $\epsilon > 0$ , if  $d \geq \frac{80000000}{\epsilon^3}$  then for the integrality gap instance described above (Section 3.3),

1. There is a solution to the SDP that gives a value of at least  $1 - \epsilon$  to each constraint that is not discarded.
2. At most  $6e^{-\frac{10}{\epsilon}}$  of the constraints that were sampled are discarded.
3. All assignments violate at least  $\frac{1}{10^{60} \lceil \log(\frac{1}{\epsilon}) \rceil}$  of the sampled NAE-SAT constraints (this includes constraints that were discarded).

By Raghavendra's theorem [Rag08], this integrality gap immediately translates into UGC hardness.

**Corollary 3.8.** Assuming UGC, for all  $\epsilon$  such that  $0 < \epsilon \leq \frac{1}{100}$ ,  $\text{fiPCSP}(1\text{-IN-3-SAT}, \text{NAE-SAT})$  is not  $(\epsilon, \frac{1}{10^{61} \log(1/\epsilon)})$ -robust.

*Proof sketch for the first statement of Theorem 3.7:* For each vector  $\mathbf{v}$  that is a center of a region that is not the outer region, we can take the vector  $\mathbf{v}' = \sqrt{1 - \frac{\epsilon}{4}} \frac{\mathbf{v}}{\|\mathbf{v}\|} + \frac{\sqrt{\epsilon}}{2} \mathbf{z}_v$  for some vector  $\mathbf{z}_v$  which is orthogonal to everything else. We assign  $\mathbf{v}_0$  to itself.

To show that this gives a valid SDP solution which has a value of at least  $1 - \epsilon$  for each constraint which is not discarded, we first observe that for each triple of unit vectors  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  such that  $\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 = \pm \mathbf{v}_0$ , there is a probability distribution of  $D$  of satisfying assignments to the corresponding constraint such that

1. For all  $i \in [3]$ ,  $\mathbb{E}_D[x_i] = \mathbf{v}_i \cdot \mathbf{v}_0$ .
2. For all distinct  $i, j \in [3]$ ,  $\mathbb{E}_D[x_i x_j] = \mathbf{v}_i \cdot \mathbf{v}_j$ .

We then observe that if  $\mathbf{v}'_1, \mathbf{v}'_2$ , and  $\mathbf{v}'_3$  are unit vectors which are close to  $\mathbf{v}_1, \mathbf{v}_2$ , and  $\mathbf{v}_3$  then there is a pseudo-distribution  $D'$  (which may give negative probabilities to some assignments  $(x_1, x_2, x_3) \in \{-1, 1\}^3$ ) which is close to  $D$  such that

1. For all  $i \in [3]$ ,  $\mathbb{E}_{D'}[x_i] = \mathbf{v}'_i \cdot \mathbf{v}_0$ .
2. For all distinct  $i, j \in [3]$ ,  $\mathbb{E}_{D'}[x_i x_j] = \mathbf{v}'_i \cdot \mathbf{v}'_j$ .

Finally, we observe that replacing each vector  $\mathbf{v}$  with  $\sqrt{1 - \frac{\epsilon}{4}} \frac{\mathbf{v}}{\|\mathbf{v}\|} + \frac{\sqrt{\epsilon}}{2} \mathbf{z}_v$  modifies each pseudo-distribution  $D'$  by replacing each probability  $p$  with  $(1 - \frac{\epsilon}{4})p + \frac{\epsilon}{32}$  which is sufficient to make all of the probabilities non-negative.  $\square$

The full details of the proof of the first statement of Theorem 3.7 are given in Appendix A.

*Proof of the second statement of Theorem 3.7:* We can prove the second statement of Theorem 3.7 using Chernoff bounds. For completeness, we include a proof of these bounds in Appendix A.

**Lemma 3.9.** The following tail bounds hold:

1. For all  $t \in [0, 1]$ , if  $\mathbf{w} \sim \mathcal{N}(0, 1/d)^d$  then  $P(|\|\mathbf{w}\|^2 - 1| \geq t) \leq 2e^{-\frac{dt^2}{8}}$ .

2. Given a vector  $\mathbf{w}$ , for all  $t \geq 0$ , if  $\mathbf{u} \sim \mathcal{N}(0, 1/d)^d$  is independent of  $\mathbf{w}$  then  $P(|\mathbf{w} \cdot \mathbf{u}| \geq t) \leq 2e^{-\frac{dt^2}{2\|\mathbf{w}\|^2}}$ .

By Lemma 3.9, since  $d \geq \frac{80000000}{\epsilon^3}$ , if  $\mathbf{w}$  and  $\mathbf{u}$  are drawn independently from  $\mathcal{N}(0, 1/d)^d$  then

1.  $P(|(\|\mathbf{w}\|^2 - 1)| \geq \frac{\epsilon}{1000}) = P(|(\|\mathbf{u}\|^2 - 1)| \geq \frac{\epsilon}{1000}) \leq 2e^{-\frac{10}{\epsilon}}$
2. Whenever  $\|\mathbf{w}\| \leq 2$ ,  $P(|\mathbf{w} \cdot \mathbf{u}| \geq \frac{\epsilon}{1000}) \leq 2e^{-\frac{10}{\epsilon}}$ .

so the proportion of discarded constraints is at most  $6e^{-\frac{10}{\epsilon}}$ .  $\square$

### 3.4 Soundness Analysis

We now prove the third statement of Theorem 3.7. To do this, we view an assignment to the points in our integrality gap instance as a map  $f : \mathbb{R}^{d+1} \rightarrow \{-1, 1\}$  where  $f(\mathbf{v})$  is the value given to the point corresponding to the center of the region  $\mathbf{v}$  is contained in.

**Definition 3.10.** Given a map  $f : \mathbb{R}^{d+1} \rightarrow \{-1, 1\}$ , for each vector  $\mathbf{w} \in \mathbb{R}^d$ , we define  $f_x(\mathbf{w})$  to be the value given to  $\mathbf{p} = x\mathbf{v}_0 + \sqrt{1-x^2}\mathbf{w}$ .

**Definition 3.11.** Given  $\rho \in [-1, 1]$  and a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , we define

$$(N_\rho f)(\mathbf{x}) = E_{\mathbf{y} \sim \mathcal{N}(0, 1/d)^d} [f(\rho\mathbf{x} + \sqrt{1-\rho^2}\mathbf{y})].$$

The third statement of Theorem 3.7 follows from the following lemma.

**Lemma 3.12.** For all  $\delta \in [0, 10^{-55}]$ , if there is an  $x = 1 - 2^k$  such that  $E_{\mathbf{w} \sim \mathcal{N}(0, 1/d)^d} [f_x(\mathbf{w})] > 1 - 2\delta$  but  $E_{\mathbf{w} \sim \mathcal{N}(0, 1/d)^d} [f_{1-2x}(\mathbf{w})] \geq 2\delta - 1$  then the proportion of monochromatic triples  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$  in  $T'_x$  is at least  $\frac{\delta}{8}$ .

*Proof.* Assume that this is not true. Observe that whenever  $f_{1-2x}(\mathbf{w}) = 1$ , the proportion of triples  $(\mathbf{p}_1, \mathbf{p}_2, (1-2x)\mathbf{v}_0 + \sqrt{1-(2|x|-1)^2}\mathbf{w})$  which are monochromatic is at least

$$1 - 2P_{\mathbf{w}' \sim \mathcal{N}(0, 1/d)^d} (f_x(\rho_x\mathbf{w}' + \sqrt{1-\rho_x^2}\mathbf{w}') = -1) = (N_{\rho_x} f_x)(-\mathbf{w}).$$

Since  $E_{\mathbf{w} \sim \mathcal{N}(0, 1/d)^d} [f_{1-2x}(\mathbf{w})] \geq 2\delta - 1$ ,  $f_{1-2x}(\mathbf{w}) = 1$  for at least  $\delta$  proportion of  $\mathbf{w}$ . This implies that in order to avoid having  $\frac{\delta}{8}$  proportion of monochromatic triples, we must have that  $(N_{\rho_x} f_x)(-\mathbf{w}) < \frac{1}{4}$  for at least  $\frac{\delta}{2}$  proportion of  $\mathbf{w} \sim \mathcal{N}(0, 1/d)^d$ . We now show that this cannot occur. To do this, we write  $f_x = \mathbb{E}[f_x] + f'_x$  and show that  $\|(N_{\rho_x} f'_x)\|^2$  is small.

**Lemma 3.13.** For all  $t \in \mathbb{N}$ , if  $E_{\mathbf{w} \sim \mathcal{N}(0, 1/d)^d} [f_x(\mathbf{w})] > 1 - 2\delta$  then

$$\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0, 1/d)^d} [(N_{\rho_x} f'_x)(\mathbf{w})^2] \leq 10(3^t)\delta^{\frac{3}{2}} + 4\rho^{2t}\delta$$

*Proof.* Let  $S = \{\mathbf{w} \in \mathbb{R}^d : f_x(\mathbf{w}) = -1\}$  and let  $\delta' = \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0, 1/d)^d} [1_{\mathbf{w} \in S}] < \delta$  be the probability that  $f_x(\mathbf{w}) = -1$  for a random  $\mathbf{w} \sim \mathcal{N}(0, 1/d)^d$ . We can write  $f_x = (1 - 2\delta') + f'_{x,low} + f'_{x,high}$  where  $f'_{x,low}$  is the part of  $f'$  which has at most degree  $t$  and  $f'_{x,high}$  is the part of  $f'$  which has degree greater than  $t$ .

We observe that if we have a bound  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0, 1/d)^d} [f'_{x,low}(\mathbf{w})^2] \leq B$  then we can make the following deductions.

1. As described in Chapter 9 of Ryan O'Donnell's textbook "Analysis of Boolean Functions" [O'D14], the Bonami Lemma (which is a special case of the Hypercontractivity Theorem) says that if  $p$  is a polynomial of degree at most  $t$  then  $\mathbb{E}_{\mathbf{x} \in \{-1,1\}^n} [p(\mathbf{x})^4] \leq 9^t (\mathbb{E}_{\mathbf{x} \in \{-1,1\}^n} [p(\mathbf{x})^2])^2$ . As observed by Gross [Gro75], since Gaussian inputs can be approximated by a linear combination of a large number of Boolean variables, the Hypercontractivity Theorem (and in particular the Bonami Lemma) apply for Gaussian inputs as well as Boolean inputs. In other words,  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)} [p(\mathbf{w})^4] \leq 9^t (\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)} [p(\mathbf{w})^2])^2$ . Thus,  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)} [(f'_{x,low}(\mathbf{w}))^4] \leq 9^t B^2$  which implies that  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=-1} [|f'_{x,low}(\mathbf{w})|] \leq \sqrt[4]{\frac{9^t B^2}{\delta'}}$ .

For a discussion of the history of Bonami's Lemma and the Hypercontractivity Theorem, see Chapter 9.7 of O'Donnell's book [O'D14].

2. For all  $\mathbf{w} \in S^{d-1}$  such that  $f_x(\mathbf{w}) = -1$ ,  $f'_{x,high}(\mathbf{w}) = 2\delta' - 1 - f'_{x,low}(\mathbf{w})$ . This implies that

$$\begin{aligned} \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=-1} [f'_{x,low}(\mathbf{w}) f'_{x,high}(\mathbf{w})] &= \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=-1} [(2\delta' - 1) f'_{x,low}(\mathbf{w}) - f'_{x,low}(\mathbf{w})^2] \\ &\leq \sqrt[4]{\frac{9^t B^2}{\delta'}} - \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=-1} [f'_{x,low}(\mathbf{w})^2] \end{aligned}$$

3. For all  $\mathbf{w} \in S^{d-1}$  such that  $f_x(\mathbf{w}) = 1$ ,  $f'_{x,high}(\mathbf{w}) = 2\delta' - f'_{x,low}(\mathbf{w})$ . This implies that

$$\begin{aligned} \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=1} [f'_{x,low}(\mathbf{w}) f'_{x,high}(\mathbf{w})] &= \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=1} [2\delta' f'_{x,low}(\mathbf{w}) - f'_{x,low}(\mathbf{w})^2] \\ &\leq 2(\delta')^2 - \frac{1}{2} \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=1} [f'_{x,low}(\mathbf{w})^2] \end{aligned}$$

where the inequality uses the observation that

$$2\delta' f'_{x,low}(\mathbf{w}) - f'_{x,low}(\mathbf{w})^2 = 2\delta'^2 - \frac{1}{2} (2\delta' - f'_{x,low}(\mathbf{w}))^2 - \frac{1}{2} f'_{x,low}(\mathbf{w})^2 \leq 2\delta'^2 - \frac{1}{2} f'_{x,low}(\mathbf{w})^2$$

4. Putting these observations together,

$$\begin{aligned} 0 &= \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)} [f'_{x,low}(\mathbf{w}) f'_{x,high}(\mathbf{w})] = \delta' \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=-1} [f'_{x,low}(\mathbf{w}) f'_{x,high}(\mathbf{w})] \\ &\quad + (1 - \delta') \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d): f_x(\mathbf{w})=1} [f'_{x,low}(\mathbf{w}) f'_{x,high}(\mathbf{w})] \\ &\leq \delta' \sqrt[4]{\frac{9^t B^2}{\delta'}} + 2(\delta')^2 - \frac{1}{2} \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)} [f'_{x,low}(\mathbf{w})^2] \end{aligned}$$

which implies that

$$\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)} [f'_{x,low}(\mathbf{w})^2] \leq 2 \left( \sqrt[4]{9^t (\delta')^3 B^2} + 2(\delta')^2 \right)$$

Thus, if  $B > 2 \sqrt[4]{9^t (\delta')^3 B^2} + 4(\delta')^2$  then we can improve our bound on  $\mathbb{E}_{\mathbf{w} \in S^{d-1}} [f'_{x,low}(\mathbf{w})^2]$ . Letting  $B^*$  be the infimum of the bounds we can obtain, we must have that  $B^* = 2 \sqrt[4]{9^t (\delta')^3 (B^*)^2} + 4(\delta')^2$ . Solving for  $B^*$  gives

$$\sqrt{B^*} = \sqrt[4]{C(\delta')^3} + \sqrt{\sqrt{9^t (\delta')^3} + 4(\delta')^2} \leq \max \{3 \sqrt[4]{9^t (\delta')^3}, 3\delta'\} \leq 3 \sqrt[4]{9^t (\delta')^3}$$

where the first inequality follows by considering whether  $4(\delta')^2 \leq 3 \sqrt{9^t (\delta')^3}$  or  $4(\delta')^2 > 3 \sqrt{9^t (\delta')^3}$ . Thus,  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)} [f'_{x,low}(\mathbf{w})^2] \leq 10 \sqrt{9^t \delta^3}$  which implies that  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)} [(N_{\rho_x} f'_{x,low})(\mathbf{w})^2] \leq 10 \sqrt{9^t \delta^3}$ .

Since  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(f(\mathbf{w}) - (1 - 2\delta'))^2] = 4(1 - \delta')\delta'^2 + \delta'(-2 + 2\delta')^2 \leq 4\delta'$ , we have that  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(f'_{x,high}(\mathbf{w}))^2] \leq 4\delta'$  which implies that  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(N_{\rho_x} f'_{x,high})(\mathbf{w})^2] \leq 4\delta$ . The result now follows as

$$\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(N_{\rho_x} f'_x)(\mathbf{w})^2] = \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(N_{\rho_x} f'_{x,low})(\mathbf{w})^2] + \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(N_{\rho_x} f'_{x,high})(\mathbf{w})^2].$$

□

We now show that there must be at least  $\frac{\delta}{8}$  monochromatic triples. As we observed earlier, in order to avoid having  $\frac{\delta}{8}$  proportion of monochromatic triples, we must have that  $(N_{\rho_x} f_x)(-\mathbf{w}) < \frac{1}{4}$  for at least  $\frac{\delta}{2}$  proportion of  $\mathbf{w} \sim \mathcal{N}(0, 1/d)^d$ . Since  $N_{\rho_x} f_x = \mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[f_x(\mathbf{w})] + N_{\rho_x} f'_x \geq 1 - 2\delta + N_{\rho_x} f'_x$ , we have that  $|(N_{\rho_x} f'_x)(\mathbf{w})| \geq \frac{3}{4} - 4\delta$  for at least  $\frac{\delta}{2}$  proportion of  $\mathbf{w} \in S^{d-1}$ . Since  $\delta \leq \frac{1}{16}$ , this implies that  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(N_{\rho_x} f'_x)(\mathbf{w})^2] \geq \frac{\delta}{2} \cdot (\frac{1}{2})^2 = \frac{\delta}{8}$ .

However, Lemma 3.13 implies that if  $\delta > 0$  is sufficiently small then  $\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(N_{\rho_x} f'_x)] \leq \frac{\delta}{10}$ . To see this, take  $t = 50$  and observe that since each  $\rho$  that appears is at most .9,  $\rho^{2t} \leq \frac{1}{80}$ . Since  $\delta < \frac{1}{40000(9^t)}$ ,  $10(3^t)\delta^{\frac{3}{2}} \leq \frac{\delta}{20}$ . By Lemma 3.13,

$$\mathbb{E}_{\mathbf{w} \sim \mathcal{N}(0,1/d)^d}[(N_{\rho_x} f'_x)(\mathbf{w})^2] \leq 10(3^t)\delta^{\frac{3}{2}} + 4\rho^{2t}\delta \leq \frac{\delta}{20} + \frac{\delta}{20} = \frac{\delta}{10}. \quad \square$$

## 4 Improved Analysis for MAJ Polymorphism

In this section, we prove that the robust approximate algorithm for 2-SAT of Charikar–Makarychev–Makarychev [CMM09] has the same  $f(\epsilon) = O(\sqrt{\epsilon})$  loss guarantee for any PCSP template with the MAJ polymorphism family. This improves over the analysis of Brakensiek–Guruswami–Sandeep [BGS23b], which only proved  $f(\epsilon) = \tilde{O}(\sqrt[3]{\epsilon})$  loss. Furthermore, in the next section, we shall extend the result (with a slightly worse loss guarantee) to PCSP templates on arbitrary domains with more general types of polymorphisms.

**Theorem 4.1.** *Let  $\Gamma$  be a promise template with  $\text{MAJ} \subseteq \text{Pol}(\Gamma)$ . Then,  $\text{PCSP}(\Gamma)$  has a uniform robust algorithm with loss function  $f(\epsilon) = O_{\Gamma}(\sqrt{\epsilon})$ .*

### 4.1 CMM Algorithm

We state the algorithm of Charikar–Makarychev–Makarychev (henceforth “CMM”) as follows.

- Input:  $\epsilon > 0$ , weighted instance of  $\text{PCSP}(\Gamma)$  on variables  $x_1, \dots, x_n$  and clauses  $C_1, \dots, C_m$  as in Section 2.1.
- Solve the basic SDP to find a vector solution  $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^{n+1}$  with objective value at least  $1 - \epsilon$ .
- Sample a random Gaussian  $\mathbf{r} \in \mathcal{N}(0_{n+1}, I_{n+1})$ .
- For all  $i \in [n]$ , round  $x_i$  to  $+1$  if  $\langle \mathbf{v}_i, \mathbf{v}_0 + \mathbf{r}\sqrt{\epsilon} \rangle \geq 0$  and  $-1$  otherwise.

We call this the  $\epsilon$ -CMM algorithm to make the choice of  $\epsilon$  explicit. We make use of the following structural lemma proved by [BGS23b].

**Lemma 4.2** (Lemma 4.5 of [BGS23b], adapted). *Let  $P \subseteq Q \subseteq \{-1, 1\}^k$  be such that  $\text{MAJ} \subseteq \text{Pol}(P, Q)$ . For any  $\mathbf{b} \notin Q$ , there exists  $\mathbf{w} \in \mathbb{R}^k$  with the following properties.*

- $\|\mathbf{w}\|_1 = 1$ .
- For all  $i \in [k]$ ,  $b_i w_i \leq 0$ .
- For all  $\mathbf{a} \in P$ ,  $\langle \mathbf{a}, \mathbf{w} \rangle \geq 0$ .

We note that this lemma is proven in much more generality in Proposition 5.5. As a corollary, we have that an approximate version of Lemma 4.2 holds for approximate solutions to the SDP.

**Lemma 4.3** (Implicit in [BGS23b]). *Let  $P \subseteq Q \subseteq \{-1, 1\}^k$  be such that  $\text{MAJ} \subseteq \text{Pol}(P, Q)$ . Fix  $\mathbf{b} \in \{-1, 1\}^k \setminus Q$  and let  $\mathbf{w} \in \mathbb{R}^k$  be as guaranteed by Lemma 4.2. Then, for any  $1 - \gamma$  approximate vector assignment  $(\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_k)$  to  $P$  we have that*

$$\sum_{i=1}^k w_i \langle \mathbf{v}_i, \mathbf{v}_0 \rangle \geq -\gamma. \quad (7)$$

*Proof.* Since  $(\mathbf{v}_0, \dots, \mathbf{v}_k)$  is a  $1 - \gamma$  approximate vector assignment to  $P$ , there exists a probability distribution  $\mu$  over  $\{-1, 1\}^k$  such that  $\mu(P) \geq 1 - \gamma$  and for all  $i \in [k]$ ,

$$\langle \mathbf{v}_i, \mathbf{v}_0 \rangle = \sum_{\mathbf{a} \in \{-1, 1\}^k} a_i \mu(\mathbf{a}).$$

Thus, we have that

$$\begin{aligned} \sum_{i=1}^k w_i \langle \mathbf{v}_i, \mathbf{v}_0 \rangle &= \sum_{i=1}^k w_i \sum_{\mathbf{a} \in \{-1, 1\}^k} a_i \mu(\mathbf{a}) \\ &= \sum_{\mathbf{a} \in \{-1, 1\}^k} \langle \mathbf{a}, \mathbf{w} \rangle \mu(\mathbf{a}). \end{aligned}$$

By Lemma 4.2,  $\langle \mathbf{a}, \mathbf{w} \rangle \geq 0$  for all  $\mathbf{a} \in P$ . Furthermore,  $\langle \mathbf{a}, \mathbf{w} \rangle \geq -1$  for all  $\mathbf{a} \in \{-1, 1\}^k$  since  $\|\mathbf{w}\|_1 = 1$ . Thus, (7) follows since  $\mu(\{-1, 1\}^k \setminus P) \leq \gamma$ .  $\square$

We now state our main technical result.

**Theorem 4.4.** *Let  $P \subseteq Q \subseteq \{-1, 1\}^k$  with  $\text{MAJ} \subseteq \text{Pol}(P, Q)$ . Let  $(\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_k)$  be a  $1 - \gamma$  approximate vector assignment to  $P$ . For any  $\mathbf{b} \in \{-1, 1\}^k \setminus Q$ , the probability that the  $\epsilon$ -CMM algorithm when run on  $(\mathbf{v}_0, \dots, \mathbf{v}_k)$  outputs  $\mathbf{b}$  is at most*

$$600k \frac{\max(\epsilon, \gamma)}{\sqrt{\epsilon}} \quad (8)$$

From this, Theorem 4.1 easily follows.

*Proof of Theorem 4.1.* Our Basic SDP solution  $(\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n)$  has the property that each clause  $C_j$  is  $1 - \gamma_j$  approximately satisfied with  $\mathbb{E}_j[\gamma_j] \leq \epsilon$ . Let  $k_j$  denote the arity of the clause  $C_j$ , and

define  $k_{\max} = \max_j(k_j)$ . By Theorem 4.4, the probability that  $C_j$  is assigned some specific weakly unsatisfying assignment is at most

$$600k_j \frac{\max(\epsilon, \gamma_j)}{\sqrt{\epsilon}}.$$

Therefore, by the union bound, the probability that  $C_j$  is weakly satisfied is at least

$$1 - 2^{k_j} \left( 600k_j \frac{\max(\epsilon, \gamma_j)}{\sqrt{\epsilon}} \right) \geq 1 - 2^{k_j} \left( 600k_j \frac{\epsilon + \gamma_j}{\sqrt{\epsilon}} \right) \geq 1 - 2^{k_{\max}} \left( 600k_{\max} \frac{\epsilon + \gamma_j}{\sqrt{\epsilon}} \right).$$

Thus, the expected fraction of clauses that are weakly satisfied is at least

$$1 - 1200 \cdot 2^{k_{\max}} k_{\max} \sqrt{\epsilon},$$

as desired. □

## 4.2 Proof of Theorem 4.4

We seek to bound the probability that

$$b_i \langle \mathbf{v}_i, \mathbf{r} + \mathbf{v}_0 / \sqrt{\epsilon} \rangle \geq 0, \quad \forall i \in [k]. \quad (9)$$

Let  $\mathbf{w}$  correspond to  $\mathbf{b}$  as in Lemma 4.2. Let  $\mathbf{u} := w_1 \mathbf{v}_1 + \dots + w_k \mathbf{v}_k$ . By Lemma 4.3, we have that  $\langle \mathbf{u}, \mathbf{v}_0 \rangle \geq -\gamma$ . Furthermore, since  $b_i w_i \leq 0$  for all  $i \in [k]$ , we have that (9) holding implies that

$$\langle w_i \mathbf{v}_i, \mathbf{r} + \mathbf{v}_0 / \sqrt{\epsilon} \rangle \leq 0, \quad \forall i \in [k]. \quad (10)$$

Summing over all  $i \in [k]$ , this implies that

$$\langle \mathbf{u}, \mathbf{r} + \mathbf{v}_0 / \sqrt{\epsilon} \rangle \leq 0. \quad (11)$$

To work toward this, we first show that  $\langle \mathbf{u}, \mathbf{u} \rangle$  cannot stray too much from  $\langle \mathbf{u}, \mathbf{v}_0 \rangle$ .

**Lemma 4.5.**  $\langle \mathbf{u}, \mathbf{u} \rangle \leq \langle \mathbf{u}, \mathbf{v}_0 \rangle + 2\gamma$ .

*Proof.* Let  $\mu$  be the probability distribution over  $\{-1, 1\}^k$  with  $\mu(P) \geq 1 - \gamma$  that corresponds to  $(\mathbf{v}_0, \dots, \mathbf{v}_k)$ . Recall from the proof of Lemma 4.3 that

$$\langle \mathbf{u}, \mathbf{v}_0 \rangle = \sum_{\mathbf{a} \in \{-1, 1\}^k} \langle \mathbf{a}, \mathbf{w} \rangle \mu(\mathbf{a}).$$

Likewise, we can compute that

$$\begin{aligned} \langle \mathbf{u}, \mathbf{u} \rangle &= \sum_{i=1}^k \sum_{j=1}^k w_i w_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle \\ &= \sum_{i=1}^k \sum_{j=1}^k w_i w_j \sum_{\mathbf{a} \in \{-1, 1\}^k} a_i a_j \mu(\mathbf{a}) \\ &= \sum_{\mathbf{a} \in \{-1, 1\}^k} \langle \mathbf{a}, \mathbf{w} \rangle^2 \mu(\mathbf{a}). \end{aligned}$$

Thus,

$$\langle \mathbf{u}, \mathbf{u} \rangle - \langle \mathbf{u}, \mathbf{v}_0 \rangle = \sum_{\mathbf{a} \in \{-1, 1\}^k} \langle \mathbf{a}, \mathbf{w} \rangle (\langle \mathbf{a}, \mathbf{w} \rangle - 1) \mu(\mathbf{a}).$$

Recall that  $\|\mathbf{w}\|_1 = 1$ . For any  $\mathbf{a} \in P$ , we have that  $\langle \mathbf{a}, \mathbf{w} \rangle \in [0, 1]$ , so  $\langle \mathbf{a}, \mathbf{w} \rangle (\langle \mathbf{a}, \mathbf{w} \rangle - 1) \leq 0$ . Furthermore, for any  $\mathbf{a} \in \{-1, 1\}^k$ ,  $\langle \mathbf{a}, \mathbf{w} \rangle \in [-1, 1]$ , which implies that  $\langle \mathbf{a}, \mathbf{w} \rangle (\langle \mathbf{a}, \mathbf{w} \rangle - 1) \leq 2$ . Thus,  $\langle \mathbf{u}, \mathbf{u} \rangle - \langle \mathbf{u}, \mathbf{v}_0 \rangle \leq 2\gamma$ .  $\square$

For each  $i \in [k]$ , let  $\mathbf{v}_i^\parallel$  be the component of  $\mathbf{v}_i$  parallel to  $\mathbf{u}$ , and let  $\mathbf{v}_i^\perp$  be the component of  $\mathbf{v}_i$  perpendicular to  $\mathbf{u}$ . That is,

$$\begin{aligned} \mathbf{v}_i^\parallel &:= \frac{\langle \mathbf{v}_i, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}, \\ \mathbf{v}_i^\perp &:= \mathbf{v}_i - \frac{\langle \mathbf{v}_i, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}. \end{aligned}$$

Let  $\eta := \max(\gamma, \epsilon)$ . We now split into cases.

**Case 1,**  $\|\mathbf{u}\|^2 \geq 8\eta \ln(1/\epsilon)$ . In that case, we have that  $\|\mathbf{u}\|^2 \geq 4\gamma$ . Therefore, by Lemma 4.5,

$$\langle \mathbf{u}, \frac{1}{\sqrt{\epsilon}} \mathbf{v}_0 \rangle \geq \frac{1}{\sqrt{\epsilon}} (\|\mathbf{u}\|^2 - 2\gamma) \geq \frac{\|\mathbf{u}\|^2}{2\sqrt{\epsilon}}.$$

Therefore, (11) holds only if

$$\langle \mathbf{r}, -\frac{\mathbf{u}}{\|\mathbf{u}\|} \rangle > \frac{\|\mathbf{u}\|}{2\sqrt{\epsilon}}.$$

Note that the LHS is normally distributed with mean 0 and variance 1. Therefore, by Proposition 2.2, we deduce that (9) holds with probability at most

$$\frac{1}{2} \exp\left(-\frac{1}{2} \left(\frac{\|\mathbf{u}\|}{2\sqrt{\epsilon}}\right)^2\right) \leq \frac{1}{2} \exp\left(-\frac{\eta}{\epsilon} \log(1/\epsilon)\right) \leq \frac{\epsilon}{2} \leq \frac{\eta}{\sqrt{\epsilon}}.$$

**Case 2, there exists  $i \in [k]$  such that  $\|w_i \mathbf{v}_i^\perp\| \geq 1/(10k)$ .** First, note that by Proposition 2.2, for all  $\delta \geq 0$  we have that

$$\Pr[\delta \leq |\langle \mathbf{r}, \mathbf{u} \rangle| \leq 2\delta] = \Pr\left[\frac{\delta}{\|\mathbf{u}\|} \leq \left\langle \mathbf{r}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \leq \frac{2\delta}{\|\mathbf{u}\|}\right] \leq 2 \frac{\delta}{\|\mathbf{u}\| \sqrt{2\pi}} e^{-\frac{\delta^2}{2\|\mathbf{u}\|^2}} \leq \frac{\delta}{\|\mathbf{u}\|} e^{-\frac{\delta^2}{2\|\mathbf{u}\|^2}}. \quad (12)$$

Assume that (9) holds. Then, by (10), (11), and Lemma 4.5, we have that

$$0 \geq \left\langle w_i \mathbf{v}_i, \mathbf{r} + \frac{\mathbf{v}_0}{\sqrt{\epsilon}} \right\rangle \geq \langle \mathbf{u}, \mathbf{r} \rangle + \frac{1}{\sqrt{\epsilon}} \langle \mathbf{u}, \mathbf{v}_0 \rangle \geq \langle \mathbf{u}, \mathbf{r} \rangle + \frac{1}{\sqrt{\epsilon}} (\|\mathbf{u}\|^2 - 2\gamma) \geq \langle \mathbf{u}, \mathbf{r} \rangle - \frac{2}{\sqrt{\epsilon}} \gamma. \quad (13)$$

Consider now the probability distribution of the random variable  $\langle w_i \mathbf{v}_i, \mathbf{r} + \frac{\mathbf{v}_0}{\sqrt{\epsilon}} \rangle$  conditional to the value of  $\langle \mathbf{u}, \mathbf{r} \rangle$ . Observe that this is a Gaussian distribution with standard deviation  $\|w_i \mathbf{v}_i^\perp\| \geq \frac{1}{10k}$  (and arbitrary mean). By Proposition 2.2, the probability that (13) happens is at most

$$10k \left( |\langle \mathbf{u}, \mathbf{r} \rangle| + \frac{2}{\sqrt{\epsilon}} \gamma \right) \leq 10k \left( 2\delta + \frac{2}{\sqrt{\epsilon}} \gamma \right).$$

Thus, the probability that both  $\delta \leq |\langle \mathbf{r}, \mathbf{u} \rangle| \leq 2\delta$  and (9) hold is at most

$$\frac{\delta}{\|\mathbf{u}\|} e^{-\frac{\delta^2}{2\|\mathbf{u}\|^2}} \cdot 10k \left( 2\delta + \frac{2}{\sqrt{\epsilon}} \gamma \right) \leq \frac{40k\delta e^{-\frac{\delta^2}{2\|\mathbf{u}\|^2}}}{\|\mathbf{u}\|} \max \left\{ \delta, \frac{\gamma}{\sqrt{\epsilon}} \right\}.$$

Call the RHS of the above expression  $f(\delta)$ . We now split our analysis into subcases based on the size of  $\|\mathbf{u}\|^2$ .

**Case 2a**,  $\|\mathbf{u}\|^2 \leq 4\eta$ . We seek to bound  $\sum_{i=-\infty}^{\infty} f(2^i \|\mathbf{u}\|)$ . If  $i \leq 1$ , note that

$$f(2^i \|\mathbf{u}\|) \leq \frac{40k \cdot 2^i \|\mathbf{u}\|}{\|\mathbf{u}\|} \max \left\{ 2^i \|\mathbf{u}\|, \frac{\gamma}{\sqrt{\epsilon}} \right\} \leq 40k \cdot 2^i \cdot \frac{\eta}{\sqrt{\epsilon}}.$$

Summing over all  $i \leq 1$ , we obtain a contribution of at most  $160k\eta/\sqrt{\epsilon}$ . Otherwise, if  $i \geq 2$ , we have that

$$f(2^i \|\mathbf{u}\|) \leq 40k \cdot 2^i \cdot e^{-2^{2i-1}} \cdot \max \left\{ 2^{i+1} \sqrt{\eta}, \frac{\eta}{\sqrt{\epsilon}} \right\} \leq \frac{160k\eta}{\sqrt{\epsilon}} \cdot 2^{2i-1} e^{-2^{2i-1}}.$$

Using that the function  $x \mapsto x \cdot e^{-x}$  is nonincreasing in  $[1, \infty)$ , we find

$$\sum_{i=2}^{\infty} 2^{2i-1} e^{-2^{2i-1}} \leq \sum_{i=2}^{\infty} i e^{-i} \leq \int_0^{\infty} x \cdot e^{-x} dx = 1$$

and, thus,

$$\sum_{i=2}^{\infty} f(2^i \|\mathbf{u}\|) \leq \frac{160k\eta}{\sqrt{\epsilon}}.$$

As a consequence, the total probability bound is at most  $\frac{320k\eta}{\sqrt{\epsilon}}$ .

**Case 2b**,  $\|\mathbf{u}\|^2 > 4\eta$ . By logic similar to Case 1, we have that (9) holds only if  $\langle \mathbf{r}, -\frac{\mathbf{u}}{\|\mathbf{u}\|} \rangle > \frac{\|\mathbf{u}\|}{2\sqrt{\epsilon}}$ . Thus, we only need to consider  $\delta \geq \frac{\|\mathbf{u}\|^2}{2\sqrt{\epsilon}}$ . Consider any such  $\delta = \lambda \|\mathbf{u}\| \geq \|\mathbf{u}\|^2/(2\sqrt{\epsilon})$  so  $2\lambda\sqrt{\epsilon} \geq \|\mathbf{u}\|$ . We have that

$$f(\lambda \|\mathbf{u}\|) = 40k\lambda e^{-\frac{\lambda^2}{2}} \max \left\{ \lambda \|\mathbf{u}\|, \frac{\gamma}{\sqrt{\epsilon}} \right\} \leq 160k\lambda^3 e^{-\frac{\lambda^2}{2}} \sqrt{\epsilon}.$$

Since  $2\lambda\sqrt{\epsilon} \geq \|\mathbf{u}\| > \sqrt{4\eta}$ , we have that  $\lambda > 1$ . Thus, the relevant probability of (9) occurring is at most

$$\sum_{i=0}^{\infty} 160k \cdot 2^{3i} \cdot e^{-2^{2i-1}} \sqrt{\epsilon} \leq 600k\sqrt{\epsilon} \leq 600k\eta/\sqrt{\epsilon},$$

as desired.

**Case 3**,  $\|\mathbf{u}\|^2 < 8\eta \ln(\frac{1}{\epsilon})$  and there is no  $i$  such that  $\|w_i \mathbf{v}_i^\perp\| \geq \frac{1}{10k}$ . In this case, since  $\mathbf{u} = \sum_{i=1}^k w_i \mathbf{v}_i$  and  $\sum_{i=1}^k \|w_i \mathbf{v}_i\| = \sum_{i=1}^k |w_i| \|\mathbf{v}_i\| = \sum_{i=1}^k |w_i| = 1$ , we have that

$$1 = \sum_{i=1}^k \|w_i \mathbf{v}_i\| \leq \sum_{i=1}^k \|w_i \mathbf{v}_i^\parallel\| + \sum_{i=1}^k \|w_i \mathbf{v}_i^\perp\| \leq \sum_{i=1}^k \|w_i \mathbf{v}_i^\parallel\| + \frac{1}{10},$$

and

$$\left\| \sum_{i=1}^k w_i \mathbf{v}_i \right\|^2 + \left\| \sum_{i=1}^k w_i \mathbf{v}_i^\perp \right\|^2 = \|\mathbf{u}\|^2 \leq 8\eta \ln \left( \frac{1}{\epsilon} \right).$$

Putting these observations together, there must be an  $i$  such that  $w_i \mathbf{v}_i = c \frac{\mathbf{u}}{\|\mathbf{u}\|} + w_i \mathbf{v}_i^\perp$  where  $|c| \geq \frac{1}{2k}$ . We now define  $t, x, y$  as follows

$$\begin{aligned} t &:= -\frac{\langle w_i \mathbf{v}_i, \mathbf{v}_0 \rangle}{c\sqrt{\epsilon}}, \\ x &:= \left\langle \frac{\mathbf{u}}{\|\mathbf{u}\|}, \mathbf{r} \right\rangle, \\ y &:= -\sum_i \langle w_i \mathbf{v}_i, \mathbf{r} + \mathbf{v}_0/\sqrt{\epsilon} \rangle. \end{aligned}$$

Note that  $t$  is a constant, but  $x$  and  $y$  are random variables depending on  $\mathbf{r}$ . In particular,  $x$  is a Gaussian random variable with mean zero and variable one. Furthermore, note that

$$\langle w_i \mathbf{v}_i, \mathbf{r} + \mathbf{v}_0/\sqrt{\epsilon} \rangle = c(x - t) + \langle w_i \mathbf{v}_i^\perp, \mathbf{r} \rangle.$$

We next observe the following. First, in order for (9) to hold, we must have that  $-y \leq \langle w_i \mathbf{v}_i, \mathbf{r} + \mathbf{v}_0/\sqrt{\epsilon} \rangle \leq 0$  which implies that

$$|\langle w_i \mathbf{v}_i^\perp, \mathbf{r} \rangle + c(x - t)| \leq y. \quad (14)$$

Second, since  $-y = \sum_i \langle w_i \mathbf{v}_i, \mathbf{r} + \mathbf{v}_0/\sqrt{\epsilon} \rangle = \langle \mathbf{u}, \mathbf{r} \rangle + \frac{1}{\sqrt{\epsilon}} \langle \mathbf{u}, \mathbf{v}_0 \rangle \geq \langle \mathbf{u}, \mathbf{r} \rangle + \frac{1}{\sqrt{\epsilon}} (\|\mathbf{u}\|^2 - 2\gamma) = \|\mathbf{u}\|x + \frac{1}{\sqrt{\epsilon}} (\|\mathbf{u}\|^2 - 2\gamma)$ , we have

$$|x| \geq -x \geq \frac{\|\mathbf{u}\|}{\sqrt{\epsilon}} - \frac{2\gamma}{\|\mathbf{u}\|\sqrt{\epsilon}} + \frac{y}{\|\mathbf{u}\|}. \quad (15)$$

To finish the proof, we split into cases based on the values of  $(y, c(x - t))$ .

**Case 3a**,  $y \leq 4\frac{\eta}{\sqrt{\epsilon}}$ . Since  $|c| \geq \frac{1}{2k}$ , but  $\|w_i \mathbf{v}_i^\perp\| \leq \frac{1}{10k}$ , we have that  $\langle w_i \mathbf{v}_i, \mathbf{r} + \mathbf{v}_0/\sqrt{\epsilon} \rangle = c(x - t) + \langle w_i \mathbf{v}_i^\perp, \mathbf{r} \rangle$  has standard deviation at least  $|c|/2 \geq \frac{1}{4k}$ . Thus, by Proposition 2.2

$$\Pr \left[ |\langle w_i \mathbf{v}_i^\perp, \mathbf{r} \rangle + c(x - t)| \leq 4\frac{\eta}{\sqrt{\epsilon}} \right] \leq \frac{4\eta}{|c|\sqrt{\epsilon}/2} \leq 16k \frac{\eta}{\sqrt{\epsilon}}.$$

That is, if we condition on  $y \leq \frac{4\eta}{\sqrt{\epsilon}}$ , the probability that (14) holds (and thus the probability that (9) holds) is at most  $16k \frac{\eta}{\sqrt{\epsilon}}$ .

**Case 3b**, there is  $\delta \geq 4\eta/\sqrt{\epsilon}$  such that  $y \geq \delta$  and  $|c(x - t)| \leq 4\delta$ . Note that  $\delta/2 \geq 2\gamma/\sqrt{\epsilon}$ , so by (15), we have that

$$|x| \geq \frac{\|\mathbf{u}\|}{\sqrt{\epsilon}} + \frac{\delta}{2\|\mathbf{u}\|}. \quad (16)$$

Thus, by the Proposition 2.2, we have that

$$\begin{aligned} \Pr[|c(x - t)| \leq 4\delta \wedge y \geq \delta] &\leq \Pr[|c(x - t)| \leq 4\delta \wedge (16)] \\ &\leq \frac{8\delta}{|c|} e^{-\frac{1}{2} \left( \frac{\|\mathbf{u}\|}{\sqrt{\epsilon}} + \frac{\delta}{2\|\mathbf{u}\|} \right)^2} \leq \frac{8\delta}{|c|} e^{-\frac{\delta}{\sqrt{\epsilon}}} \leq 16k\delta e^{-\frac{\delta}{\sqrt{\epsilon}}}. \end{aligned}$$

**Case 3c, there is  $\delta' \geq 4\delta \geq 16\eta/\sqrt{\epsilon}$  such that  $y \in [\delta, 2\delta]$  and  $|c(x-t)| \in [\delta', 2\delta']$ .** Note that (16) still holds, so we have that

$$\begin{aligned} \Pr[|c(x-t)| \in [\delta', 2\delta'] \wedge y \in [\delta, 2\delta]] &\leq \Pr[|c(x-t)| \in [\delta', 2\delta'] \wedge (16)] \\ &\leq \frac{2\delta'}{|c|} e^{-\frac{1}{2}\left(\frac{\|\mathbf{u}\|}{\sqrt{\epsilon}} + \frac{\delta}{2\|\mathbf{u}\|}\right)^2} \leq \frac{2\delta'}{|c|} e^{-\frac{\delta}{\sqrt{\epsilon}}}. \end{aligned}$$

Let  $d = \|w_i \mathbf{v}_i^\perp\|$ . We then have by Proposition 2.2 that

$$\begin{aligned} \Pr[(9) \mid |c(x-t)| \in [\delta', 2\delta'] \wedge y \in [\delta, 2\delta]] &\leq \Pr[(14) \mid |c(x-t)| \in [\delta', 2\delta'] \wedge y \in [\delta, 2\delta]] \\ &\leq \frac{4\delta}{d} e^{-\frac{\delta'^2}{8d^2}}, \end{aligned}$$

where the last inequality follows from the fact that  $\langle w_i \mathbf{v}_i^\perp, \mathbf{r} \rangle$  must land in an interval of length at most  $2\delta$  with minimum values of  $\delta' - 2\delta \geq \frac{\delta'}{2}$ . Therefore, combining the previous two calculations, we have that

$$\Pr[(9) \wedge |c(x-t)| \in [\delta', 2\delta'] \wedge y \in [\delta, 2\delta]] \leq \frac{8\delta\delta'}{|c|d} e^{-\frac{\delta}{\sqrt{\epsilon}} - \frac{\delta'^2}{8d^2}}.$$

**Finishing Case 3.** We now put these subcases together. First, condition on  $\delta \geq 4\eta/\sqrt{\epsilon}$ . We have by cases 3b and 3c that

$$\begin{aligned} \Pr[(9) \wedge y \in [\delta, 2\delta]] &\leq \Pr[|c(x-t)| \in [\delta', 2\delta'] \wedge y \in [\delta, 2\delta]] \\ &\quad + \sum_{i=2}^{\infty} \Pr[(9) \mid |c(x-t)| \in [2^i\delta, 2^{i+1}\delta] \wedge y \in [\delta, 2\delta]] \\ &\leq 6k\delta e^{-\frac{\delta}{\sqrt{\epsilon}}} + \sum_{i=2}^{\infty} \frac{2^{i+3}\delta^2}{|c|d} e^{-\frac{\delta}{\sqrt{\epsilon}} - \frac{2^{2i}\delta^2}{8d^2}} \\ &= 6k\delta e^{-\frac{\delta}{\sqrt{\epsilon}}} + 16k \cdot \delta e^{-\delta/\sqrt{\epsilon}} \sum_{i=2}^{\infty} \frac{2^i\delta}{d} e^{-\frac{2^{2i}\delta^2}{8d^2}}. \end{aligned}$$

Now,

$$\sum_{i=2}^{\infty} \frac{2^i\delta}{d} e^{-\frac{2^{2i}\delta^2}{8d^2}} \leq \sum_{i=2}^{\infty} 2 \int_{2^{i-1}}^{2^i} \frac{\delta}{d} e^{-\frac{s^2\delta^2}{8d^2}} ds \leq \int_{-\infty}^{\infty} \frac{\delta}{d} e^{-\frac{s^2\delta^2}{8d^2}} ds = \int_{-\infty}^{\infty} e^{-\frac{s^2}{8}} ds = 2\sqrt{2\pi} \leq \frac{11}{2},$$

so

$$\Pr[(9) \wedge y \in [\delta, 2\delta]] \leq 100k\delta e^{-\frac{\delta}{\sqrt{\epsilon}}}.$$

We next incorporate the bound from case 3a to get that

$$\begin{aligned} \Pr[(9)] &\leq \Pr[(9) \wedge y \leq 4\eta/\sqrt{\epsilon}] + \sum_{i=2}^{\infty} \Pr[(9) \wedge y \in [2^i\eta/\sqrt{\epsilon}, 2^{i+1}\eta/\sqrt{\epsilon}]] \\ &\leq 16k \frac{\eta}{\sqrt{\epsilon}} + 100k \cdot \frac{\eta}{\sqrt{\epsilon}} \sum_{i=2}^{\infty} 2^i e^{-\frac{2^i\eta}{\epsilon}} \\ &\leq 16k \frac{\eta}{\sqrt{\epsilon}} + 100k \cdot \frac{\eta}{\sqrt{\epsilon}} \sum_{i=2}^{\infty} 2^i e^{-2^i} \\ &\leq 50k \cdot \frac{\eta}{\sqrt{\epsilon}}, \end{aligned}$$

as desired. This completes the proof of Theorem 4.4 and thus also of Theorem 4.1.

## 5 Rounding schemes for separable PCSPs

The goal of this section is to extend the results of Section 4 to the case of promise CSPs on arbitrary domains. We begin with two definitions.

Given a tuple  $\mathbf{a} \in A^k$ , we denote by  $\Pi_{\mathbf{a}} \in \mathbb{R}^{k \times A}$  the one-hot encoding of  $\mathbf{a}$ ; i.e., the  $(i, a)$ -th entry of the matrix  $\Pi_{\mathbf{a}}$  is 1 if  $a_i = a$ , 0 otherwise. Also, we let  $\langle -, - \rangle_{\mathbb{F}}$  denote the Frobenius inner product of matrices, given by  $\langle M, N \rangle_{\mathbb{F}} = \text{Tr}(M^{\top} N)$ .

**Definition 5.1.** Fix a PCSP template  $(\mathbf{A}, \mathbf{B})$  and a function  $\rho : \mathbb{R}^A \rightarrow B$ . We say that  $(\mathbf{A}, \mathbf{B})$  is  $\rho$ -separable if, for each promise relation  $(P, Q)$  of some arity  $k$  and each  $\mathbf{b} \in B^k \setminus Q$ , there exists a matrix  $W \in \mathbb{R}^{k \times A}$  such that

1.  $\|W\|_1 = 1$ ;
2.  $W\mathbf{1} = \mathbf{0}$ ;
3.  $\langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}} \geq 0$  for each  $\mathbf{a} \in P$ ; and
4.  $\langle W, M \rangle_{\mathbb{F}} \leq 0$  for each  $M \in \rho^{-1}(\mathbf{b})$ .

Note that, in part 4. of the above definition, the expression  $\rho^{-1}(\mathbf{b})$  indicates the set of  $k \times A$  real matrices whose  $i$ -th row is in  $\rho^{-1}(b_i)$  for each  $i \in [k]$ .

Intuitively, in Definition 5.1,  $\rho$  should be thought of as an arbitrary rounding scheme, and the idea is to consider polymorphism families that are compatible with the scheme in the sense that they can be used to avoid the occurrence of large integrality gaps. We shall additionally need that the rounding function  $\rho$  be well-behaved in the following weak sense. Henceforth we will assume, without loss of generality, that  $A \subseteq B$ .

**Definition 5.2.** Fix some parameter  $\alpha > 0$ . We say that a function  $\rho : \mathbb{R}^A \rightarrow B$  is  $\alpha$ -conservative if, for each  $\mathbf{q} \in \mathbb{R}^A$ , the following holds:

- if  $\rho(\mathbf{q}) \in A$ , then  $q_{\rho(\mathbf{q})} \geq \alpha \sum_{b \in A} q_b$ ;
- if  $\rho(\mathbf{q}) \notin A$ , then  $q_a \geq \alpha \sum_{b \in A} q_b$  for each  $a \in A$ .

**Theorem 5.3.** Let  $(\mathbf{A}, \mathbf{B})$  be a  $\rho$ -separable PCSP template for an  $\alpha$ -conservative function  $\rho : \mathbb{R}^A \rightarrow B$ . Then  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  is robustly solvable via SDP with loss  $O_{\alpha, A}(\sqrt{\epsilon} \log(1/\epsilon))$ .

*Proof.* Without loss of generality, we shall consider the case in which  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  contains a single promise relation  $(P, Q)$  of some arity  $k$ . Let  $(\mathbf{v}_0, \{\mathbf{v}_{1,a} : a \in A\}, \dots, \{\mathbf{v}_{k,a} : a \in A\})$  be a  $1 - \gamma$  approximate vector assignment to  $P$  for some  $\gamma > 0$ , as per Section 2.2. Fix  $\epsilon > 0$ , and run the  $\epsilon$ -CMM algorithm on such vector assignment, using  $\rho$  as the rounding function (see Section 4.1). More precisely, we apply the following procedure:

1. Sample a random Gaussian  $\mathbf{r} \in \mathcal{N}(0_n, I_n)$  (where  $n$  is the dimension of the space where the SDP vectors live).
2. Define  $\mathbf{z} = \mathbf{r} + \frac{\mathbf{v}_0}{\sqrt{\epsilon}}$ , and let  $\mathbf{c}^{(i)} \in \mathbb{R}^A$  be the vector whose  $a$ -th entry is  $c_a^{(i)} = \langle \mathbf{v}_{i,a}, \mathbf{z} \rangle$  for each  $i \in [k]$ .

3. Let  $b_i = \rho(\mathbf{c}^{(i)})$  for each  $i \in [k]$ .

Consider now the vector  $\mathbf{b} = (b_i)_{i \in [k]} \in B^k$ , and suppose that  $\mathbf{b} \notin Q$ . Take a corresponding matrix  $W \in \mathbb{R}^{k,A}$  as guaranteed from the fact that  $(\mathbf{A}, \mathbf{B})$  is  $\rho$ -separable, and define the vector

$$\mathbf{u} = \sum_{i \in [k]} \sum_{a \in A} w_{i,a} \mathbf{v}_{i,a}.$$

Consider also the matrix  $M \in \mathbb{R}^{k \times A}$  whose  $(i, a)$ -th entry is  $c_a^{(i)}$ . Observe that  $M \in \rho^{-1}(\mathbf{b})$ , so part 4. of Definition 5.1 guarantees that

$$\langle \mathbf{u}, \mathbf{z} \rangle = \langle W, M \rangle_{\mathbb{F}} \leq 0. \quad (17)$$

Observe now that, for each  $\mathbf{a} \in A^k$ , it holds that

$$\langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}} = \sum_{i \in [k]} w_{i,a_i}.$$

Hence, we find

$$\begin{aligned} \langle \mathbf{u}, \mathbf{u} \rangle &= \sum_{i,i' \in [k]} \sum_{a,a' \in A} w_{i,a} w_{i',a'} \langle \mathbf{v}_{i,a}, \mathbf{v}_{i',a'} \rangle = \sum_{i,i' \in [k]} \sum_{a,a' \in A} w_{i,a} w_{i',a'} \sum_{\substack{\mathbf{a} \in A^k \\ a_i = a \\ a_{i'} = a'}} \mu(\mathbf{a}) \\ &= \sum_{\mathbf{a} \in A^k} \mu(\mathbf{a}) \sum_{i,i' \in [k]} w_{i,a_i} w_{i',a_{i'}} = \sum_{\mathbf{a} \in A^k} \mu(\mathbf{a}) \langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}}^2. \end{aligned}$$

Similarly,

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v}_0 \rangle &= \sum_{i \in [k]} \sum_{a \in A} w_{i,a} \langle \mathbf{v}_{i,a}, \mathbf{v}_0 \rangle = \sum_{i \in [k]} \sum_{a \in A} w_{i,a} \sum_{\substack{\mathbf{a} \in A^k \\ a_i = a}} \mu(\mathbf{a}) = \sum_{\mathbf{a} \in A^k} \mu(\mathbf{a}) \sum_{i \in [k]} w_{i,a_i} \\ &= \sum_{\mathbf{a} \in A^k} \mu(\mathbf{a}) \langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}}. \end{aligned} \quad (18)$$

Now, part 1. of Definition 5.1 implies that, for each  $\mathbf{a} \in A^k$ ,

$$|\langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}}| = \left| \sum_{i \in [k]} w_{i,a_i} \right| \leq \sum_{i \in [k], a \in A} |w_{i,a}| = \|W\|_1 = 1,$$

while part 3. implies that  $0 \leq \langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}}$  for each  $\mathbf{a} \in P$ . Hence, the quantity  $\langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}}^2 - \langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}}$  is at most 2 for every  $\mathbf{a} \in A^k$ , and it is nonpositive if  $\mathbf{a} \in P$ . It follows that

$$\langle \mathbf{u}, \mathbf{u} \rangle - \langle \mathbf{u}, \mathbf{v}_0 \rangle = \sum_{\mathbf{a} \in A^k} \mu(\mathbf{a}) (\langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}}^2 - \langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}}) \leq 2\gamma \quad (19)$$

(where we have used that the vectors  $\mathbf{v}_0, \{\mathbf{v}_{i,a}\}$  form a  $1 - \gamma$  approximate vector assignment to  $P$ ).

Take now  $\delta = 2\epsilon \log(1/\epsilon)$  and  $\eta = \sqrt{2(\delta + 2\gamma) \log(1/\epsilon) + \frac{\delta + \gamma}{\sqrt{\epsilon}}}$ . Suppose first that  $\langle \mathbf{u}, \mathbf{v}_0 \rangle > \delta$ . Note that the random variable  $\langle \mathbf{u}, \mathbf{z} \rangle = \langle \mathbf{u}, \mathbf{r} + \frac{\mathbf{v}_0}{\sqrt{\epsilon}} \rangle$  is distributed as a Gaussian with mean  $\mu = \frac{1}{\sqrt{\epsilon}} \langle \mathbf{u}, \mathbf{v}_0 \rangle$

and standard deviation  $\sigma = \|\mathbf{u}\|$ . Applying part (b) of Proposition 2.1 and assuming without loss of generality that  $\gamma \leq \frac{\delta}{2}$ , we deduce that the probability that (17) happens is upper bounded by

$$\exp\left(-\frac{\langle \mathbf{u}, \mathbf{v}_0 \rangle^2}{2\epsilon \|\mathbf{u}\|^2}\right) \leq \exp\left(-\frac{\langle \mathbf{u}, \mathbf{v}_0 \rangle^2}{2\epsilon(\langle \mathbf{u}, \mathbf{v}_0 \rangle + 2\gamma)}\right) \leq \exp\left(-\frac{\delta^2}{2\epsilon(\delta + 2\gamma)}\right) \leq \exp\left(-\frac{\delta}{4\epsilon}\right) = \sqrt{\epsilon},$$

where the second inequality uses the fact that  $f(x) = \frac{x^2}{x+c}$  is an increasing function if  $x, c \geq 0$ . Thus, the probability that  $\mathbf{b} \notin Q$  is at most  $\sqrt{\epsilon}$ .

Suppose now that  $\langle \mathbf{u}, \mathbf{v}_0 \rangle \leq \delta$ . Using (19), we deduce that, in this case,  $\langle \mathbf{u}, \mathbf{u} \rangle \leq 2\gamma + \delta$ . Recall that  $\mathbf{r}$  is a random Gaussian sampled from the distribution  $\mathcal{N}(0_n, I_n)$ , so  $\langle \mathbf{u}, \mathbf{r} \rangle$  is distributed as a Gaussian with mean 0 and standard deviation  $\|\mathbf{u}\|$ . Using part (b) of Proposition 2.1, we find that

$$\Pr\left(|\langle \mathbf{u}, \mathbf{r} \rangle| \geq \sqrt{2(\delta + 2\gamma) \log(1/\epsilon)}\right) \leq 2 \exp\left(-\frac{2(\delta + 2\gamma) \log(1/\epsilon)}{2\|\mathbf{u}\|^2}\right) \leq 2 \exp(-\log(1/\epsilon)) = 2\epsilon.$$

Observe now that  $\langle \mathbf{u}, \mathbf{v}_0 \rangle \geq -\gamma$ . Indeed, we have

$$\langle \mathbf{u}, \mathbf{v}_0 \rangle = \sum_{\mathbf{a} \in A^k} \mu(\mathbf{a}) \langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}} \geq \sum_{\mathbf{a} \in A^k \setminus P} \mu(\mathbf{a}) \langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}} \geq -\|W\|_1 \sum_{\mathbf{a} \in A^k \setminus P} \mu(\mathbf{a}) \geq -\gamma,$$

where the first equality is (18), while the first and last inequalities come from part 3. and part 1. of Definition 5.1, respectively. In particular, this implies that  $|\langle \mathbf{u}, \mathbf{v}_0 \rangle| \leq \max(\delta, \gamma) < \delta + \gamma$ . Therefore, with probability at least  $1 - 2\epsilon$  it holds that

$$|\langle \mathbf{u}, \mathbf{z} \rangle| = |\langle \mathbf{u}, \mathbf{r} \rangle + \frac{1}{\sqrt{\epsilon}} \langle \mathbf{u}, \mathbf{v}_0 \rangle| \leq |\langle \mathbf{u}, \mathbf{r} \rangle| + \frac{1}{\sqrt{\epsilon}} |\langle \mathbf{u}, \mathbf{v}_0 \rangle| \leq \sqrt{2(\delta + 2\gamma) \log(1/\epsilon)} + \frac{\delta + \gamma}{\sqrt{\epsilon}} = \eta. \quad (20)$$

Consider now, for each  $i \in [k]$ , the sets  $A_{i,+} = \{a \in A : w_{i,a} > 0\}$  and  $A_{i,-} = \{a \in A : w_{i,a} \leq 0\}$ , and let  $c_i = \sum_{a \in A} |w_{i,a}|$ . Using part 2. of Definition 5.1, we find that  $\sum_{a \in A} w_{i,a} = 0$ , so

$$\sum_{a \in A_{i,+}} w_{i,a} = - \sum_{a \in A_{i,-}} w_{i,a}.$$

Fix  $i \in [k]$  and take some  $\lambda > 0$ . Consider the  $k \times A$  matrix  $M^{(i,\lambda)}$  having  $(i, b_i)$ -th entry 1,  $(j, b_j)$ -th entry  $\lambda$  for each  $j \neq i \in [k]$ , and all other entries 0. Recall that the function  $\rho$  is assumed to be  $\alpha$ -conservative for some  $\alpha > 0$ . It follows that  $M^{(i,\lambda)} \in \rho^{-1}(\mathbf{b})$ , so  $\langle W, M^{(i,\lambda)} \rangle_{\mathbb{F}} \leq 0$  by part 4. of Definition 5.1. Since this holds for each positive  $\lambda$ , we deduce that  $w_{i,b_i} \leq 0$ , so  $b_i \in A_{i,-}$  for each  $i \in [k]$ .

Let now  $\beta = \min(\frac{\alpha}{2}, \frac{1}{2|A|})$ . Consider the matrix  $N^{(i,\lambda)}$  defined by  $N_{i,b_i}^{(i,\lambda)} = 1 - \beta|A_{i,+}|$ ,  $N_{i,a}^{(i,\lambda)} = \beta$  for each  $a \in A_{i,+}$ ,  $N_{i,a}^{(i,\lambda)} = 0$  for each  $a \in A_{i,-} \setminus \{b_i\}$ ,  $N_{j,b_j}^{(i,\lambda)} = \lambda$  for each  $j \neq i \in [k]$ , and  $N_{j,a}^{(i,\lambda)} = 0$  elsewhere. Using again the  $\alpha$ -conservativity of  $\rho$ , we find that  $N^{(i,\lambda)} \in \rho^{-1}(\mathbf{b})$  and, thus,  $\langle W, N^{(i,\lambda)} \rangle_{\mathbb{F}} \leq 0$ . Taking the limit as  $\lambda \rightarrow 0$ , this means that  $\beta \sum_{a \in A_{i,+}} w_{i,a} \leq (\beta|A_{i,+}| - 1)w_{i,b_i}$ .

Therefore, for each  $i \in [k]$ , we have

$$c_i = \sum_{a \in A} |w_{i,a}| = \sum_{a \in A_{i,+}} w_{i,a} - \sum_{a \in A_{i,-}} w_{i,a} = 2 \sum_{a \in A_{i,+}} w_{i,a} \leq 2 \frac{\beta|A_{i,+}| - 1}{\beta} w_{i,b_i} = 2 \frac{1 - \beta|A_{i,+}|}{\beta} |w_{i,b_i}|. \quad (21)$$

The  $\alpha$ -conservativity of  $\rho$  also guarantees that

$$\langle \mathbf{v}_{i,b_i}, \mathbf{z} \rangle = c_{b_i}^{(i)} \geq \alpha \sum_{a \in A} c_a^{(i)} = \alpha \sum_{a \in A} \langle \mathbf{v}_{i,a}, \mathbf{z} \rangle = \alpha \langle \mathbf{v}_0, \mathbf{z} \rangle. \quad (22)$$

Using Proposition 2.3, we find that

$$\Pr(\|\mathbf{r}\| > \frac{1}{2\sqrt{\epsilon}}) \leq O(\exp(-1/\sqrt{\epsilon})).$$

Hence, from Cauchy–Schwarz, we deduce that

$$\langle \mathbf{v}_0, \mathbf{z} \rangle = \langle \mathbf{v}_0, \mathbf{r} + \frac{\mathbf{v}_0}{\sqrt{\epsilon}} \rangle = \langle \mathbf{v}_0, \mathbf{r} \rangle + \frac{1}{\sqrt{\epsilon}} \geq -|\langle \mathbf{v}_0, \mathbf{r} \rangle| + \frac{1}{\sqrt{\epsilon}} \geq -\|\mathbf{r}\| + \frac{1}{\sqrt{\epsilon}} \geq \frac{1}{2\sqrt{\epsilon}} \quad (23)$$

with probability  $1 - O(\exp(-1/\sqrt{\epsilon}))$ . Similarly, we have

$$\langle \mathbf{v}_{i,b_i}, \mathbf{z} \rangle \leq \|\mathbf{v}_{i,b_i}\| \cdot \|\mathbf{z}\| = \|\mathbf{v}_{i,b_i}\| \cdot \|\mathbf{r} + \frac{\mathbf{v}_0}{\sqrt{\epsilon}}\| \leq \|\mathbf{v}_{i,b_i}\| \cdot (\|\mathbf{r}\| + \frac{1}{\sqrt{\epsilon}}) \leq \frac{2}{\sqrt{\epsilon}} \|\mathbf{v}_{i,b_i}\| \quad (24)$$

with probability  $1 - O(\exp(-1/\sqrt{\epsilon}))$ . Combining (22), (23), and (24), we deduce that

$$\|\mathbf{v}_{i,b_i}\| \geq \frac{\sqrt{\epsilon}}{2} \langle \mathbf{v}_{i,b_i}, \mathbf{z} \rangle \geq \frac{\alpha\sqrt{\epsilon}}{2} \langle \mathbf{v}_0, \mathbf{z} \rangle \geq \frac{\alpha}{4} \quad (25)$$

with probability  $1 - O(\exp(-1/\sqrt{\epsilon}))$ .

Let us now define  $\mathbf{y}_i = \sum_{a \in A} w_{i,a} \mathbf{v}_{i,a}$  for each  $i \in [k]$ . Observe that  $\beta|A_{i,+}| \leq \beta|A| \leq \frac{1}{2}$ . Combining (21) with (25), we obtain that

$$\|\mathbf{y}_i\| = \sqrt{\sum_{a \in A} w_{i,a}^2 \|\mathbf{v}_{i,a}\|^2} \geq |w_{i,b_i}| \|\mathbf{v}_{i,b_i}\| \geq \frac{\beta c_i}{2(1 - \beta|A_{i,+}|)} \cdot \frac{\alpha}{4} = \frac{\alpha\beta c_i}{8(1 - \beta|A_{i,+}|)} \geq \frac{\alpha\beta c_i}{8}$$

with probability  $1 - O(\exp(-1/\sqrt{\epsilon}))$ . Note now that

$$\Pr\left(|\langle \mathbf{y}_i, \mathbf{z} \rangle| < \frac{8}{\alpha\beta} \eta \|\mathbf{y}_i\|\right) \leq \frac{\frac{16}{\alpha\beta} \eta \|\mathbf{y}_i\|}{2\|\mathbf{y}_i\|} = \frac{8}{\alpha\beta} \eta$$

by part (a) of Proposition 2.1. Using part 4. of Definition 5.1 in a similar fashion as before, we find that  $\langle \mathbf{y}_i, \mathbf{z} \rangle \leq 0$ ; thus,

$$\Pr\left(\langle \mathbf{y}_i, \mathbf{z} \rangle < -\frac{8}{\alpha\beta} \eta \|\mathbf{y}_i\|\right) \geq 1 - \frac{8}{\alpha\beta} \eta.$$

As a consequence, by the union bound, the property

$$\langle \mathbf{y}_i, \mathbf{z} \rangle < -\frac{8}{\alpha\beta} \eta \|\mathbf{y}_i\| \quad \text{for each } i \in [k]$$

holds with probability  $1 - O_{\alpha,A}(\eta)$ . Putting all together, it follows that, with probability  $1 - O_{\alpha,A}(\eta)$ ,

$$\langle \mathbf{u}, \mathbf{z} \rangle = \sum_{i \in [k]} \langle \mathbf{y}_i, \mathbf{z} \rangle < -\frac{8}{\alpha\beta} \eta \sum_{i \in [k]} \|\mathbf{y}_i\| \leq -\eta \sum_{i \in [k]} c_i = -\eta \|W\|_1 = -\eta,$$

thus contradicting (20). It follows that, in this case,  $\mathbf{b} \in Q$  with probability  $1 - O_{\alpha,A}(\eta)$ , as needed.

In summary, in both cases, the loss of SDP is at most  $O_{\alpha,A}(\sqrt{\epsilon} \log(1/\epsilon))$ , as required.  $\square$

It is easy to check that a Boolean PCSP admitting majority polymorphisms of all (odd) arities satisfies the hypotheses of Theorem 5.3. (Note, however, that the loss in the theorem above is slightly worse—by a  $\log(1/\epsilon)$  factor—than the  $O(\sqrt{\epsilon})$  loss we managed to obtain in the Boolean majority case, see Theorem 4.1.) More interestingly, Theorem 5.3 can be applied to other classes of PCSPs, whose robust solvability was not known. This is done by relating the polymorphism families defining such PCSPs to corresponding rounding functions as per Definition 5.1. In particular, we now show that any PCSP containing *plurality* polymorphisms of all arities satisfies the hypotheses of Theorem 5.3 and, thus, it is robustly solvable.

**Definition 5.4** ([BWZ21]). *A function  $f : A^n \rightarrow A$  is a plurality if*

$$f(\mathbf{a}) = \operatorname{argmax}_{a \in A} \{\# \text{ of occurrences of } a \text{ in } \mathbf{a}\}$$

for all  $\mathbf{a} \in A^n$ , with ties broken in a way that  $f$  is symmetric.

**Proposition 5.5.** *Let  $(\mathbf{A}, \mathbf{B})$  be a PCSP template such that  $\operatorname{Pol}(\mathbf{A}, \mathbf{B})$  contains plurality polymorphisms of all arities. Then  $\operatorname{PCSP}(\mathbf{A}, \mathbf{B})$  is robustly solvable by SDP with loss  $O(\sqrt{\epsilon} \log(1/\epsilon))$ .*

*Proof.* Consider any function  $\rho : \mathbb{R}^A \rightarrow A$  satisfying  $q_{\rho(\mathbf{q})} \geq q_a$  for each  $\mathbf{q} \in \mathbb{R}^A$ ,  $a \in A$  (i.e.,  $\rho$  is the  $\operatorname{argmax}$  function, with ties broken arbitrarily). Note that  $\rho$  is  $\alpha$ -conservative for  $\alpha = \frac{1}{|A|}$ . We claim that  $(\mathbf{A}, \mathbf{B})$  is  $\rho$ -separable.

Fix a promise relation  $(P, Q)$  of  $(\mathbf{A}, \mathbf{B})$  of arity  $k$ , and choose some tuple  $\mathbf{b} \in B^k \setminus Q$ . Let  $K \subseteq \mathbb{R}^{k \times A}$  be the convex hull of the set  $\{\Pi_{\mathbf{a}} : \mathbf{a} \in P\}$ . Let also  $L$  be the subset of  $\mathbb{R}^{k \times A}$  for which  $\mathbf{b}$  is the vector of (strict) plurality coordinates. That is,  $M \in L$  if and only if, for each  $i \in [k]$  and each  $a \in A \setminus \{b_i\}$ , it holds that  $M_{i,b_i} > M_{i,a}$ . Note that  $K$  is closed and  $L$  is open, and they are both convex sets. We claim that they are disjoint. Suppose by contradiction that  $M \in K \cap L$ . Since  $M$  is constrained by a system of inequalities with rational coefficients, we may assume there exists a family  $\{\lambda_{\mathbf{a}} : \mathbf{a} \in P\}$  of rational nonnegative weights summing up to 1 such that  $M = \sum_{\mathbf{a} \in P} \lambda_{\mathbf{a}} \Pi_{\mathbf{a}}$ . Take a positive integer  $n$  such that  $n\lambda_{\mathbf{a}}$  is an integer for each  $\mathbf{a}$ , and choose a plurality polymorphism  $f \in \operatorname{Pol}(\mathbf{A}, \mathbf{B})$  of arity  $n$ . Let  $\mathbf{c} \in A^k$  be the tuple obtained by applying  $f$  coordinatewise to a list of tuples  $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots, \mathbf{a}^{(n)} \in P$ , where each  $\mathbf{a}$  appears  $n\lambda_{\mathbf{a}}$ -many times. Since  $f$  is a polymorphism,  $\mathbf{c} \in Q$ . On the other hand,  $M \in L$  implies that  $M_{i,b_i} > M_{i,a}$  for each  $i \in [k]$  and each  $b_i \neq a \in A$ . By definition of plurality, we deduce that  $\mathbf{c} = \mathbf{b} \notin Q$ , a contradiction. It follows that  $K \cap L = \emptyset$ , as claimed.

Using well-known results on hyperplane separation of convex bodies [BV04], since  $\mathbb{R}^{k \times A}$  is finite dimensional, there exists a linear operator  $W \in \mathbb{R}^{k \times A}$  and a scalar  $\mu \in \mathbb{R}$  such that for every  $M \in K$  it holds that  $\langle W, M \rangle_{\mathbb{F}} \geq \mu$ , and for every  $M \in L$  it holds that  $\langle W, M \rangle_{\mathbb{F}} < \mu$ . We claim that  $W$  is our desired matrix of weights as required by Definition 5.1.

First, observe that, since  $L$  is an open cone, we have that  $\langle W, M \rangle_{\mathbb{F}} < \mu$  for all  $M \in L$  if and only if  $\langle W, \lambda M \rangle_{\mathbb{F}} < \mu$  for all  $M \in L$  and  $\lambda > 0$ . By sending  $\lambda \rightarrow \infty$ , we have that  $\langle W, M \rangle_{\mathbb{F}} < 0$  for all  $M \in L$  and the RHS of this inequality cannot be made any smaller. In other words, we can assume without loss of generality that  $\mu = 0$ . Then, we may also assume without loss of generality that the 1-norm of  $W$  is 1, by simply rescaling it if necessary. Hence, 1. holds. The condition  $\langle W, M \rangle_{\mathbb{F}} \geq 0$  for all  $M \in K$  is equivalent to  $\langle W, \Pi_{\mathbf{a}} \rangle_{\mathbb{F}} \geq 0$  for all  $\mathbf{a} \in P$ , so 3. holds. To prove 4., just observe that  $\rho^{-1}(\mathbf{b})$  is a subset of the topological closure of  $L$ . Hence, 4. follows from the continuity of the map  $\bullet \mapsto \langle W, \bullet \rangle_{\mathbb{F}}$ .

In order to prove condition 2., choose  $i \in [k]$ . For each  $\alpha, \beta \geq 0$ , consider the following matrices:

- $D^{(\alpha)}$  is the  $k \times k$  diagonal matrix whose  $i$ -th diagonal entry is 1 and whose  $j$ -th diagonal entry is  $\alpha$  for each  $j \neq i$ ;
- $J$  is the  $k \times A$  all-one matrix;
- $M^{(\alpha,\beta)} = D^{(\alpha)}(J + \beta\Pi_{\mathbf{b}})$ ;
- $N^{(\alpha,\beta)} = D^{(\alpha)}(-J + \beta\Pi_{\mathbf{b}})$ .

Observe that  $\rho(M^{(\alpha,\beta)}) = \rho(N^{(\alpha,\beta)}) = \mathbf{b}$  for each  $\alpha > 0, \beta > 0$ . Hence, by condition 4., it holds that  $\langle W, M^{(\alpha,\beta)} \rangle_{\mathbb{F}} \leq 0$  and  $\langle W, N^{(\alpha,\beta)} \rangle_{\mathbb{F}} \leq 0$ . Using the continuity of the Frobenius inner product, we deduce that

$$0 \geq \langle W, M^{(0,0)} \rangle_{\mathbb{F}} = \sum_{a \in A} w_{i,a} \quad \text{and} \quad 0 \geq \langle W, N^{(0,0)} \rangle_{\mathbb{F}} = - \sum_{a \in A} w_{i,a},$$

thus proving that  $W\mathbf{1} = \mathbf{0}$ , as required.

In summary, we have shown that  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  is  $\rho$ -separable for a  $\frac{1}{|A|}$ -conservative function  $\rho$ . Hence, the result follows from Theorem 5.3.  $\square$

## 6 Equality Preserves Robustness

Given a domain  $A$ , we define  $\text{EQ}_A := \{(a, a) : a \in A\}$ . Likewise, for a domain pair  $(A, B)$ , we define  $\text{EQ}_{A,B}$  to be the promise relation  $(\text{EQ}_A, \text{EQ}_B)$ . As an informal shorthand, given a promise template  $(\mathbf{A}, \mathbf{B})$ , we let  $(\mathbf{A}, \mathbf{B}) + \text{EQ}$  denote the promise template with  $\text{EQ}_{A,B}$  added. Fix parameters  $0 < \alpha < \beta < 1$ . We say that  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  is  $(\alpha, \beta)$ -robust if there exists a polynomial-time algorithm that on input an instance on variable set  $V$  with the promise that there is a map  $g : V \rightarrow A$  that strongly satisfies at least  $1 - \alpha$  fraction of the constraints, outputs the description of a map  $g' : V \rightarrow B$  that weakly satisfies at least  $1 - \beta$  fraction of the constraints.

The main goal of this section is to prove the following theorem.

**Theorem 6.1.** *Assume UGC. If  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  is  $(\alpha, \beta)$ -robust for some  $0 < \alpha < \beta < 1$ , then  $\text{PCSP}((\mathbf{A}, \mathbf{B}) + \text{EQ})$  is  $(\alpha_{\text{EQ}}, \beta_{\text{EQ}})$ -robust for*

$$\alpha_{\text{EQ}} = \Omega_{\mathbf{A}}(\alpha^2\beta^4), \tag{26}$$

$$\beta_{\text{EQ}} = O_{\mathbf{B}}(\beta). \tag{27}$$

It directly follows from the result above that the existence of a *minion homomorphism*<sup>16</sup>  $\text{Pol}(\mathbf{A}, \mathbf{B}) \rightarrow \text{Pol}(\mathbf{C}, \mathbf{D})$  between the polymorphism sets of two PCSP templates preserves robust solvability under UGC.

**Corollary 6.2.** *Under UGC, if  $\text{Pol}(\mathbf{A}, \mathbf{B}) \rightarrow \text{Pol}(\mathbf{C}, \mathbf{D})$  and  $(\mathbf{A}, \mathbf{B})$  is  $f$ -robust, then  $(\mathbf{C}, \mathbf{D})$  is  $g$ -robust, where  $g(\epsilon) = O_{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}}(f(\epsilon^{1/6}))$ .*

*Proof.* We start by introducing some terminology from algebraic PCSP theory, following [BG21a, BBKO21]. Fix a relational signature  $\sigma$  (i.e., a finite set of relation symbols  $R_1, R_2, \dots$ , each with an integer arity  $r_1, r_2, \dots$ ). A *pp-formula* over  $\sigma$  is a formal expression  $\psi$  consisting of an existentially

<sup>16</sup>For the definition of minions and their homomorphisms, we refer the reader to [BBKO21].

quantified conjunction of predicates of the form (i) “ $x = y$ ”, or (ii) “ $(x_{i_1}, \dots, x_{i_r}) \in R$ ” for some  $R \in \sigma$  of arity  $r$ , where  $x, y, x_{i_1}, \dots, x_{i_r}$  are variables. Let  $k$  be the number of free (i.e., unquantified) variables in  $\psi$ . Given a  $\sigma$ -structure  $\mathbf{A}$ , the interpretation of  $\psi$  in  $\mathbf{A}$  is the set  $\psi(\mathbf{A}) \subseteq A^k$  containing all tuples  $(a_1, \dots, a_k) \in A^k$  that satisfy  $\psi$ , where each symbol  $R$  appearing in  $\psi$  is interpreted in  $\mathbf{A}$ . Let now  $\mathbf{A}'$  be a  $\sigma'$ -structure for some possibly different signature  $\sigma'$ , such that  $A' = A$ . We say that  $\mathbf{A}'$  is *pp-definable* from  $\mathbf{A}$  if for each symbol  $S \in \sigma'$  it holds that  $S^{\mathbf{A}'} = \psi_S(\mathbf{A})$  for some pp-formula  $\psi_S$  over  $\sigma$ . Suppose now that  $A' = A^n$  for some  $n \in \mathbb{N}$ , and let  $\text{vec}_n(\mathbf{A}')$  be the structure with domain  $A$  and relations defined as follows: For each  $r$ -ary symbol  $S \in \sigma'$ ,  $\text{vec}_n(\mathbf{A}')$  has an  $rn$ -ary relation containing the tuple  $(b_1^{(1)}, \dots, b_n^{(1)}, \dots, b_1^{(r)}, \dots, b_n^{(r)})$  for each tuple  $(\mathbf{b}^{(1)}, \dots, \mathbf{b}^{(r)}) \in S^{\mathbf{A}'}$ . We say that  $\mathbf{A}'$  is an  $n$ -fold *pp-power* of  $\mathbf{A}$  if  $\text{vec}_n(\mathbf{A}')$  is pp-definable from  $\mathbf{A}$ .

Take now  $\mathbf{A}, \mathbf{B}, \mathbf{C}$ , and  $\mathbf{D}$  as in the statement of the corollary, let  $\sigma$  be the signature of  $\mathbf{A}, \mathbf{B}$ , and let  $\sigma'$  be the signature of  $\mathbf{C}, \mathbf{D}$ . It was shown in [BBKO21, Theorem 4.12] that the existence of a minion homomorphism  $\text{Pol}(\mathbf{A}, \mathbf{B}) \rightarrow \text{Pol}(\mathbf{C}, \mathbf{D})$  is equivalent to the existence of two  $\sigma'$ -structures  $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}$  such that (i)  $\mathbf{C} \rightarrow \tilde{\mathbf{A}} \rightarrow \tilde{\mathbf{B}} \rightarrow \mathbf{D}$ , and (ii)  $\tilde{\mathbf{A}}$  and  $\tilde{\mathbf{B}}$  are pp-powers of  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, for the same power  $n \in \mathbb{N}$  and via the same pp-formulae.

Recall that we are assuming that  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  is  $f$ -robust. Using Theorem 6.1 we have that  $\text{PCSP}((\mathbf{A}, \mathbf{B}) + \text{EQ})$  is  $g$ -robust for  $g(\epsilon) = f(\epsilon^{1/6})$ . Using [BGS23b, Proposition 2.10], we deduce that  $\text{PCSP}(\text{vec}_n(\tilde{\mathbf{A}}), \text{vec}_n(\tilde{\mathbf{B}}))$  is  $O_{\mathbf{A}, \mathbf{B}, \tilde{\mathbf{A}}, \tilde{\mathbf{B}}}(1) \cdot g$ -robust. Consider now the (standard) reduction from  $\text{PCSP}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$  to  $\text{PCSP}(\text{vec}_n(\tilde{\mathbf{A}}), \text{vec}_n(\tilde{\mathbf{B}}))$  defined as follows. For an instance  $\mathbf{X}$  of  $\text{PCSP}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ , output an instance  $\hat{\mathbf{X}}$  of  $\text{PCSP}(\text{vec}_n(\tilde{\mathbf{A}}), \text{vec}_n(\tilde{\mathbf{B}}))$  whose domain contains  $n$  copies  $x^{(1)}, \dots, x^{(n)}$  of each  $x \in X$ , and whose relations  $R^{\hat{\mathbf{X}}}$  contain the tuple  $(x_1^{(1)}, \dots, x_1^{(n)}, \dots, x_r^{(1)}, \dots, x_r^{(n)})$  for each tuple  $(x_1, \dots, x_r)$  in  $R^{\mathbf{X}}$ . Clearly, such reduction preserves robust solvability with the same loss, so we deduce that  $\text{PCSP}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$  is  $O_{\mathbf{A}, \mathbf{B}, \tilde{\mathbf{A}}, \tilde{\mathbf{B}}}(1) \cdot g$ -robust. Finally, since  $(\mathbf{C}, \mathbf{D})$  is a homomorphic relaxation of  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ , we conclude that the same holds for  $\text{PCSP}(\mathbf{C}, \mathbf{D})$ , as required.  $\square$

As discussed in the introduction, Corollary 6.2 partially confirms Question 1.4 due to Barto and Kozik [BK16]. The proof of Theorem 6.1 is rather intricate as it involves substantial modifications to the algorithm of Brown-Cohen and Raghavendra [BR16], which itself is a modified version of Raghavendra’s algorithm [Rag08]. We begin this section with an overview of Brown-Cohen and Raghavendra’s algorithm (henceforth called the *BCR algorithm*).

## 6.1 An Overview of the Algorithm of Brown-Cohen and Raghavendra

In this section, we describe the BCR algorithm and its main correctness guarantee. We begin with the necessary preliminaries.

### 6.1.1 BCR Preliminaries

A key ingredient in the BCR algorithm is a notion of *approximate polymorphisms*. For a promise template  $(\mathbf{A}, \mathbf{B})$  and a positive integer  $R$ , consider a probability distribution  $\mathcal{P}$  of maps  $p : A^R \rightarrow B$ .

**Definition 6.3** (Approximate Polymorphism [BR16]). *We say that  $\mathcal{P}$  is a  $(c, s)$ -approximate polymorphism of  $(\mathbf{A}, \mathbf{B})$  if for all (normalized) weighted instances  $\mathcal{X}$  of  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  on variable set  $X$ , given  $R$  assignments  $a_1, \dots, a_R : X \rightarrow A$  each with value at least  $c$ , then*

$$\mathbb{E}_{p \sim \mathcal{P}}[\text{val}_{\mathcal{X}}(p \circ (a_1, \dots, a_R))] \geq s,$$

where for all  $x \in X$ , we have that  $(p \circ (a_1, \dots, a_R))(x) = p(a_1(x), \dots, a_R(x))$ .

We also define  $\mathcal{P}$  to be an  $(\alpha, \beta)$ -robust polymorphism if  $\mathcal{P}$  is a  $(1 - \alpha, 1 - \beta)$ -approximate polymorphism. Observe that there are many trivial approximate polymorphisms with good parameters, such as probability distributions supported only on dictators. To identify the “meaningful” approximate polymorphisms, BCR use a notion of *quasirandomness* to ensure that the robust/approximate polymorphism is sufficiently nontrivial.

We start by recalling some standard Fourier-analytic notions. Fix a probability distribution  $\mu \in \Delta_A$ , and let  $\{\chi_0 = 1, \dots, \chi_{|A|-1}\}$  be an orthonormal basis for the vector space  $L_2(A, \mu)$ . Any function  $f : A^R \rightarrow \mathbb{R}$  can be written as  $f = \sum_{\sigma \in \mathbb{N}^R} \hat{f}_\sigma \chi_\sigma$ , where  $\chi_\sigma(x) = \prod_{j=1}^R \chi_{\sigma_j}(x_j)$ , and the  $\hat{f}_\sigma$  are the Fourier coefficients of  $f$ . Let  $|\sigma|$  denote the number of non-zero coordinates of  $\sigma$ . The *degree  $d$  influence* of the  $i$ -th coordinate of  $f$  (under the probability measure  $\mu$ ) is defined as

$$\text{Inf}_{i,\mu}^{<d}(f) = \sum_{\substack{\sigma \in \mathbb{N}^R \\ \sigma_i \neq 0 \\ |\sigma| < d}} \hat{f}_\sigma^2.$$

We say that an approximate polymorphism  $\mathcal{P}$  is  $(\tau, d)$ -*quasirandom* if, for every probability measure  $\mu \in \Delta_A$ , it holds that

$$\mathbb{E}_{p \sim \mathcal{P}} \left[ \max_{i \in [R]} \text{Inf}_{i,\mu}^{<d}(p) \right] \leq \tau.$$

In order to describe the BCR algorithm, we shall also need to use *noise operators*. Given a parameter  $\rho \in [0, 1]$ , we let  $T_\rho$  be the operator defined as follows: for each function  $f : A^R \rightarrow \mathbb{R}$ ,

$$T_\rho f(x) = \mathbb{E}_{y \sim_\rho x} [f(y)],$$

where the expression “ $y \sim_\rho x$ ” denotes that, for each  $i \in [R]$ ,  $y_i$  equals  $x_i$  with probability  $\rho$ , and  $y_i$  is independently sampled according to  $\mu$  with probability  $1 - \rho$ . It follows from this definition that the Fourier expansion of the operator is given by  $T_\rho f = \sum_{\sigma \in \mathbb{N}^R} \hat{f}_\sigma \rho^{|\sigma|} \chi_\sigma$ .

### 6.1.2 The BCR Algorithm

The main result of Brown-Cohen and Raghavendra [BR16] is as follows.

**Theorem 6.4.** *For any PCSP template  $(\mathbf{A}, \mathbf{B})$  and any  $\theta > 0$ , there exists  $\tau > 0$  and  $d \in \mathbb{N}$  such that, for each  $0 < \alpha < \beta < 1$ , if  $(\mathbf{A}, \mathbf{B})$  admits an  $(\alpha, \beta)$ -robust  $(\tau, d)$ -quasirandom polymorphism, then the BCR algorithm is  $(\alpha - \theta, \beta + \theta)$ -robust for PCSP $(\mathbf{A}, \mathbf{B})$ . Conversely, assuming UGC, if PCSP $(\mathbf{A}, \mathbf{B})$  has any  $(\alpha, \beta)$ -robust algorithm, then for any  $\theta > 0$ , PCSP $(\mathbf{A}, \mathbf{B})$  admits an  $(\alpha - \theta, \beta + \theta)$ -robust algorithm via the aforementioned algorithmic procedure.*

We now present the BCR algorithm. We let  $\Psi : \mathbb{R}^A \rightarrow \Delta_A$  be a Lipschitz-continuous function extending the identity function on  $\Delta_A$ .

- Let  $\mathcal{P}$  be an  $R$ -arity  $(d, \tau)$ -quasirandom  $(\alpha, \beta)$ -robust polymorphism. Pick  $\eta > 0$  suitably small.
- Let  $\mathcal{X}$  be a weighted instance of PCSP $(\mathbf{A}, \mathbf{B})$  on variable set  $X$ .

- Solve the basic SDP for  $\mathcal{X}$  with respect to  $\text{CSP}(\mathbf{A})$  to get a unit vector  $\mathbf{v}_0 \in \mathbb{R}^N$  and vectors  $\mathbf{v}_{x,a} \in \mathbb{R}^N$  for all  $x \in X$  and  $a \in A$  with completeness at least  $1 - \alpha$  (see Section 6.2.2).
- Sample  $p \sim \mathcal{P}$  and independently sample  $R$  vectors  $\zeta^{(1)}, \dots, \zeta^{(R)} \sim \mathcal{N}(0^N, I_N)$ .
- For all  $x \in X$ ,  $a \in A$ , and  $i \in [R]$ , compute

$$g_{x,a,i} := \langle \mathbf{v}_{x,a}, \mathbf{v}_0 \rangle + \langle \mathbf{v}_{x,a} - \mathbf{v}_0 \langle \mathbf{v}_0, \mathbf{v}_{x,a} \rangle, \zeta^{(i)} \rangle.$$

- For each  $x \in X$  and  $i \in [R]$ , compute

$$h_{x,i} := \Psi(g_{x,a,i} : a \in A).$$

- Let  $H_\eta$  be the multilinear polynomial corresponding to the function  $T_{1-\eta}p : \Delta_A^R \rightarrow \Delta_A$  (recall the definition of the noise operator  $T_{1-\eta}$ ). For each  $x \in X$ ,

$$q_x := (H_\eta p)(h_{x,1}, \dots, h_{x,R}) \in \Delta_B.$$

- For each  $x \in X$ , assign  $\sigma(x)$  according to the probability distribution  $q_x$ .

## 6.2 Configurations and Rounding Schemes

In order to describe our modifications to the BCR algorithm, we need to first abstract out the essential properties of the SDP rounding scheme.

### 6.2.1 Configurations

Given a unit vector  $\mathbf{v}_0 \in \mathbb{R}^N$ , we define  $\mathbb{S}_A^N(\mathbf{v}_0) \subseteq (\mathbb{R}^N)^A$  to be the set of all  $A$ -tuples of vectors  $\{\mathbf{v}_a : a \in A\}$  satisfying the following criteria:

$$\begin{aligned} \langle \mathbf{v}_a, \mathbf{v}_{a'} \rangle &= 0 \quad \forall a \neq a' \in A \\ \sum_{a \in A} \mathbf{v}_a &= \mathbf{v}_0. \end{aligned}$$

We call these  $A$ -tuples *local configurations* and typically denote them by  $\mathbf{V} \in \mathbb{S}_A^N$ . Note for each  $a \in A$ , there is an *integral* local configuration  $\mathbf{V} \in \mathbb{S}_A^N$  for which  $\mathbf{v}_a = \mathbf{v}_0$  and  $\mathbf{v}_{a'} = 0$  for all  $a' \in A \setminus \{a\}$ . We let  $\mathbf{I}_a$  denote this integral local configuration.

When  $\mathbf{v}_0$  is fixed in a context, for succinctness we let  $\mathbb{S}_A^N$  denote  $\mathbb{S}_A^N(\mathbf{v}_0)$ . We define a (normed) metric on  $\mathbb{S}_A^N$  by

$$\|\mathbf{V} - \mathbf{V}'\|_2 = \sqrt{\sum_{a \in A} \|\mathbf{v}_a - \mathbf{v}'_a\|_2^2}.$$

Note that, if  $\mathbf{V}$  and  $\mathbf{V}'$  are interpreted as matrices in  $\mathbb{R}^{N \times A}$ , the above is the Frobenius matrix norm.

We now define a *global configuration* to be a tuple of local configurations

$$\mathcal{V} := (\mathbf{V}_x \in \mathbb{S}_A^N : x \in X) \in (\mathbb{S}_A^N)^X.$$

In particular, given  $\mathbf{a} \in A^X$ , we let  $\mathcal{I}_{\mathbf{a}} := (\mathbf{I}_{a_x} : x \in X)$  denote the integral global configuration. Given  $\mathcal{V} \in (\mathbb{S}_A^N)^X$ , we let  $\text{Mat}(\mathcal{V}) \in \mathbb{R}^{(X \times A)^2}$  denote the Gram matrix of all dot products. That is, for any  $(x, a), (x', a') \in X \times A$ , we have that

$$\text{Mat}(\mathcal{V})_{(x,a),(x',a')} := \langle \mathbf{v}_{x,a}, \mathbf{v}_{x',a'} \rangle.$$

For  $\mathbf{a} \in A^X$ , we let  $\text{Mat}(\mathbf{a})$  be shorthand for  $\text{Mat}(\mathcal{I}_{\mathbf{a}})$ . In other words  $\text{Mat}(\mathbf{a})_{(x,a),(x',a')} = 1$  iff  $\mathbf{a}_x = a$  and  $\mathbf{a}_{x'} = a'$ , and 0 otherwise. Observe that  $\text{Mat}(\mathbf{a})$  can be identified with the Kronecker product  $\Pi_{\mathbf{a}} \otimes \Pi_{\mathbf{a}}$ , with  $\Pi_{\mathbf{a}}$  defined as in Section 5.

It shall be useful to notice that, for any global configuration  $\mathcal{V}$ , it holds that

$$\text{Tr}(\text{Mat}(\mathcal{V})) = \sum_{(x,a) \in X \times A} \|\mathbf{v}_{x,a}\|_2^2 = \sum_{x \in X} \|\mathbf{v}_0\|_2^2 = |X|. \quad (28)$$

### 6.2.2 SDP solutions and Rounding Schemes

Given a (weighted) instance  $\mathcal{X}$  of PCSP( $\mathbf{A}, \mathbf{B}$ ) on variable set  $X$ , any valid solution to the SDP relaxation of  $\mathcal{X}$  can be viewed as a global configuration  $\mathcal{V} \in (\mathbb{S}_A^N)^X$  for some suitably large dimension  $N$  (which can be made at most  $|X| \cdot |A|$ , although we will often consider  $N$  to be larger). However, not every  $\mathcal{V} \in (\mathbb{S}_A^N)^X$  is a valid solution to the SDP as each clause  $(Y, P, Q)$  of  $\mathcal{X}$  imposes some structure on  $\mathcal{V}$ . More precisely, if we let  $\mathcal{V}|_Y := (\mathbf{V}_y : y \in Y)$ , then there exists a probability distribution  $\Lambda_{Y,P}$  on  $A^Y$  such that

$$\text{Mat}(\mathcal{V}|_Y) = \mathbb{E}_{\mathbf{a} \sim \Lambda_{Y,P}} [\text{Mat}(\mathbf{a})]. \quad (29)$$

We say that  $\Lambda_{Y,P}$  is *consistent* with  $\mathcal{V}$  if (29) holds. Note that this is the translation of the second-moment condition (3), while the first-moment condition (2) is implicit in the definition of local configurations.

We let  $\mathbb{V}_{\mathcal{X},A}^N \subseteq (\mathbb{S}_A^N)^X$  be the set of all  $\mathcal{V} \in (\mathbb{S}_A^N)^X$  that satisfy (29) for some choice of  $\Lambda_{Y,P}$  for all constraints  $(Y, P, Q)$  of  $\mathcal{X}$ . We further define the map  $\text{comp}_{\mathcal{X},A} : \mathbb{V}_{\mathcal{X},A}^N \rightarrow [0, 1]$  to be

$$\text{comp}_{\mathcal{X},A}(\mathcal{V}) := \max_{\substack{\{\Lambda_{Y,P} | (Y,P,Q) \in \mathcal{X}\} \\ \text{consistent with } \mathcal{V}}} \mathbb{E}_{(Y,P,Q) \sim \mathcal{X}} \left[ \Pr_{\mathbf{a} \sim \Lambda_{Y,P}} [\mathbf{a} \in P] \right].$$

Analogously, we define  $\text{sound}_{\mathcal{X},B} : \Delta_B^X \rightarrow [0, 1]$  to be

$$\text{sound}_{\mathcal{X},B}(\mathbf{z}) := \mathbb{E}_{(Y,P,Q) \sim \mathcal{X}} \left[ \sum_{\mathbf{q} \in Q} \prod_{y \in Y} z_y(q_y) \right].$$

We now introduce the notion of an (oblivious) rounding scheme.

**Definition 6.5** (Rounding scheme). *Let  $\Xi$  be a probability distribution. We define an (oblivious) rounding scheme to be a function  $S : \mathbb{S}_A^N \times \Xi \rightarrow \Delta_B$ . If  $S$  can be computed efficiently, then we say that  $S$  is a rounding algorithm.*

**Definition 6.6** (Robust rounding scheme). *We say that a rounding scheme  $S : \mathbb{S}_A^N \times \Xi \rightarrow \Delta_B$  is  $(\alpha, \beta)$ -robust for an instance  $\mathcal{X}$  of PCSP( $\mathbf{A}, \mathbf{B}$ ) if for all global configurations  $\mathcal{V} \in \mathbb{V}_{\mathcal{X},A}^N$ , we have that*

$$\text{comp}_{\mathcal{X},A}(\mathcal{V}) \geq 1 - \alpha \implies \mathbb{E}_{\xi \sim \Xi} [\text{sound}_{\mathcal{X},B}(S(\mathbf{V}_x, \xi) : x \in X)] \geq 1 - \beta.$$

For the BCR algorithm, we consider the probability distribution  $\Xi_{N,\theta}^{\text{BCR}} := \mathcal{P} \times \mathcal{N}(0^N, I_N)^R$ , where  $\mathcal{P}$  is the  $R$ -arity  $(\alpha, \beta)$ -robust, quasirandom polymorphism that gives a loss of  $\theta$  in Theorem 6.4. Then, we can think of the BCR algorithm as a rounding algorithm  $\text{BCR}_\theta^N : \mathbb{S}_A^N \times \Xi_{N,\theta}^{\text{BCR}} \rightarrow \Delta_B$  for any choices of  $N \in \mathbb{N}$  and  $\theta > 0$ . In particular, once the SDP solution is found, the assigned probability distribution to each variable only depends on the global randomness chosen by  $\Xi_{N,\theta}^{\text{BCR}}$ . We can then restate the second implication in Theorem 6.4 as follows.

**Corollary 6.7.** *Assume UGC. If  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  is  $(\alpha, \beta)$ -robust, then for any  $\theta > 0$  and for all sufficiently large  $N$  (as a function of  $\theta$ ),  $\text{BCR}_\theta^N : \mathbb{S}_A^N \times \Xi_{N,\theta}^{\text{BCR}} \rightarrow \Delta_B$  is  $(\alpha - \theta, \beta + \theta)$ -robust for any instance  $\mathcal{X}$  of  $\text{PCSP}(\mathbf{A}, \mathbf{B})$ .*

Using Corollary 6.7 as the only property of the BCR algorithm we need, we shall construct another rounding procedure meeting the requirements of Theorem 6.1.

### 6.3 Smoothing Configurations

We say that  $\mathcal{V}, \mathcal{W} \in (\mathbb{S}_A^N)^X$  are  $\delta$ -modifications of each other for some  $\delta \geq 0$  if for all  $x \in X$ , we have that  $\|\mathbf{V}_x - \mathbf{W}_x\|_2^2 \leq \delta$ . We let  $B(\mathcal{V}, \delta) \subseteq \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$  denote the set of all  $\delta$ -modifications of  $\mathcal{V}$ . Note that if  $\mathcal{V}$  and  $\mathcal{W}$  are  $\delta$ -modifications with  $\mathcal{V} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$ , it may not be the case that  $\mathcal{W} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$ , there may be a clause  $(Y, P, Q)$  of  $\mathcal{X}$  for which no marginal distribution  $\Lambda_{Y,P}$  is consistent with  $\mathcal{W}$ .

As a key definition, let  $\mathbb{V}_{\mathcal{X}, \mathbf{A}}^{N, \delta} := \bigcup_{\mathcal{V} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N} B(\mathcal{V}, \delta)$ . As an important step in constructing our rounding scheme, we first seek to extend  $\text{BCR}_\theta^N$  from being robust for  $\mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$  to being robust for  $\mathbb{V}_{\mathcal{X}, \mathbf{A}}^{N, \delta}$ .

**Definition 6.8** (Spaciously Robust rounding scheme). *Given an instance  $\mathcal{X}$  of  $\text{PCSP}(\mathbf{A}, \mathbf{B})$ , we say that a rounding scheme  $S : \mathbb{S}_A^N \times \Xi \rightarrow \Delta_B$  is  $(\delta, \alpha, \beta)$ -spaciously robust for  $\mathcal{X}$  if for all global configurations  $\mathcal{V} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$ , and all  $\mathcal{W} \in B(\mathcal{V}, \delta)$ , we have that*

$$\text{comp}_{\mathcal{X}, \mathbf{A}}(\mathcal{V}) \geq 1 - \alpha \implies \mathbb{E}_{\xi \sim \Xi} [\text{sound}_{\mathcal{X}, \mathbf{B}}(S(\mathbf{W}_x, \xi) : x \in X)] \geq 1 - \beta.$$

Clearly, for all weighted instances of the PCSP, any spaciously robust rounding scheme is in particular robust, as  $\mathcal{V} \in B(\mathcal{V}, \delta)$  for any  $\delta$ . Note also that in general  $\text{comp}_{\mathcal{X}, \mathbf{A}}(\mathcal{W})$  is not defined (since, as noted above, it may be that  $\mathcal{W} \notin \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$ ), so we use the completeness of  $\mathcal{V}$  as a proxy.

#### 6.3.1 Smooth Lifting

Given an  $(\alpha, \beta)$ -robust scheme  $S$ , we would like a general method to turn it into a  $(\delta, \alpha, \beta)$ -spaciously robust scheme  $S'$ . To that end, we use a padding trick similar to one used by Raghavendra and Steurer [RS09].

**Definition 6.9.** *Given  $\mathbf{u} \in \mathbb{R}^M$ ,  $\mathbf{v} \in \mathbb{R}^N$ , and  $\rho \in [0, 1]$ , we let  $\mathbf{u} \oplus_\rho \mathbf{v} \in \mathbb{R}^{M+N}$  be the  $\rho$ -concatenation of  $\mathbf{u}$  and  $\mathbf{v}$ ; i.e., the vector defined by*

$$(\mathbf{u} \oplus_\rho \mathbf{v})_i := \begin{cases} \sqrt{\rho} u_i & i \in [M] \\ \sqrt{1 - \rho} v_i & i \in [M + N] \setminus [M]. \end{cases}$$

for each  $i \in [M + N]$ .

**Definition 6.10.** Given  $\mathbf{U} \in \mathbb{S}_A^M(\mathbf{u}_0)$ ,  $\mathbf{V} \in \mathbb{S}_A^N(\mathbf{v}_0)$ , and  $\rho \in [0, 1]$ , we define the  $\rho$ -concatenation of  $\mathbf{U}$  and  $\mathbf{V}$  as the list  $\mathbf{U} \oplus_\rho \mathbf{V} := (\mathbf{u}_a \oplus_\rho \mathbf{v}_a : a \in A)$ .

**Proposition 6.11.**  $\mathbf{U} \oplus_\rho \mathbf{V} \in \mathbb{S}_A^{M+N}(\mathbf{u}_0 \oplus_\rho \mathbf{v}_0)$ .

*Proof.* First, for any distinct  $a, a' \in A$ , we have that

$$\langle \mathbf{u}_a \oplus_\rho \mathbf{v}_a, \mathbf{u}_{a'} \oplus_\rho \mathbf{v}_{a'} \rangle = \rho \langle \mathbf{u}_a, \mathbf{u}_{a'} \rangle + (1 - \rho) \langle \mathbf{v}_a, \mathbf{v}_{a'} \rangle = 0.$$

Second,

$$\sum_{a \in A} \mathbf{u}_a \oplus_\rho \mathbf{v}_a = \left( \sum_{a \in A} \mathbf{u}_a \right) \oplus_\rho \left( \sum_{a \in A} \mathbf{v}_a \right) = \mathbf{u}_0 \oplus_\rho \mathbf{v}_0. \quad \square$$

Likewise, for  $\mathcal{U} \in (\mathbb{S}_A^M)^X$  and  $\mathcal{V} \in (\mathbb{S}_A^N)^X$ , we can define  $\mathcal{U} \oplus_\rho \mathcal{V} := (\mathbf{U}_x \oplus_\rho \mathbf{V}_x : x \in X)$ . Recall the Gram matrix  $\text{Mat}(\mathcal{V})$  of a global configuration  $\mathcal{V}$  defined in Section 6.2.1.

**Proposition 6.12.**  $\text{Mat}(\mathcal{U} \oplus_\rho \mathcal{V}) = \rho \text{Mat}(\mathcal{U}) + (1 - \rho) \text{Mat}(\mathcal{V})$ .

*Proof.* For any  $x, x' \in X$  and  $a, a' \in A$ , we have that

$$\begin{aligned} \text{Mat}(\mathcal{U} \oplus_\rho \mathcal{V})_{(x,a),(x',a')} &= \langle \mathbf{u}_{x,a} \oplus_\rho \mathbf{v}_{x,a}, \mathbf{u}_{x',a'} \oplus_\rho \mathbf{v}_{x',a'} \rangle \\ &= \rho \langle \mathbf{u}_{x,a}, \mathbf{u}_{x',a'} \rangle + (1 - \rho) \langle \mathbf{v}_{x,a}, \mathbf{v}_{x',a'} \rangle \\ &= \rho \text{Mat}(\mathcal{U})_{(x,a),(x',a')} + (1 - \rho) \text{Mat}(\mathcal{V})_{(x,a),(x',a')}, \end{aligned}$$

as desired.  $\square$

We now construct a specific global configuration  $\mathcal{U}$  that will be useful for our rounding scheme.

**Definition 6.13.** We say that  $\mathcal{U}_{X,A} \in (\mathbb{S}_A^M)^X$  is a uniform configuration if for all  $x, x' \in X$  and  $a, a' \in A$ , we have that

$$\langle \mathbf{u}_{x,a}, \mathbf{u}_{x',a'} \rangle = \begin{cases} \frac{1}{|A|} & x = x' \text{ and } a = a' \\ 0 & x = x' \text{ and } a \neq a' \\ \frac{1}{|A|^2} & \text{otherwise.} \end{cases}$$

**Proposition 6.14.** For  $M = |X| \cdot |A|$ ,  $\mathcal{U}_{X,A}$  is well-defined. Further, for any instance  $\mathcal{X}$  of  $\text{CSP}(\mathbf{A})$ , we have that  $\mathcal{U}_{X,A} \in \mathbb{V}_{\mathcal{X},\mathbf{A}}^M$ .

*Proof.* First, it is straightforward to verify that

$$\text{Mat}(\mathcal{U}_{X,A}) = \frac{1}{|A|^{|X|}} \sum_{\mathbf{a} \in A^X} \text{Mat}(\mathbf{a}).$$

Thus,  $\text{Mat}(\mathcal{U}_{X,A})$  is PSD and thus has a realization in  $\mathbb{R}^{|X| \cdot |A|}$ . Now, for any clause  $P$  of  $\mathcal{X}$  on variable set  $Y \subseteq X$ , we can also verify that

$$\text{Mat}(\mathcal{U}_{X,A})|_{(Y \times A)^2} = \frac{1}{|A|^{|Y|}} \sum_{\mathbf{a} \in A^Y} \text{Mat}(\mathbf{a}).$$

Thus, for each clause of  $\mathcal{X}$ ,  $\mathcal{U}_{X,A}$  is consistent with a uniformly random assignment. Thus,  $\mathcal{U}_{X,A} \in \mathbb{V}_{\mathcal{X},\mathbf{A}}^M$ .  $\square$

Given  $\mathcal{V} \in (\mathbb{S}_A^N)^X$  and  $\rho \in [0, 1]$ , we let  $\mathcal{U}_{X,A} \oplus_\rho \mathcal{V} \in (\mathbb{S}_A^{M+N})^X$  be the  $\rho$ -lift of  $\mathcal{V}$ . We next use  $\rho$ -lifting to make (29) more likely to be true.

### 6.3.2 Lifting $\delta$ -modifications

Our current goal is to prove the following lemma.

**Lemma 6.15.** *Let  $\mathcal{X}$  be a weighted instance of  $\text{PCSP}(\mathbf{A}, \mathbf{B})$ . For any  $\delta > 0$ , there exists  $\rho = O_{\mathbf{A}}(\sqrt{\delta}) \in [0, 1]$  such that for all  $\mathcal{V} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$  and  $\mathcal{W} \in B(\mathcal{V}, \delta)$ , we have that*

- (1)  $\mathcal{U}_{\mathcal{X}, \mathbf{A}} \oplus_{\rho} \mathcal{W} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^{|\mathcal{X}| \cdot |\mathbf{A}| + N}$ .
- (2)  $\text{comp}_{\mathcal{X}, \mathbf{A}}(\mathcal{U}_{\mathcal{X}, \mathbf{A}} \oplus_{\rho} \mathcal{W}) \geq \text{comp}_{\mathcal{X}, \mathbf{A}}(\mathcal{V}) - O_{\mathbf{A}}(\sqrt{\delta})$ .

To prove this lemma, we need some machinery from convex geometry. Some ideas are inspired by Raghavendra–Steurer [RS09] and Appendix A of [BGS23b]. We start with a simple proposition.

**Proposition 6.16.** *Let  $\mathbf{u}, \mathbf{u}', \mathbf{v}, \mathbf{v}' \in \mathbb{R}^N$  be vectors of length at most 1, such that  $\|\mathbf{u} - \mathbf{u}'\|_2^2 \leq \delta$  and  $\|\mathbf{v} - \mathbf{v}'\|_2^2 \leq \delta$ . Then  $|\langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}', \mathbf{v}' \rangle| \leq 2\sqrt{\delta}$ .*

*Proof.* Using Cauchy–Schwarz, notice that

$$\begin{aligned} |\langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}', \mathbf{v}' \rangle| &\leq |\langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{v}' \rangle| + |\langle \mathbf{u}, \mathbf{v}' \rangle - \langle \mathbf{u}', \mathbf{v}' \rangle| \\ &\leq \|\mathbf{u}\|_2 \|\mathbf{v} - \mathbf{v}'\|_2 + \|\mathbf{u} - \mathbf{u}'\|_2 \|\mathbf{v}'\|_2 \\ &\leq 2\sqrt{\delta}. \end{aligned} \quad \square$$

Next, we fix a finite (nonempty) set  $Y$ . We let  $K_{Y, \mathbf{A}}, K_{Y, \mathbf{A}}^0 \subseteq \mathbb{R}^{(Y \times \mathbf{A})^2}$  denote the following two convex sets, that are linear subspaces of  $\mathbb{R}^{(Y \times \mathbf{A})^2}$ :

$$\begin{aligned} K_{Y, \mathbf{A}} &:= \left\{ \sum_{\mathbf{a} \in \mathbf{A}^Y} \lambda_{\mathbf{a}} \text{Mat}(\mathbf{a}) : \lambda_{\mathbf{a}} \in \mathbb{R} \right\}, \\ K_{Y, \mathbf{A}}^0 &:= \left\{ \sum_{\mathbf{a} \in \mathbf{A}^Y} \lambda_{\mathbf{a}} \text{Mat}(\mathbf{a}) : \sum_{\mathbf{a} \in \mathbf{A}^Y} \lambda_{\mathbf{a}} = 0 \right\}. \end{aligned}$$

We will need the following key lemma. It can be directly derived from a combinatorial characterisation of low-dimensional projections of a hypermatrix proved in [CŽ23a]. Moreover, we also give a self-contained proof involving the dual space  $K_{Y, \mathbf{A}}^*$  of  $K_{Y, \mathbf{A}}$ . Full details of both arguments are postponed to Appendix A.

**Lemma 6.17.** *For any  $\mathcal{V}, \mathcal{W} \in (\mathbb{S}_{\mathbf{A}}^N)^Y$ , we have that*

- (1)  $\text{Mat}(\mathcal{V}) \in K_{Y, \mathbf{A}}$ .
- (2)  $\text{Mat}(\mathcal{V}) - \text{Mat}(\mathcal{W}) \in K_{Y, \mathbf{A}}^0$ .

We just need a bit more machinery, then we can prove Lemma 6.15. Define the set

$$L_{Y, \mathbf{A}} := \{M \in K_{Y, \mathbf{A}}^0 : \|M\|_{\infty} \leq 1\},$$

where  $\|M\|_\infty$  is the absolute value of the greatest entry of  $M$ . Note that  $L_{Y,A}$  is convex and compact. Define also the function  $f : K_{Y,A}^0 \rightarrow \mathbb{R}_{\geq 0}$  by

$$f(M) := \min_{\substack{\sum_{\mathbf{a}} \lambda_{\mathbf{a}} = 0 \\ \sum_{\mathbf{a}} \lambda_{\mathbf{a}} \text{Mat}(\mathbf{a}) = M}} \max_{\mathbf{a} \in A^Y} |\lambda_{\mathbf{a}}|.$$

We now prove the following.

**Proposition 6.18.**  $\kappa_{Y,A} := \sup_{M \in L_{Y,A}} f(M)$  is finite.

*Proof.* Pick an arbitrary basis  $M_1, \dots, M_r$  of  $K_{Y,A}^0$ . It is straightforward to see that  $f(M_i) < \infty$  for all  $i \in [r]$ . Furthermore, if  $M = \beta_1 M_1 + \dots + \beta_r M_r$ , then

$$f(M) \leq \sum_{i=1}^r |\beta_i| f(M_i). \quad (30)$$

Let  $g(M)$  be the RHS of (30). Since  $M_1, \dots, M_r$  form a basis, the map  $M \mapsto (\beta_1, \dots, \beta_r)$  is continuous. Thus,  $g(M)$  is continuous. Thus, since  $L$  is compact,  $\max_{M \in L_{Y,A}} g(M)$  exists and is finite. Thus,  $\kappa_{Y,A} := \sup_{M \in L_{Y,A}} f(M)$  is finite as well.  $\square$

We also need the following fact.

**Proposition 6.19.** We have that  $\mathcal{U}_{X,A} \oplus_\rho \mathcal{V} \in \mathbb{V}_{\mathcal{X},\mathbf{A}}^{|X|-|A|+N}$  and

$$\text{comp}_{\mathcal{X},\mathbf{A}}(\mathcal{U}_{X,A} \oplus_\rho \mathcal{V}) \geq \text{comp}_{\mathcal{X},\mathbf{A}}(\mathcal{V}) - O_{\mathbf{A}}(\rho). \quad (31)$$

*Proof.* Fix a clause  $(Y, P)$  of  $\mathcal{X}$  with  $Y \subseteq X$  and  $P \in \mathbf{A}$ . Pick a probability distribution  $(p_{\mathbf{a}} : \mathbf{a} \in A^Y)$  such that  $\text{Mat}(\mathcal{V}|_Y) = \sum_{\mathbf{a} \in A^Y} p_{\mathbf{a}} \text{Mat}(\mathbf{a})$  and  $\text{comp}_{Y,P}(\mathcal{V}|_Y) = \sum_{\mathbf{a} \in P} p_{\mathbf{a}}$ . By methods similar to that of Proposition 6.14, we can observe that

$$\text{Mat}((\mathcal{U}_{X,A} \oplus_\rho \mathcal{V})|_Y) = \sum_{\mathbf{a} \in A^Y} \left( \frac{\rho}{|A|^{|Y|}} + (1 - \rho)p_{\mathbf{a}} \right) \text{Mat}(\mathbf{a}). \quad (32)$$

Thus,  $\mathcal{U}_{X,A} \oplus_\rho \mathcal{V} \in \mathbb{V}_{\mathcal{X},\mathbf{A}}^{|X|-|A|+N}$  and

$$\begin{aligned} \text{comp}_{Y,P}((\mathcal{U}_{X,A} \oplus_\rho \mathcal{V})|_Y) &\geq \sum_{\mathbf{a} \in P} \left( \frac{\rho}{|A|^{|Y|}} + (1 - \rho)p_{\mathbf{a}} \right) \text{Mat}(\mathbf{a}) \\ &\geq (1 - \rho) \text{comp}_{Y,P}(\mathcal{V}|_Y) \\ &\geq \text{comp}_{Y,P}(\mathcal{V}|_Y) - O_{\mathbf{A}}(\rho), \end{aligned}$$

where we use that the completeness is at most 1. Taking the average of the above inequality for all clauses of  $\mathcal{X}$ , we get (31).  $\square$

*Proof of Lemma 6.15.* Keep  $\rho \in [0, 1]$  arbitrary for the moment. Fix a clause  $(Y, P, Q)$  of  $\mathcal{X}$ . Let  $\mathcal{V}' := (\mathcal{U}_{X,A} \oplus_\rho \mathcal{V})|_Y$  and  $\mathcal{W}' := (\mathcal{U}_{X,A} \oplus_\rho \mathcal{W})|_Y$ . To prove (1) it suffices to show that  $\mathcal{W}'$  has a solution to (29). Recall from Proposition 6.19 and (32) that  $\mathcal{V}' \in \mathbb{V}_{Y,P}^{|X|-|A|+N}$  and

$$\text{Mat}(\mathcal{V}') = \sum_{\mathbf{a} \in A^Y} \left( \frac{\rho}{|A|^{|Y|}} + (1 - \rho)p_{\mathbf{a}} \right) \text{Mat}(\mathbf{a}).$$

Since  $\mathcal{W} \in B(\mathcal{V}, \delta)$ , we have that  $\|(\mathcal{U}_{X,A} \oplus_\rho \mathcal{V}) - (\mathcal{U}_{X,A} \oplus_\rho \mathcal{W})\|_2^2 \leq (1 - \rho)\delta$ . By Proposition 6.16, we have that

$$\|\text{Mat}(\mathcal{V}') - \text{Mat}(\mathcal{W}')\|_\infty \leq 2\sqrt{(1 - \rho)\delta}.$$

That is,  $\frac{1}{2\sqrt{(1 - \rho)\delta}}(\text{Mat}(\mathcal{W}') - \text{Mat}(\mathcal{V}')) \in L_{Y,A}$ . Thus, by Lemma 6.17 and Proposition 6.18, there exists a choice of  $\lambda_{\mathbf{a}} \in \mathbb{R}$  for all  $\mathbf{a} \in A^Y$  such that  $\text{Mat}(\mathcal{W}') - \text{Mat}(\mathcal{V}') = \sum_{\mathbf{a} \in A^Y} \lambda_{\mathbf{a}} \text{Mat}(\mathbf{a})$  and  $|\lambda_{\mathbf{a}}| \leq 2\kappa_{Y,A}\sqrt{(1 - \rho)\delta}$ . Pick  $\rho$  minimal such that

$$2\kappa_{Y,A}\sqrt{(1 - \rho)\delta} \leq \frac{\rho}{|A|^{|Y|}}$$

for every clause  $(Y, P, Q)$  of  $\mathcal{X}$ . Note that  $\rho = O_{\mathbf{A}}(\sqrt{\delta})$  and  $\rho \in [0, 1]$  because the LHS is 0 when  $\rho = 1$ . As such, we now have that

$$\text{Mat}(\mathcal{W}') = \sum_{\mathbf{a} \in A^Y} \left( \frac{\rho}{|A|^{|Y|}} + \lambda_{\mathbf{a}} + (1 - \rho)p_{\mathbf{a}} \right) \text{Mat}(\mathbf{a}),$$

where every coefficient is nonzero. Thus,  $\mathcal{W}'$  satisfies (29), so  $\mathcal{U}_{X,A} \oplus_\rho \mathcal{W} \in \mathbb{V}_{\mathbf{X}, \mathbf{A}}^{|X| \cdot |A| + N}$ . This proves (1).

To prove (2), note that for a fixed clause  $(Y, P, Q)$  of  $\mathcal{X}$ , we have that

$$\begin{aligned} \text{comp}_{Y,P}(\mathcal{U}_{X,A} \oplus_\rho \mathcal{W}) &\geq \sum_{\mathbf{a} \in P} \left( \frac{\rho}{|A|^{|Y|}} + \lambda_{\mathbf{a}} + (1 - \rho)p_{\mathbf{a}} \right) \\ &\geq (1 - \rho) \sum_{\mathbf{a} \in P} p_{\mathbf{a}} \\ &= (1 - \rho) \text{comp}_{Y,P}(\mathcal{V}) \\ &\geq \text{comp}_{Y,P}(\mathcal{V}) - O_{\mathbf{A}}(\sqrt{\delta}). \end{aligned}$$

Averaging this inequality over all clauses  $(Y, P, Q)$  of  $\mathcal{X}$  proves (2), as desired.  $\square$

### 6.3.3 The Pullback Scheme

Recall by Corollary 6.7 that, for any set  $Z$ , we have that  $\text{BCR}_\theta^{|Z| \cdot |A| + N}$  is an  $(\alpha - \theta, \beta + \theta)$ -robust scheme for an instance  $\mathcal{X}$  of  $\text{PCSP}(\mathbf{A}, \mathbf{B})$ . We now define a *pullback scheme*  $\widehat{\text{BCR}}_{\theta, \rho}^{N, Z}$  for any  $\delta \in (0, 1)$  as follows.

**Definition 6.20** (Pullback Scheme). *For any set  $Z$ , we define  $\widehat{\text{BCR}}_{\theta, \delta}^{N, Z} : \mathbb{S}_{\mathbf{A}}^N \times \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}} \rightarrow \Delta_B$  to be*

$$\widehat{\text{BCR}}_{\theta, \delta}^{N, Z}(\mathbf{V}, \xi) = \mathbb{E}_{z \sim Z} \left[ \text{BCR}_\theta^{|Z| \cdot |A| + N}(\mathbf{U}_z \oplus_\rho \mathbf{V}, \xi) \right],$$

where  $z \sim Z$  is the uniform distribution,  $\mathbf{U}_z$  is the  $z$ -th local configuration of  $\mathcal{U}_{Z,A}$ , and  $\rho = O_{\mathbf{A}}(\sqrt{\delta})$  is chosen according to Lemma 6.15.

Note that we call this a ‘‘pullback’’ scheme as we are mapping a higher-dimensional rounding scheme to a lower dimension. We now show this pullback is spaciously robust.

**Lemma 6.21.** *If  $\text{BCR}_\theta^{|X| \cdot |A| + N}$  is an  $(\alpha - \theta, \beta + \theta)$ -robust scheme for an instance  $\mathcal{X}$  of  $\text{PCSP}(\mathbf{A}, \mathbf{B})$ , then  $\widehat{\text{BCR}}_{\theta, \delta}^{N, Z}$  is a  $(\delta, \alpha', \beta + \theta + \frac{|X|^2}{|Z|})$ -spaciously robust scheme for  $\mathcal{X}$ , where  $\alpha' = \alpha - \theta - O_{\mathbf{A}}(\sqrt{\delta})$ .*

*Proof.* Let  $\mathcal{V} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$  satisfy  $\text{comp}_{\mathcal{X}, \mathbf{A}}(\mathcal{V}) \geq 1 - \alpha'$ . Let  $\mathcal{W} \in B(\mathcal{V}, \delta)$ . By Definition 6.20 and the fact that  $\text{sound}_{\mathcal{X}, \mathbf{B}}$  is multilinear, we have that

$$\begin{aligned} & \mathbb{E}_{\xi \sim \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} [\text{sound}_{\mathcal{X}, \mathbf{B}}(\widehat{\text{BCR}}_{\theta, \delta}^{N, Z}(\mathbf{W}, \xi) : x \in X)] \\ &= \mathbb{E}_{\xi \sim \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}} \left( \mathbb{E}_{z \sim Z} \left[ \text{BCR}_{\theta}^{|Z| \cdot |A| + N}(\mathbf{U}_z \oplus_{\rho} \mathbf{W}, \xi) \right] : x \in X \right) \right] \\ &= \mathbb{E}_{\substack{\xi \sim \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}} \\ \mathbf{z} \sim Z^X}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}} \left( \text{BCR}_{\theta}^{|Z| \cdot |A| + N}(\mathbf{U}_{z_x} \oplus_{\rho} \mathbf{W}, \xi) : x \in X \right) \right], \end{aligned}$$

where  $\mathbf{z} \sim Z^X$  is chosen uniformly at random. For each  $\mathbf{z} \in Z^X$ , let  $\mathcal{W}_{\mathbf{z}} := (\mathbf{U}_{z_x} \oplus_{\rho} \mathbf{W} : x \in X)$ . If  $\mathbf{z}$  is injective—that is, if  $z_x = z_{x'}$  implies  $x = x'$ —then we may compute that

$$\text{Mat}(\mathcal{W}_{\mathbf{z}}) = \text{Mat}(\mathcal{U}_{X, A} \oplus_{\rho} \mathcal{W}).$$

Thus, by Lemma 6.15, for all injective  $\mathbf{z}$ , we have that  $\mathcal{W}_{\mathbf{z}} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^{|Z| \cdot |A| + N}$  and

$$\text{comp}_{\mathcal{X}, \mathbf{A}}(\mathcal{W}_{\mathbf{z}}) \geq \text{comp}_{\mathcal{X}, \mathbf{A}}(\mathcal{V}) - O_{\mathbf{A}}(\sqrt{\delta}) = 1 - \alpha + \theta.$$

Thus, since  $\text{BCR}_{\theta}^{|X| \cdot |A| + N}$  is  $(\alpha - \theta, \beta + \theta)$ -robust, we have that

$$\begin{aligned} 1 - \beta - \theta &\leq \mathbb{E}_{\substack{\xi \sim \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}} \\ \mathbf{z} \sim Z^X}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}} \left( \text{BCR}_{\theta}^{|Z| \cdot |A| + N}(\mathbf{U}_{z_x} \oplus_{\rho} \mathbf{W}, \xi) : x \in X \right) \middle| \mathbf{z} \text{ injective} \right] \\ &\leq \frac{1}{\Pr_{\mathbf{z} \sim Z^X}[\mathbf{z} \text{ injective}]} \mathbb{E}_{\substack{\xi \sim \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}} \\ \mathbf{z} \sim Z^X}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}} \left( \text{BCR}_{\theta}^{|Z| \cdot |A| + N}(\mathbf{U}_{z_x} \oplus_{\rho} \mathbf{W}, \xi) : x \in X \right) \right] \\ &= \frac{1}{\Pr_{\mathbf{z} \sim Z^X}[\mathbf{z} \text{ injective}]} \mathbb{E}_{\xi \sim \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} [\text{sound}_{\mathcal{X}, \mathbf{B}}(\widehat{\text{BCR}}_{\theta, \delta}^{N, Z}(\mathbf{W}, \xi) : x \in X)]. \end{aligned}$$

To finish, we note that

$$\Pr_{\mathbf{z} \sim Z^X}[\mathbf{z} \text{ injective}] = \prod_{i=1}^{|X|} \left( 1 - \frac{i-1}{|Z|} \right) \leq 1 - \frac{|X|^2}{|Z|},$$

so

$$\mathbb{E}_{\xi \sim \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} [\text{sound}_{\mathcal{X}, \mathbf{B}}(\widehat{\text{BCR}}_{\theta, \delta}^{N, Z}(\mathbf{W}, \xi) : x \in X)] \geq (1 - \beta - \theta) \cdot \left( 1 - \frac{|X|^2}{|Z|} \right) \geq 1 - \beta - \theta - \frac{|X|^2}{|Z|},$$

as desired.  $\square$

## 6.4 Measure Theory Fundamentals

Our next step toward the proof of Theorem 6.1 is to modify  $\widehat{\text{BCR}}$  to another scheme (which we will call  $\text{REQ}$ ) so that similar inputs give similar outputs. In order to do this, we first need some fundamentals of measure theory. We refer to the textbook of DiBenedetto [DiB16] for foundational concepts.

Let  $R$  be a possibly infinite set, and let  $\mathcal{R} := (R, \mu)$  be a probability measure over  $R$ . See Chapter 3 of [DiB16] for a precise definition. Recall that a function  $f : R \rightarrow \mathbb{R}^d$  is *measurable* if the preimage under  $f$  of every measurable subset of  $\mathbb{R}^d$  is measurable. Let also  $\int_R$  denote the Lebesgue integral over  $R$ . We shall consider the set  $L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  consisting of the quotient space of the set of all measurable functions  $f : R \rightarrow \mathbb{R}^d$  such that  $\int_R \|f(x)\|_2^2 d\mu(x) < \infty$ , modulo the equivalence relation “ $\sim$ ” defined by  $f \sim g$  if and only if  $f - g$  is only nonzero on a set of zero measure. Observe that  $L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  is a normed space, with the norm

$$\|f\|_{2,2} = \sqrt{\int_R \|f(x)\|_2^2 d\mu(x)}. \quad (33)$$

We say that  $f \in L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  is a limit of a sequence of functions  $f_1, f_2, \dots \in L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  if, for all  $\epsilon > 0$ , there exists  $N \in \mathbb{N}$  such that  $\|f - f_i\|_{2,2} \leq \epsilon$  for all  $i \geq N$ . It is clear that every sequence has at most one limit. A *Cauchy sequence* is a sequence of functions  $f_1, f_2, \dots \in L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  such that, for all  $\epsilon > 0$ , there exists  $N \in \mathbb{N}$  such that  $\|f_i - f_j\|_{2,2} \leq \epsilon$  for any  $i, j \geq N$ . We say that  $L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  is a *Banach space* if every Cauchy sequence has a limit. The following is a simple consequence of the Riesz–Fischer theorem.

**Theorem 6.22** (Riesz–Fischer, see Chapter 6, Theorem 5.1 of [DiB16]).  $L^{2,2}(\mathcal{R}, \mathbb{R}) = L^2(\mathcal{R})$  is a Banach space.<sup>17</sup>

**Corollary 6.23.** For all  $d \geq 1$ ,  $L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  is a Banach space.

*Proof.* Consider any Cauchy sequence  $f_1, f_2, \dots \in L^{2,2}(\mathcal{R}, \mathbb{R}^d)$ . For each  $i \in [d]$ , note that  $f_{1,i}, f_{2,i}, \dots$  is a Cauchy sequence in  $L^{2,2}(\mathcal{R}, \mathbb{R})$ . Thus, by Theorem 6.22, this sequence converges to  $g_i \in L^{2,2}(\mathcal{R}, \mathbb{R})$ . Let  $g = (g_1, \dots, g_d)$ . It is easy to check that  $g \in L^{2,2}(\mathcal{R}, \mathbb{R}^d)$ . Furthermore, since  $d$  is finite, we have that

$$\begin{aligned} 0 &= \sqrt{\sum_{j=1}^d 0} = \sqrt{\sum_{j=1}^d \lim_{i \rightarrow \infty} \|f_{i,j} - g_j\|_{2,2}^2} \\ &= \lim_{i \rightarrow \infty} \sqrt{\sum_{j=1}^d \|f_{i,j} - g_j\|_{2,2}^2} \\ &= \lim_{i \rightarrow \infty} \|f_i - g\|_{2,2}. \end{aligned}$$

Thus,  $g$  is indeed the limit of the Cauchy sequence  $f_1, f_2, \dots$ , as desired.  $\square$

We say that  $K \subseteq L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  is *convex* if, for any  $f_1, f_2 \in K$  and  $\lambda \in [0, 1]$ , we have that  $\lambda f_1 + (1 - \lambda)f_2 \in K$ . We let  $L^{2,2}(\mathcal{R}, \Delta_d) \subseteq L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  be the subset of functions whose range lies in  $\Delta_d$ . Note that  $L^{2,2}(\mathcal{R}, \Delta_d)$  is convex.

Given  $K \subseteq L^{2,2}(\mathcal{R}, \mathbb{R}^d)$ , we let  $\overline{K} \subseteq L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  denote its (sequential) *closure*. That is,  $f \in \overline{K}$  if and only if, for all  $\epsilon > 0$ , there exists  $g \in K$  with  $\|f - g\|_{2,2} \leq \epsilon$ . Clearly if  $K$  is convex then so is  $\overline{K}$ . The following fact is well known. To get some familiarity with the definitions, we include a simple proof.

<sup>17</sup>Technically, the treatment in [DiB16] allows for  $f$  to take on the value  $\pm\infty$ , but it is clear from (33) that  $f^{-1}(\{-\infty, \infty\})$  must have zero measure in order for  $\|f\|_{2,2} < \infty$ . Thus, we can assume  $f$  has finite output.

**Proposition 6.24.** *If  $K \subseteq L^{2,2}(\mathcal{R}, \Delta_d)$ , then  $\overline{K} \subseteq L^{2,2}(\mathcal{R}, \Delta_d)$ .*

*Proof.* Assume for sake of contradiction that there is  $f \in \overline{K}$  with  $f \notin L^{2,2}(\mathcal{R}, \Delta_d)$ . Let  $S = f^{-1}(\mathbb{R}^d \setminus \Delta_d)$ . Note that  $\mu(S) > 0$ . Since  $f$  is measurable, there must exist  $\delta > 0$  and  $T \subseteq S$  such that  $\mu(T) > 0$  and, for all  $x \in T$ , the  $\ell_1$  distance between  $x$  and  $\Delta_d$  is greater than  $d$ . Therefore, for all  $g \in L^{2,2}(\mathcal{R}, \Delta_d)$ , we have that

$$\|f - g\|_{2,2}^2 \geq \int_T \|f(x) - g(x)\|_1^2 d\mu(x) \geq \delta^2 \mu(T).$$

Since  $\delta^2 \mu(T) > 0$ , this contradicts the fact that  $f \in \overline{K}$ .  $\square$

We next observe an elementary fact about our norm for  $L^{2,2}(\mathcal{R}, \mathbb{R}^d)$ .

**Proposition 6.25.** *If  $f, g \in L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  satisfy  $\|f - g\|_{2,2} \geq \delta$  then*

$$\|(f + g)/2\|_{2,2}^2 \leq \max(\|f\|_{2,2}^2, \|g\|_{2,2}^2) - \delta^2/4.$$

*Proof.* Note that

$$\begin{aligned} \|(f + g)/2\|_{2,2}^2 &= \frac{1}{4} \int_{\mathcal{R}} \|f(x) + g(x)\|_2^2 d\mu(x) \\ &= \frac{1}{4} \int_{\mathcal{R}} 2\|f(x)\|_2^2 + 2\|g(x)\|_2^2 - \|f(x) - g(x)\|_2^2 d\mu(x) \\ &= \frac{1}{4} (2\|f\|_{2,2}^2 + 2\|g\|_{2,2}^2 - \|f - g\|_{2,2}^2) \\ &\leq \max(\|f\|_{2,2}^2, \|g\|_{2,2}^2) - \delta^2/4. \end{aligned} \quad \square$$

We now show that we can uniquely solve a suitable optimization problem over  $L^{2,2}(\mathcal{R}, \Delta_d)$ .

**Lemma 6.26.** *Let  $K \subseteq L^{2,2}(\mathcal{R}, \mathbb{R}^d)$  be nonempty and convex, then there exists a unique  $f \in \overline{K}$  for which  $\|f\|_{2,2}$  is minimized.*

*Proof.* First, note that if a minimizer exists it must be unique. Otherwise, there are  $f, g \in \overline{K}$  for which  $\|f\|_{2,2} = \|g\|_{2,2}$  is the minimum, but Proposition 6.25 shows that  $(f + g)/2 \in \overline{K}$  must have a smaller norm for  $\|f - g\|_{2,2} > 0$  by definition of being distinct (recall that for  $f$  and  $g$  to be distinct we require that  $f \neq g$  on a set of nonzero measure).

Thus, it suffices to prove that a minimizer exists. Let  $\alpha = \inf\{\|f\|_{2,2}^2 : f \in \overline{K}\} < \infty$ . By definition of the infimum, for all  $n \in \mathbb{N}$ , there exists  $f_n \in \overline{K}$  for which  $\|f_n\|_{2,2}^2 \leq \alpha + \frac{1}{n}$ . Now consider any positive integers  $n \geq m$  and observe by Proposition 6.25 that

$$\alpha \leq \|(f_n + f_m)/2\|_{2,2}^2 \leq \max(\|f_n\|_{2,2}^2, \|f_m\|_{2,2}^2) - \frac{1}{4}\|f_n - f_m\|_{2,2}^2 \leq \alpha + \frac{1}{m} - \frac{1}{4}\|f_n - f_m\|_{2,2}^2.$$

Thus,  $\|f_n - f_m\|_{2,2}^2 \leq 4/m$ . Hence  $f_1, f_2, \dots$  is a Cauchy sequence. By Theorem 6.22, we have that its limit  $f$  lives in  $\overline{K}$ . Furthermore, since  $\|f_m\|_{2,2}^2$  approaches  $\alpha$  as  $m \rightarrow \infty$ , it is straightforward to show that  $\|f\|_{2,2}^2 = \alpha$ . Thus,  $f$  is the unique minimizer of the norm in  $\overline{K}$ .  $\square$

## 6.5 Defining REQ

Recall we previously defined a *rounding scheme* to be a function  $S : \mathbb{S}_A^N \times \Xi \rightarrow \Delta_B$  (Definition 6.5). Equivalently, we can think of  $S$  as having signature  $\mathbb{S}_A^N \rightarrow L^{2,2}(\Xi, \Delta_B)$ . We assume this latter presentation from now on.

**Definition 6.27** ( $\delta$ -spread). *Recall that, for any  $\mathbf{V} \in \mathbb{S}_A^N$ ,  $B(\mathbf{V}, \delta)$  denotes the set of all  $\mathbf{W} \in \mathbb{S}_A^N$  with  $\|\mathbf{V} - \mathbf{W}\|_2^2 \leq \delta$ . We define a  $\delta$ -spread of  $\mathbf{V}$  to be a finitely supported probability distribution  $\Lambda$  over  $B(\mathbf{V}, \delta)$  such that for all  $p \in [0, 1]$ , we have that*

$$\Pr_{\mathbf{W} \sim \Lambda} [\|\mathbf{V} - \mathbf{W}\|_2^2 \leq p^2 \delta] \geq p.$$

We say that  $S' : \mathbb{S}_A^N \rightarrow L^{2,2}(\Xi, \Delta_B)$  is a  $\delta$ -smoothing of  $S : \mathbb{S}_A^N \rightarrow L^{2,2}(\Xi, \Delta_B)$  if, for all  $\mathbf{V} \in \mathbb{S}_A^N$ , there exists a  $\delta$ -spread  $\Lambda_{\mathbf{V}}$  of  $\mathbf{V}$  such that

$$S'(\mathbf{V}) = \mathbb{E}_{\mathbf{W} \sim \Lambda_{\mathbf{V}}} [S(\mathbf{W})].$$

We let  $B(S, \mathbf{V}, \delta) \subseteq L^{2,2}(\Xi, \Delta_B)$  denote all possible values of  $S'(\mathbf{V})$ .

**Proposition 6.28.**  *$B(S, \mathbf{V}, \delta)$  is convex.*

*Proof.* Consider any  $S', S'' \in B(S, \mathbf{V}, \delta)$  and let  $\Lambda', \Lambda''$  be  $\delta$ -spreads of  $\mathbf{V}$  such that  $S'(\mathbf{V}) = \mathbb{E}_{\mathbf{W} \sim \Lambda'} [S(\mathbf{W})]$  and  $S''(\mathbf{V}) = \mathbb{E}_{\mathbf{W} \sim \Lambda''} [S(\mathbf{W})]$ . For any  $\theta \in [0, 1]$ , let  $\Lambda_{\theta}$  be the probability distribution that samples from  $\Lambda'$  with probability  $\theta$  and from  $\Lambda''$  with probability  $1 - \theta$ . It is not hard to see that

$$(\theta S' + (1 - \theta) S'')(\mathbf{V}) = \mathbb{E}_{\mathbf{W} \sim \Lambda_{\theta}} [S(\mathbf{W})].$$

It remains to verify that  $\Lambda_{\theta}$  is a  $\delta$ -spread of  $\mathbf{V}$ . To see why, note that for any  $p \in [0, 1]$ , we have that

$$\begin{aligned} \Pr_{\mathbf{W} \sim \Lambda_{\theta}} [\|\mathbf{V} - \mathbf{W}\|_2^2 \leq p^2 \delta] &= \theta \Pr_{\mathbf{W} \sim \Lambda'} [\|\mathbf{V} - \mathbf{W}\|_2^2 \leq p^2 \delta] + (1 - \theta) \Pr_{\mathbf{W} \sim \Lambda''} [\|\mathbf{V} - \mathbf{W}\|_2^2 \leq p^2 \delta] \\ &\geq \theta p + (1 - \theta)p = p \end{aligned}$$

as desired. □

We now define  $\text{REQ}_{\theta, \delta}^{N, Z} : \mathbb{S}_A^N \rightarrow L^{2,2}(\Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}, \Delta_B)$  in terms of  $\widehat{\text{BCR}}_{\theta, \delta}^{N, Z} : \mathbb{S}_A^N \rightarrow L^{2,2}(\Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}, \Delta_B)$  as follows.

**Definition 6.29.** *For all  $\mathbf{V} \in \mathbb{S}_A^N$  and sets  $Z$ , we let*

$$\text{REQ}_{\theta, \delta}^{N, Z}(\mathbf{V}) := \overline{\text{argmin}\{\|f\|_{2,2} : f \in B(\widehat{\text{BCR}}_{\theta, \delta}^{N, Z}, \mathbf{V}, \delta)\}}.$$

Note this definition is well posed by Lemma 6.26. The analysis of  $\text{REQ}_{\theta, \delta}^{N, Z}$  splits into two parts, which shall be proved in the next two subsections.

**Lemma 6.30.** *Let  $\mathcal{X}$  be a weighted instance of PCSP( $\mathbf{A}, \mathbf{B}$ ). Assume  $\text{BCR}_{\theta}^{|Z| \cdot |A| + N}$  is  $(\alpha - \theta, \beta + \theta)$ -robust for  $\mathcal{X}$ . Then  $\text{REQ}_{\theta, \delta}^{N, Z}$  is  $(\alpha - \theta - O_{\mathbf{A}}(\sqrt{\delta}), \beta + \theta + \frac{|X|^2}{|Z|})$ -robust for  $\mathcal{X}$ , where  $O_{\mathbf{A}}(\sqrt{\delta})$  is from Lemma 6.21.*

**Lemma 6.31.** *Assume  $\mathbf{V}, \mathbf{W} \in \mathbb{S}_A^N$  with  $\|\mathbf{V} - \mathbf{W}\|_2^2 \leq \delta \cdot \eta^2$  for some  $\eta > 0$ . Then,  $\|\text{REQ}_{\theta, \delta}^{N, Z}(\mathbf{V}) - \text{REQ}_{\theta, \delta}^{N, Z}(\mathbf{W})\|_{2,2} \leq 8\sqrt{\eta}$ .*

### 6.5.1 Robustness is Preserved: Lemma 6.30

*Proof of Lemma 6.30.* By definition of  $\text{REQ}_{\theta,\delta}^{N,Z}$  and Lemma 6.26, for all  $\epsilon > 0$ , there exists a  $\delta$ -smoothing  $S$  of  $\widehat{\text{BCR}}_{\theta,\delta}^{N,Z}$  such that for all  $\mathbf{V} \in \mathbb{S}_A^N$ , we have that

$$\|\text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V}) - S(\mathbf{V})\|_{2,2} = \sqrt{\mathbb{E}_{\xi \in \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} [(\text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V}, \xi) - S(\mathbf{V}, \xi))^2]} < \epsilon.$$

Let  $\alpha' = \alpha - \theta - O_{\mathbf{A}}(\sqrt{\delta})$ . Fix  $\mathcal{V} \in \mathbb{V}_{\mathcal{X}, \mathbf{A}}^N$  for which  $\text{comp}_{\mathcal{X}, \mathbf{A}}(\mathcal{V}) \geq 1 - \alpha'$ . For each  $x \in X$ , let  $\Lambda_x$  be a probability distribution over  $B(\mathbf{V}_x, \delta)$  such that

$$S(\mathbf{V}) = \mathbb{E}_{\mathbf{W} \sim \Lambda_x} \left[ \widehat{\text{BCR}}_{\theta,\delta}^{N,Z}(\mathbf{W}) \right]. \quad (34)$$

Let  $\mathbb{W}$  be a finite probability distribution over  $B(\mathcal{V}, \delta)$ , where for each  $x \in X$ ,  $\mathbf{W}_x$  is an independent sample from  $\Lambda_x$ . Note that every  $\mathcal{W} \in \text{supp } \mathbb{W}$  is a  $\delta$ -modification of  $\mathcal{V}$ , thus by Lemma 6.21 we have that

$$\mathbb{E}_{\xi \in \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}, \mathbf{W} \sim \mathbb{W}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}}(\widehat{\text{BCR}}_{\theta,\delta}^{N,Z}(\mathbf{W}, \xi) : x \in X) \right] \geq 1 - \beta - \theta - \frac{|X|}{|Z|^2}.$$

Like in the proof of Lemma 6.21, we use the fact that  $\text{sound}_{\mathcal{X}, \mathbf{B}} : \Delta_B^X \rightarrow [0, 1]$  is a multilinear function. Thus, since  $\mathbb{W}$  is a product distribution, we have that

$$\begin{aligned} 1 - \beta - \theta - \frac{|X|^2}{|Z|} &\leq \mathbb{E}_{\xi \in \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}}(\mathbb{E}_{\mathbf{W}_x \sim \Lambda_x} [\widehat{\text{BCR}}_{\theta,\delta}^{N,Z}(\mathbf{W}_x, \xi)] : x \in X) \right] \\ &= \mathbb{E}_{\xi \in \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}}(S(\mathbf{W}_x, \xi) : x \in X) \right], \end{aligned}$$

where the last equality uses (34). Since  $\text{sound}_{\mathcal{X}, \mathbf{B}} : \Delta_B^X \rightarrow [0, 1]$  is bounded and multilinear, we have for any  $\mathbf{p}, \mathbf{q} \in \Delta_B^X$  the coarse bound

$$\text{sound}_{\mathcal{X}, \mathbf{B}}(\mathbf{p}) - \text{sound}_{\mathcal{X}, \mathbf{B}}(\mathbf{q}) \leq \sum_{x \in X} |p_x - q_x| \leq \sqrt{|X| \sum_{x \in X} (p_x - q_x)^2},$$

where the latter inequality follows from Cauchy–Schwarz. Thus, since  $\sqrt{\cdot}$  is a concave function,

$$\begin{aligned} &\mathbb{E}_{\xi \in \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}}(S(\mathbf{V}_x, \xi) : x \in X) - \text{sound}_{\mathcal{X}, \mathbf{B}}(\text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V}_x, \xi) : x \in X) \right] \\ &\leq \sqrt{|X|} \mathbb{E}_{\xi \in \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} \left[ \sqrt{\sum_{x \in X} (S(\mathbf{V}_x, \xi) - \text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V}_x, \xi))^2} \right] \\ &\leq \sqrt{|X| \cdot \sum_{x \in X} \mathbb{E}_{\xi \in \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} [(S(\mathbf{V}, \xi) - \text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V}, \xi))^2]} \\ &< \sqrt{|X| \cdot \sum_{x \in X} \epsilon^2} = |X| \epsilon. \end{aligned}$$

Therefore,

$$\mathbb{E}_{\xi \in \Xi_{|Z| \cdot |A| + N, \theta}^{\text{BCR}}} \left[ \text{sound}_{\mathcal{X}, \mathbf{B}}(\text{REQ}_{\theta,\delta,x}^N(\mathbf{V}_x, \xi) : x \in X) \right] \geq 1 - \beta - \theta - |X| \epsilon.$$

Taking the limit as  $\epsilon \rightarrow 0$  proves that  $\text{REQ}_{\theta,\delta}^N$  is  $(\alpha - \theta - O_{\mathbf{A}}(\sqrt{\delta}), \beta + \theta)$ -robust for  $\mathcal{X}$ .  $\square$

### 6.5.2 Smoothness of $\text{REQ}_{\theta,\delta}^N$ : Lemma 6.31

We now prove Lemma 6.31. We first argue the following.

**Proposition 6.32.** *Let  $S : \mathbb{S}_A^N \times \Xi \rightarrow \Delta_B$  be a rounding scheme. Consider  $\mathbf{V}, \mathbf{W} \in \mathbb{S}_A^N$  with  $\|\mathbf{V} - \mathbf{W}\|_2^2 \leq \delta \cdot \eta^2$  for  $\eta \in (0, 1)$ . For any  $f \in B(S, \mathbf{V}, \delta)$ , there is  $g \in B(S, \mathbf{W}, \delta)$  with  $\|f - g\|_{2,2} \leq 2\eta$ .*

*Proof.* Let  $\Lambda$  be a  $\delta$ -spread of  $\mathbf{V}$  such that

$$\begin{aligned} \forall p \in [0, 1], p &\leq \Pr_{\mathbf{V}' \sim \Lambda} [\|\mathbf{V} - \mathbf{V}'\|_2^2 \leq p^2 \delta] \\ f &= \mathbb{E}_{\mathbf{V}' \sim \Lambda} [S(\mathbf{V}')]. \end{aligned}$$

Recall that  $\Lambda$  is a discrete probability distribution. Let  $\mathbf{V}'_1, \dots, \mathbf{V}'_k \in B(\mathbf{V}, \delta)$  be the support of this distribution with  $p_1, \dots, p_k \in [0, 1]$  the corresponding probabilities. We assume that  $\|\mathbf{V} - \mathbf{V}'_1\|_2^2 \leq \dots \leq \|\mathbf{V} - \mathbf{V}'_k\|_2^2$ . Pick  $\ell \in [k]$  minimal such that  $p_1 + \dots + p_\ell \geq 1 - \eta$ . Let  $\Lambda'$  be a probability distribution supported on  $\{\mathbf{W}, \mathbf{V}'_1, \dots, \mathbf{V}'_\ell\}$  which samples  $\mathbf{W}$  with probability  $\eta$ ,  $\mathbf{V}'_i$  for  $i \in [\ell - 1]$  with probability  $p_i$ , and samples  $\mathbf{V}'_\ell$  otherwise (with probability  $1 - \eta - p_1 - \dots - p_{\ell-1}$ ). Informally,  $\Lambda'$  samples  $\mathbf{W}$  with probability  $\eta$  and otherwise samples randomly from the  $1 - \eta$  fraction of  $\Lambda$  that is closest to  $\mathbf{V}$ .

We next justify why  $\Lambda'$  is a  $\delta$ -spread of  $\mathbf{W}$ . First note that if  $\|\mathbf{V} - \mathbf{V}'\|_2^2 \leq p^2 \delta$  then

$$\|\mathbf{W} - \mathbf{V}'\|_2^2 \leq (\|\mathbf{W} - \mathbf{V}\|_2 + \|\mathbf{V} - \mathbf{V}'\|_2)^2 \leq (\eta\sqrt{\delta} + p\sqrt{\delta})^2 = (p + \eta)^2 \delta.$$

As such, for any  $p \in [0, 1]$ , we have that

$$\Pr_{\mathbf{V}' \sim \Lambda} [\|\mathbf{W} - \mathbf{V}'\|_2^2 \leq p^2 \delta] \geq \Pr_{\mathbf{V}' \sim \Lambda} [\|\mathbf{V} - \mathbf{V}'\|_2^2 \leq \max(p - \eta, 0)^2 \delta] \geq \max(p - \eta, 0).$$

Therefore,

$$\Pr_{\mathbf{V}' \sim \Lambda'} [\|\mathbf{W} - \mathbf{V}'\|_2^2 \leq p^2 \delta] \geq \max(p - \eta, 0) + \eta \geq p.$$

Thus,  $\Lambda'$  is a  $\delta$ -spread of  $\mathbf{W}$  and thus supported in  $B(\mathbf{W}, \delta)$ . Let  $g = \mathbb{E}_{\mathbf{V}' \sim \Lambda'} [S(\mathbf{V}')]$ . We have that

$$\begin{aligned} \|f - g\|_{2,2} &= \left\| \mathbb{E}_{\mathbf{V}' \sim \Lambda} [S(\mathbf{V}')] - \mathbb{E}_{\mathbf{V}' \sim \Lambda'} [S(\mathbf{V}')] \right\|_{2,2} \\ &= \left\| \sum_{i \in [k]} p_i S(\mathbf{V}'_i) - \eta S(\mathbf{W}) - \sum_{i \in [\ell-1]} p_i S(\mathbf{V}'_i) - (1 - \eta - p_1 - \dots - p_{\ell-1}) S(\mathbf{V}'_\ell) \right\|_{2,2} \\ &\leq 2\eta \max_{\mathbf{V}' \in B(\mathbf{V}, \delta) \cup B(\mathbf{W}, \delta)} \|S(\mathbf{V}')\|_{2,2} \\ &\leq 2\eta, \end{aligned}$$

as desired.  $\square$

*Proof of Lemma 6.31.* If  $\eta \geq 1$ , then  $\|\text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V}) - \text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{W})\|_{2,2} \leq \sqrt{2} < 8\sqrt{\eta}$ , as desired. Thus, assume  $\eta \in (0, 1)$ .

Fix  $x \in X$  and let  $f = \text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V})$  and  $g = \text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{W})$ . By Proposition 6.32, we have that there is  $f' \in B(\widehat{\text{BCR}}_{\theta,\delta}^{N,Z}, \mathbf{V}, \delta)$  and  $g' \in B(\widehat{\text{BCR}}_{\theta,\delta}^{N,Z}, \mathbf{W}, \delta)$  with  $\|f' - g\|_{2,2} \leq 2\eta$  and  $\|g' - f\|_{2,2} \leq 2\eta$ .

Let  $g'' = (g + g')/2$  and  $f'' = (f + f')/2$ . Assume WLOG that  $\|f\|_{2,2} \geq \|g\|_{2,2}$ . By Proposition 6.25 and the fact that  $\|f\|_{2,2} \leq 1$ , we have that

$$\begin{aligned} \|f''\|_{2,2}^2 &\leq \max(\|f\|_{2,2}^2, \|f'\|_{2,2}^2) - \frac{\|f - f'\|_{2,2}^2}{4} \\ &\leq \max(\|f\|_{2,2}^2, (\|g\|_{2,2} + 2\eta)^2) - \frac{(\|f - g\|_{2,2} - 2\eta)^2}{4} \\ &\leq (\|f\|_{2,2} + 2\eta)^2 - \frac{(\|f - g\|_{2,2} - 2\eta)^2}{4} \\ &\leq \|f\|_{2,2}^2 + 8\eta - \frac{(\|f - g\|_{2,2} - 2\eta)^2}{4}. \end{aligned}$$

Since  $f'' \in B(\widehat{\text{BCR}}_{\theta,\delta}^{N,Z}, \mathbf{V}, \delta)$  and  $f = \text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V})$ , we have that  $\|f''\|_{2,2}^2 \geq \|f\|_{2,2}^2$ . Thus,

$$8\eta \geq \frac{(\|f - g\|_{2,2} - 2\eta)^2}{4}.$$

Thus,  $\|f - g\|_{2,2} \leq \sqrt{32\eta} + 2\eta \leq 8\sqrt{\eta}$ , as desired.  $\square$

## 6.6 Correlated Rounding

The final necessary ingredient to prove Theorem 6.1 is to convert the *independent* rounding of BCR into a *correlated* rounding. Let  $\text{plur} : \mathbb{R}^B \rightarrow B$  be any function with the property that for all  $q \in \Delta_B$ ,  $q(\text{plur}(q)) \geq q(b)$  for all  $b \in B$ . That is,  $\text{plur}$  is a plurality function with ties broken arbitrarily (cf. Section 5).

**Definition 6.33.** Let  $S : \mathbb{S}_A^N \rightarrow L^{2,2}(\Xi, \Delta_B)$  be a rounding scheme. Let  $\Xi_{\text{COR},B}$  be the uniform distribution over  $[0, \frac{1}{2|B|}]^B$ . We then define  $S_{\text{COR}} : \mathbb{S}_A^N \rightarrow L^{2,2}(\Xi \times \Xi_{\text{COR},B}, \Delta_B)$  as follows:

$$S_{\text{COR}}(\mathbf{V}, \xi, \xi_{\text{COR}}) = \text{plur}(S(\mathbf{V}, \xi) + \xi_{\text{COR}}). \quad (35)$$

Note that  $S_{\text{COR}}$  only outputs elements of  $B$ , which we identify with the vertices of  $\Delta_B$ .

We seek to prove two facts about (35): First, if  $S$  is robust, then  $S_{\text{COR}}$  is robust; second, if  $\|S(\mathbf{V}) - S(\mathbf{W})\|_{2,2}$  is small, then  $S_{\text{COR}}(\mathbf{V})$  and  $S_{\text{COR}}(\mathbf{W})$  have mostly equal outputs.

**Lemma 6.34.** Assume  $S : \mathbb{S}_A^N \times \Xi \rightarrow \Delta_B$  is a rounding scheme that is  $(\alpha, \beta)$ -robust for a weighted instance  $\mathcal{X}$  of PCSP $(\mathbf{A}, \mathbf{B})$ . Then,  $S_{\text{COR}}$  is  $(\alpha, O_{\mathbf{B}}(\beta))$ -robust for  $\mathcal{X}$ .

*Proof.* Let  $\mathcal{V} \in \mathbb{V}_{\mathcal{X},\mathbf{A}}^N$  be an SDP solution with  $\text{comp}_{\mathcal{X},\mathbf{A}}(\mathcal{V}) \geq 1 - \alpha$ . By Definition 6.6, we have that

$$\mathbb{E}_{\xi \sim \Xi} [\text{sound}_{\mathcal{X},\mathbf{B}}(S(\mathbf{V}_x, \xi) : x \in X)] \geq 1 - \beta. \quad (36)$$

Let  $(Y, P, Q)$  be an arbitrary clause of  $\mathcal{X}$  and fix  $\xi \in \text{supp } \Xi$ . Assume that  $\text{sound}_{Y,Q}(S(\mathbf{V}_y, \xi) : y \in Y) \geq 1 - \gamma$ . We claim that

$$\mathbb{E}_{\mathbf{p} \sim [0,1/(2|B|)]^B} [\text{sound}_{Y,Q}(S_{\text{COR}}(\mathbf{V}_y, \xi, \mathbf{p}) : y \in Y)] \geq 1 - (2|B|)^{|Y|}\gamma. \quad (37)$$

For notational convenience, we let  $\mathbf{r} := (S(\mathbf{V}_y, \xi) : y \in Y) \in \Delta_B^Y$ . The assumption  $\text{sound}_{Y,Q}(\mathbf{q}) \geq 1 - \gamma$  is equivalent to

$$\sum_{\mathbf{q} \in B^Y \setminus Q} \prod_{y \in Y} r_{y, q_y} \leq \gamma. \quad (38)$$

and we seek to prove that

$$\Pr_{\mathbf{p} \sim [0, 1/(2|B|)]^B} [(\text{plur}(\mathbf{r}_y + \mathbf{p}) : y \in Y) \in Q] \geq 1 - (2|B|)^{|Y|} \gamma.$$

Consider any  $\mathbf{q} \in B^Y \setminus Q$  for which there exists  $\mathbf{p} \in [0, 1/(2|B|)]^B$  for which  $(\text{plur}(\mathbf{r}_y + \mathbf{p}) : y \in Y) = \mathbf{q}$ . If no such  $\mathbf{q}$  exists, then the LHS of (36) equals 1.

For each  $y \in Y$ , in order for  $\text{plur}(\mathbf{r}_y + \mathbf{p}) = q_y$ , we must have that  $r_{y, q_y} + \frac{1}{2|B|} \geq r_{y, b}$  for all  $b \in B$ . Since  $\sum_{b \in B} r_{y, b} \geq 1$ , this means that  $r_{y, q_y} \geq \frac{1}{2|B|}$ . Thus, by (38), we have that

$$\gamma \geq \prod_{y \in Y} r_{y, q_y} \geq \frac{1}{(2|B|)^{|Y|}}.$$

Therefore, the RHS of (37) is at most zero, so the inequality is true. To finish, we apply (37) and (36) as follows.

$$\begin{aligned} & \mathbb{E}_{\xi \sim \Xi} [\text{sound}_{\mathcal{X}, \mathbf{B}}(S_{\text{COR}}(\mathbf{V}_x, \xi) : x \in X)] \\ &= \mathbb{E}_{\substack{\xi \sim \Xi \\ (Y, P, Q) \sim \mathcal{X}}} [\text{sound}_{Y, Q}(S_{\text{COR}}(\mathbf{V}_y, \xi) : y \in Y)] \\ &\geq 1 - \mathbb{E}_{\substack{\xi \sim \Xi \\ (Y, P, Q) \sim \mathcal{X}}} [ |B|^{|Y|} (1 - \text{sound}_{Y, Q}(S(\mathbf{V}_y, \xi) : y \in Y))] && \text{by (37)} \\ &\geq 1 - \mathbb{E}_{\substack{\xi \sim \Xi \\ (Y, P, Q) \sim \mathcal{X}}} [O_{\mathbf{B}}(1)(1 - \text{sound}_{Y, Q}(S(\mathbf{V}_y, \xi) : y \in Y))] && \mathbf{B} \text{ has bounded arity} \\ &= 1 - O_{\mathbf{B}}(1) \mathbb{E}_{\xi \sim \Xi} [1 - \text{sound}_{\mathcal{X}, \mathbf{B}}(S(\mathbf{V}_x, \xi) : x \in X)] \\ &\geq 1 - O_{\mathbf{B}}(\beta), && \text{by (36)} \end{aligned}$$

as desired.  $\square$

Next, we show that  $S_{\text{COR}}$  rounds similar local configurations to near-identical outputs.

**Lemma 6.35.** *Let  $\mathbf{V}, \mathbf{W} \in \mathbb{S}_A^N$  be such that  $\|S(\mathbf{V}) - S(\mathbf{W})\|_{2,2} \leq \eta$ . Then,*

$$\Pr_{\substack{\xi \sim \Xi \\ \mathbf{p} \sim \Xi_{\text{COR}, B}}} [S_{\text{COR}}(\mathbf{V}, \xi, \mathbf{p}) \neq S_{\text{COR}}(\mathbf{W}, \xi, \mathbf{p})] \leq O_{\mathbf{B}}(\eta).$$

*Proof.* We have that

$$\begin{aligned}
& \Pr_{\substack{\xi \sim \Xi \\ \mathbf{p} \sim \Xi_{\text{COR},B}}} [S_{\text{COR}}(\mathbf{V}, \xi, \mathbf{p}) \neq S_{\text{COR}}(\mathbf{W}, \xi, \mathbf{p})] \\
&= \Pr_{\substack{\xi \sim \Xi \\ \mathbf{p} \sim \Xi_{\text{COR},B}}} [\text{plur}(S(\mathbf{V}, \xi) + \mathbf{p}) \neq \text{plur}(S(\mathbf{W}, \xi) + \mathbf{p})] \\
&= \sum_{b \neq b' \in B} \Pr_{\substack{\xi \sim \Xi \\ \mathbf{p} \sim \Xi_{\text{COR},B}}} [\text{plur}(S(\mathbf{V}, \xi) + \mathbf{p}) = b \wedge \text{plur}(S(\mathbf{W}, \xi) + \mathbf{p}) = b'] \\
&\leq \sum_{b \neq b' \in B} \Pr_{\substack{\xi \sim \Xi \\ \mathbf{p} \sim \Xi_{\text{COR},B}}} [S(\mathbf{V}, \xi)_b + p_b \geq S(\mathbf{V}, \xi)_{b'} + p_{b'} \wedge S(\mathbf{W}, \xi)_{b'} + p_{b'} \geq S(\mathbf{W}, \xi)_b + p_b] \\
&= \sum_{b \neq b' \in B} \Pr_{\substack{\xi \sim \Xi \\ \mathbf{p} \sim [0,1/(2|B|)]^B}} [S(\mathbf{V}, \xi)_b - S(\mathbf{V}, \xi)_{b'} \geq p_{b'} - p_b \geq S(\mathbf{W}, \xi)_b - S(\mathbf{W}, \xi)_{b'}] \\
&\leq \sum_{b \neq b' \in B} \mathbb{E}_{\xi \sim \Xi} [2|B| \cdot |S(\mathbf{V}, \xi)_b - S(\mathbf{V}, \xi)_{b'} - (S(\mathbf{W}, \xi)_b - S(\mathbf{W}, \xi)_{b'})|] \tag{*} \\
&\leq 2|B|^3 \mathbb{E}_{\xi \sim \Xi} [\|S(\mathbf{V}, \xi) - S(\mathbf{W}, \xi)\|_1] \tag{**} \\
&\leq 2|B|^3 \|S(\mathbf{V}) - S(\mathbf{W})\|_{2,2} = O_{\mathbf{B}}(\eta),
\end{aligned}$$

where (\*) follows from the fact that the CDF of the distribution  $p_b - p_{b'}$  is bounded by  $2|B|$  everywhere and (\*\*) follows from Hölder's inequality (and the fact that  $\mathbb{E}_{\xi \sim \Xi}[1] = 1$ ).  $\square$

## 6.7 Proof of Theorem 6.1

We now have all the pieces to prove Theorem 6.1, which we restate for convenience.

**Theorem 6.1.** *Assume UGC. If  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  is  $(\alpha, \beta)$ -robust for some  $0 < \alpha < \beta < 1$ , then  $\text{PCSP}((\mathbf{A}, \mathbf{B}) + \text{EQ})$  is  $(\alpha_{\text{EQ}}, \beta_{\text{EQ}})$ -robust for*

$$\alpha_{\text{EQ}} = \Omega_{\mathbf{A}}(\alpha^2 \beta^4), \tag{26}$$

$$\beta_{\text{EQ}} = O_{\mathbf{B}}(\beta). \tag{27}$$

*Proof.* By Theorem 6.4, it suffices to prove that, for every instance  $\mathcal{X}$  of  $\text{PCSP}((\mathbf{A}, \mathbf{B}) + \text{EQ})$  with SDP value at least  $1 - 2\alpha_{\text{EQ}}$ , there exists an integral assignment  $\mathcal{X} \rightarrow \mathbf{B}$  with soundness at least  $1 - \beta_{\text{EQ}}/2$ . In particular, we do not need to present an efficient algorithm as that automatically follows from Raghavendra's theorem.

Let  $\mathcal{X}$  be an arbitrary instance of  $\text{PCSP}((\mathbf{A}, \mathbf{B}) + \text{EQ})$  and let  $\mathcal{V} \in \mathbb{V}_{\mathcal{X}, \mathbf{A} + \text{EQ}}^N$  have completeness at least  $1 - 2\alpha_{\text{EQ}}$ . Split  $\mathcal{X}$  into two parts:  $\mathcal{X}_{\mathbf{A}, \mathbf{B}}$  as an instance of  $\text{PCSP}(\mathbf{A}, \mathbf{B})$  and  $\mathcal{X}_{\text{EQ}}$  as an instance of  $\text{PCSP}(\text{EQ})$ . Let  $p \in [0, 1]$  be the fraction of weight on  $\mathcal{X}_{\mathbf{A}, \mathbf{B}}$ . Note that any assignment to the variables that sets all variables to the same value satisfies every constraint in  $\mathcal{X}_{\text{EQ}}$ . Thus, the best integral assignment to  $\mathcal{X}$  has value at least  $1 - p$ . If  $1 - p \geq 1 - \beta$ , we are done, so we may assume that  $1 - p < 1 - \beta$  (i.e.,  $p > \beta > \alpha$ ). Note further that

$$1 - 2\alpha_{\text{EQ}} \leq \text{comp}_{\mathcal{X}, \mathbf{A} + \text{EQ}}(\mathcal{V}) = p \text{comp}_{\mathcal{X}_{\mathbf{A}, \mathbf{B}}, \mathbf{A}}(\mathcal{V}) + (1 - p) \text{comp}_{\mathcal{X}_{\text{EQ}}, \text{EQ}}(\mathcal{V}).$$

Since  $\text{comp}$  has range  $[0, 1]$ , we can deduce that

$$\text{comp}_{\mathcal{X}_{\mathbf{A},\mathbf{B},\mathbf{A}}}(\mathcal{V}) \geq 1 - \frac{2\alpha_{\text{EQ}}}{p} \geq 1 - \frac{2\alpha_{\text{EQ}}}{\alpha}, \quad (39)$$

$$\text{comp}_{\mathcal{X}_{\text{EQ},\text{EQ}}}(\mathcal{V}) \geq 1 - \frac{2\alpha_{\text{EQ}}}{1-p}. \quad (40)$$

By Corollary 6.7, we may pick  $\theta = \alpha/2$  so that, for any set  $Z$ ,  $\text{BCR}_{\theta}^{|Z|-|A|+N}$  is  $(\alpha/2, \beta + \alpha/2)$ -robust for  $\mathcal{X}$  (for sufficiently large  $N$ ). Pick  $|Z| \geq 2|X|^2/\alpha$  and  $\delta = \Theta_{\mathbf{A}}(\alpha^2)$ . By Lemma 6.30, we have that the collection of rounding schemes  $\text{REQ}_{\theta,\delta}^{N,Z}$  is  $(\alpha/2 - O_{\mathbf{A}}(\sqrt{\delta}), \beta + \alpha/2 + \frac{|X|^2}{|Z|})$ -robust, where we pick  $\delta$  sufficiently small so that  $\text{REQ}_{\theta,\delta}^{N,Z}$  is  $(\alpha/4, 2\beta)$ -robust. Since by (39) we have that  $\text{comp}_{\mathcal{X}_{\mathbf{A},\mathbf{B},\mathbf{A}}}(\mathcal{V}) \geq 1 - 2\alpha_{\text{EQ}}/\alpha \geq 1 - \alpha/4$ , we have that applying  $\text{REQ}_{\theta,\delta}^{N,Z}$  to  $\mathbf{V}$  satisfies at least  $1 - 2\beta$  fraction of the constraints of  $\mathcal{X}_{\mathbf{A},\mathbf{B}}$ .

In particular, when  $\text{REQ}_{\theta,\delta}^{N,Z}$  is applied to  $\mathbf{V}$ , we satisfy  $1 - 2\beta$  fraction of the constraints of  $\mathcal{X}_{\mathbf{A},\mathbf{B}}$  on average. Furthermore, by Lemma 6.34,  $\text{REQ}_{\theta,\delta,\text{COR}}^{N,Z}$  is  $(\alpha/4, \Theta_{\mathbf{B}}(\beta))$ -robust for  $\mathcal{X}$ . So,  $\text{REQ}_{\theta,\delta,\text{COR}}^{N,Z}$  when applied to  $\mathcal{V}$  satisfies  $1 - O_{\mathbf{B}}(\beta)$ -fraction of the constraints of  $\mathcal{X}_{\mathbf{A},\mathbf{B}}$ .

Now, we evaluate the performance of  $\text{REQ}_{\theta,\delta,\text{COR}}^{N,Z}$  rounding  $\mathcal{V}$  for  $\mathcal{X}_{\text{EQ}}$ . Fix a constraint  $(x, y)$ , and assume that  $\text{comp}_{\{x,y\},\text{EQ}}(\mathcal{V}) = 1 - \gamma$ . Observe that

$$\text{comp}_{\{x,y\},\text{EQ}}(\mathcal{V}) = 1 - \frac{\|\mathbf{V}_x - \mathbf{V}_y\|_2^2}{2},$$

so  $\|\mathbf{V}_x - \mathbf{V}_y\|_2^2 = 2\gamma$ . By Lemma 6.31 with  $\eta = \sqrt{2\gamma/\delta}$ , we have that

$$\|\text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V}_x) - \text{REQ}_{\theta,\delta}^{N,Z}(\mathbf{V}_y)\|_{2,2} \leq 8\sqrt[4]{\frac{2\gamma}{\delta}} \leq 10\sqrt[4]{\frac{\gamma}{\delta}}.$$

Thus, by Lemma 6.35, we have that

$$\Pr_{\substack{\xi \sim \Xi_{|Z|-|A|+N,\theta}^{\text{BCR}} \\ \mathbf{p} \sim \Xi_{\text{COR},\mathbf{B}}}} [\text{REQ}_{\theta,\delta,\text{COR}}^{N,Z}(\mathbf{V}_x, \xi, \mathbf{p}) \neq \text{REQ}_{\theta,\delta,\text{COR}}^{N,Z}(\mathbf{V}_y, \xi, \mathbf{p})] \leq O_{\mathbf{B}}((\gamma/\delta)^{1/4}) = O_{\mathbf{B}}(\gamma^{1/4}\alpha^{-1/2}).$$

In other words, the probability that  $x$  and  $y$  are rounded to the same value by applying  $\text{REQ}_{\theta,\delta,\text{COR}}^{N,Z}$  to  $\mathcal{V}$  is  $1 - O_{\mathbf{B}}(\gamma^{1/4}\alpha^{-1/2})$ . Since this soundness is a convex function of  $\gamma$ , we have that, if  $\text{comp}_{\mathcal{X}_{\text{EQ},\text{EQ}}}(\mathcal{V}) = 1 - \lambda$ , then  $\text{REQ}_{\theta,\delta,\text{COR}}^{N,Z}$  satisfies  $1 - O_{\mathbf{B}}(\lambda^{1/4}\alpha^{-1/2})$  fraction of the equality constraints on average. By (40), we have that  $\lambda \leq \frac{2\alpha_{\text{EQ}}}{1-p}$ , so the expected number of constraints of  $\mathcal{X}$  satisfied by applying  $\text{REQ}_{\theta,\delta,\text{COR}}^{N,Z}$  to  $\mathcal{V}$  is at least

$$\begin{aligned} p(1 - O_{\mathbf{B}}(\beta)) + (1-p)(1 - O_{\mathbf{B}}(\lambda^{1/4}\alpha^{-1/2})) &\geq 1 - p \cdot O_{\mathbf{B}}(\beta) - (1-p)O_{\mathbf{B}}\left(\frac{(2\alpha_{\text{EQ}})^{1/4}}{(1-p)^{1/4}\sqrt{\alpha}}\right) \\ &\geq 1 - O_{\mathbf{B}}\left(p\beta + (1-p)^{3/4}\frac{\alpha_{\text{EQ}}^{1/4}}{\sqrt{\alpha}}\right) \\ &\geq 1 - O_{\mathbf{B}}\left(\beta + \frac{\alpha_{\text{EQ}}^{1/4}}{\sqrt{\alpha}}\right), \end{aligned}$$

where the last line bound  $p \leq 1$  and  $(1-p)^{3/4} \leq 1$  for any  $p \in [0, 1]$ . In particular, for  $\alpha_{\text{EQ}} = \Theta_{\mathbf{B}}(\alpha^2\beta^4)$ , we get  $1 - O_{\mathbf{B}}(\beta) = 1 - \beta_{\text{EQ}}$  fraction of the constraints are satisfied, as desired.  $\square$

## 7 Conclusion

In this paper, we proved a number of new results on the nature of Robust Promise CSPs. First, our integrality gap for (1-IN-3-SAT, NAE-SAT) shows that the  $(\epsilon, O(\log \log(1/\epsilon)/\log(1/\epsilon)))$ -robust algorithm for the AT polymorphism due to Brakensiek–Guruswami–Sandeep [BGS23b] is nearly tight. Second, our algorithms for the Majority, Plurality, and separable polymorphisms show that robust algorithms due to Charikar–Makarychev–Makarychev [CMM06, CMM09] can be extended to quite general settings. Finally, our robust algorithm for a robust PCSP template with equality added shows that a rich theory of gadget reductions between PCSPs can be used in the theory of robust CSPs in a black-box manner. Moving forward, there are a number of directions for further exploration.

- In Section 3, we presented an integrality gap for  $\text{fiPCSP}(1\text{-IN-3-SAT}, \text{NAE-SAT})$ . More generally, can we get an integrality gap for  $\text{fiPCSP}(a\text{-IN-}b\text{-SAT}, \text{NAE-SAT})$  with similar asymptotics? In the case for which  $b = 2a$ , then MAJ are polymorphisms, so in general we consider  $a/b \in (0, 1) \setminus \{1/2\}$ . As far as we know, one cannot use only gadget reductions to establish this broader family of gaps.
- Recall that Brakensiek, Guruswami, and Sandeep [BGS23b] constructed robust algorithms for all Boolean PCSPs with the MAJ or AT polymorphisms. In Section 4 and Section 5, we greatly improved the analysis and scope of BGS’s MAJ algorithm. What would such a generalization look like for AT? One possible direction is to remove the “ $\alpha$ -conservative” assumption on the map  $\rho$  in Theorem 5.3.
- Can one generalize the analysis in Section 6 to fully resolve Question 1.4? With the current techniques, a loss of at least  $\epsilon^{1/2}$  seems necessary due to the use of the geometry of  $L^2$ -spaces (e.g., Proposition 6.32). Perhaps sharper asymptotics could be achieved by a direct construction of a quasirandom approximate polymorphisms for  $\Gamma \cup \{\text{EQ}\}$ .

## References

- [AFO<sup>+</sup>25] Sergey Avvakumov, Marek Filakovský, Jakub Oprsal, Gianluca Tassinato, and Uli Wagner. Hardness of 4-colouring  $k$ -colourable graphs. In *Proc. 57th Annual ACM Symposium on Theory of Computing (STOC’25)*, pages 72–83. ACM, 2025.
- [AGH17] Per Austrin, Venkatesan Guruswami, and Johan Håstad.  $(2+\epsilon)$ -Sat is NP-hard. *SIAM J. Comput.*, 46(5):1554–1573, 2017.
- [BBB21] Libor Barto, Diego Battistelli, and Kevin M. Berg. Symmetric Promise Constraint Satisfaction Problems: Beyond the Boolean Case. In *Proc. 38th International Symposium on Theoretical Aspects of Computer Science (STACS’21)*, volume 187 of *LIPICs*, pages 10:1–10:16, 2021.
- [BBKO21] Libor Barto, Jakub Bulín, Andrei A. Krokhin, and Jakub Opršal. Algebraic approach to promise constraint satisfaction. *J. ACM*, 68(4):28:1–28:66, 2021.
- [BD06] Andrei A. Bulatov and Víctor Dalmau. A simple algorithm for Mal’tsev constraints. *SIAM J. Comput.*, 36(1):16–27, 2006.

- [BG19] Joshua Brakensiek and Venkatesan Guruswami. An algorithmic blend of lps and ring equations for promise csps. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 436–455. SIAM, 2019.
- [BG21a] Joshua Brakensiek and Venkatesan Guruswami. Promise Constraint Satisfaction: Algebraic Structure and a Symmetric Boolean Dichotomy. *SIAM J. Comput.*, 50(6):1663–1700, 2021.
- [BG21b] Joshua Brakensiek and Venkatesan Guruswami. The quest for strong inapproximability results with perfect completeness. *ACM Trans. Algorithms*, 17(3):27:1–27:35, 2021.
- [BGS23a] Joshua Brakensiek, Venkatesan Guruswami, and Sai Sandeep. Conditional Dichotomy of Boolean Ordered Promise CSPs. *TheoretCS*, Volume 2:8967, January 2023.
- [BGS23b] Joshua Brakensiek, Venkatesan Guruswami, and Sai Sandeep. SDPs and robust satisfiability of promise CSP. In *Proc. 55th Annual ACM Symposium on Theory of Computing (STOC’23)*, pages 609–622. ACM, 2023.
- [BGWŽ20] Joshua Brakensiek, Venkatesan Guruswami, Marcin Wrochna, and Stanislav Živný. The power of the combined basic LP and affine relaxation for promise CSPs. *SIAM J. Comput.*, 49(6):1232–1248, 2020.
- [BJK05] Andrei A. Bulatov, Peter Jeavons, and Andrei A. Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.*, 34(3):720–742, 2005.
- [BK14] Libor Barto and Marcin Kozik. Constraint satisfaction problems solvable by local consistency methods. *J. ACM*, 61(1), 2014.
- [BK16] Libor Barto and Marcin Kozik. Robustly solvable constraint satisfaction problems. *SIAM J. Comput.*, 45(4):1646–1669, 2016.
- [BK22] Libor Barto and Marcin Kozik. Combinatorial gap theorem and reductions between promise csps. In *Proc. 2022 ACM-SIAM Symposium on Discrete Algorithms (SODA’22)*, pages 1204–1220. SIAM, 2022.
- [BK24] Demian Banakh and Marcin Kozik. Injective hardness condition for pcsps. In Pawel Sobocinski, Ugo Dal Lago, and Javier Esparza, editors, *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2024, Tallinn, Estonia, July 8-11, 2024*, pages 8:1–8:10. ACM, 2024.
- [BKM21] Mark Braverman, Subhash Khot, and Dor Minzer. On rich 2-to-1 games. In *Proc. 12th Innovations in Theoretical Computer Science Conference (ITCS’21)*, volume 185 of *LIPICs*, pages 27:1–27:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.
- [BR16] Jonah Brown-Cohen and Prasad Raghavendra. Correlation decay and tractability of CSPs. In *Proc. 43rd International Colloquium on Automata, Languages, and Programming (ICALP’16)*, volume 55 of *LIPICs*, pages 79:1–79:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

- [Bul17] Andrei A. Bulatov. A dichotomy theorem for nonuniform CSPs. In *Proc. 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS'17)*, pages 319–330, USA, 2017. IEEE.
- [BV04] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [BWŽ21] Alex Brandts, Marcin Wrochna, and Stanislav Živný. The complexity of promise SAT on non-Boolean domains. *ACM Trans. Comput. Theory*, 13(4):26:1–26:20, 2021.
- [BŽ22] Alex Brandts and Stanislav Živný. Beyond PCSP(1-in-3,NAE). *Information and Computation*, 2022.
- [CKK+25] Lorenzo Ciardo, Marcin Kozik, Andrei A. Krokhin, Tamio-Vesa Nakajima, and Stanislav Živný. 1-in-3 vs. not-all-equal: Dichotomy of a broken promise. *ACM Trans. Comput. Log.*, 26(2):10:1–10:22, 2025.
- [CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, pages 205–214, Seattle WA USA, May 2006. ACM.
- [CMM09] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. *ACM Trans. Algorithms*, 5(3):32:1–32:14, 2009.
- [CŽ23a] Lorenzo Ciardo and Stanislav Živný. Approximate graph colouring and crystals. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2256–2267. SIAM, 2023.
- [CŽ23b] Lorenzo Ciardo and Stanislav Živný. Approximate graph colouring and the hollow shadow. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 623–631. ACM, 2023.
- [CŽ23c] Lorenzo Ciardo and Stanislav Živný. CLAP: A new algorithm for promise CSPs. *SIAM J. Comput.*, 52(1):1–37, 2023.
- [CŽ23d] Lorenzo Ciardo and Stanislav Živný. Hierarchies of minion tests for PCSPs through tensors. In *Proc. 2023 ACM-SIAM Symposium on Discrete Algorithms (SODA'23)*, pages 568–580, 2023.
- [CŽ24] Lorenzo Ciardo and Stanislav Živný. Semidefinite programming and linear equations vs. homomorphism problems. In Bojan Mohar, Igor Shinkar, and Ryan O'Donnell, editors, *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 1935–1943. ACM, 2024.
- [DiB16] Emmanuele DiBenedetto. *Real Analysis*. Birkhäuser Advanced Texts Basler Lehrbücher. Springer New York, New York, NY, 2016.
- [DK13] Víctor Dalmau and Andrei A. Krokhin. Robust satisfiability for csps: Hardness and algorithmic results. *ACM Trans. Comput. Theory*, 5(4):15:1–15:25, 2013.

- [FKOS19] Miron Ficak, Marcin Kozik, Miroslav Olšák, and Szymon Stankiewicz. Dichotomy for Symmetric Boolean PCSPs. In *Proc. 46th International Colloquium on Automata, Languages, and Programming (ICALP'19)*, volume 132, pages 57:1–57:12. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019.
- [FNO<sup>+</sup>24] Marek Filakovský, Tamio-Vesa Nakajima, Jakub Opršal, Gianluca Tasinato, and Uli Wagner. Hardness of Linearly Ordered 4-Colouring of 3-Colourable 3-Uniform Hypergraphs. In *Proc. 41st International Symposium on Theoretical Aspects of Computer Science (STACS'24)*, volume 289 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:19, 2024.
- [GJ76] M. R. Garey and David S. Johnson. The complexity of near-optimal graph coloring. *J. ACM*, 23(1):43–49, 1976.
- [GLS93] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.
- [GM12] Bernd Gärtner and Jiri Matousek. *Approximation algorithms and semidefinite programming*. Springer Science & Business Media, 2012.
- [Gro75] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.
- [GS20] Venkatesan Guruswami and Sai Sandeep. d-To-1 Hardness of Coloring 3-Colorable Graphs with  $O(1)$  Colors. In *Proc. 47th International Colloquium on Automata, Languages, and Programming (ICALP'20)*, volume 168 of *LIPIcs*, pages 62:1–62:12. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020.
- [GZ12] Venkatesan Guruswami and Yuan Zhou. Tight bounds on the approximability of almost-satisfiable horn SAT and exact hitting set. *Theory Comput.*, 8(1):239–267, 2012.
- [Hås01] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.
- [HMNŽ24] Johan Håstad, Björn Martinsson, Tamio-Vesa Nakajima, and Stanislav Živný. A logarithmic approximation of linearly-ordered colourings. In Amit Kumar and Noga Ron-Zewi, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2024, August 28-30, 2024, London School of Economics, London, UK*, volume 317 of *LIPIcs*, pages 7:1–7:6. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024.
- [IMM<sup>+</sup>10] Pawel M. Idziak, Petar Markovic, Ralph McKenzie, Matthew Valeriote, and Ross Willard. Tractability and learnability arising from algebras with few subpowers. *SIAM J. Comput.*, 39(7):3023–3037, 2010.
- [Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC'02: Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775. ACM, New York, 2002.

- [KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell. Optimal Inapproximability Results for MAX-CUT and Other 2-Variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.
- [KOWŽ23] Andrei Krokhin, Jakub Opršal, Marcin Wrochna, and Stanislav Živný. Topology and adjunction in promise constraint satisfaction. *SIAM J. Comput.*, 52(1):37–79, 2023.
- [LM00] Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of statistics*, pages 1302–1338, 2000.
- [LŽ25] Alberto Larrauri and Stanislav Živný. Solving promise equations over monoids and groups. *ACM Trans. Comput. Log.*, 26(1):3:1–3:24, 2025.
- [MOO10] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. of Math. (2)*, 171(1):295–341, 2010.
- [Mot25] Antoine Mottet. Algebraic and algorithmic synergies between promise and infinite-domain csp. In *Proceedings of the 40th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2025*, 2025.
- [NVWŽ25] Tamio-Vesa Nakajima, Zephyr Verwimp, Marcin Wrochna, and Stanislav Živný. Complexity of approximate conflict-free, linearly-ordered, and nonmonochromatic hypergraph colourings. In Keren Censor-Hillel, Fabrizio Grandoni, Joël Ouaknine, and Gabriele Puppis, editors, *52nd International Colloquium on Automata, Languages, and Programming, ICALP 2025, July 8-11, 2025, Aarhus, Denmark*, volume 334 of *LIPICs*, pages 169:1–169:10. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.
- [NŽ22] Tamio-Vesa Nakajima and Stanislav Živný. Linearly ordered colourings of hypergraphs. *ACM Trans. Comput. Theory*, 13(3–4), 2022.
- [NŽ24] Tamio-Vesa Nakajima and Stanislav Živný. On the complexity of symmetric vs. functional PCSPs. *ACM Trans. Algorithms*, 20(4):33:1–33:29, 2024.
- [O’D14] Ryan O’Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.
- [Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proc. 40th Annual ACM Symposium on Theory of Computing (STOC’08)*, pages 245–254, 2008.
- [RS09] Prasad Raghavendra and David Steurer. How to round any csp. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 586–594. IEEE, 2009.
- [Sch78] Thomas Schaefer. The complexity of satisfiability problems. In *Proc. 10th Annual ACM Symposium on the Theory of Computing (STOC’78)*, pages 216–226, 1978.
- [Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2018.

- [Zhu20] Dmitriy Zhuk. A proof of the CSP dichotomy conjecture. *J. ACM*, 67(5):30:1–30:78, 2020.
- [Zwi98] Uri Zwick. Finding almost-satisfying assignments. In *Proc. 13th Annual ACM Symposium on the Theory of Computing (STOC'98)*, pages 551–560. ACM, 1998.

## A Omitted proofs

In this section, we give the proof of some of the technical results stated in the previous sections of the paper.

### A.1 Omitted proofs from Section 3

We start by formally establishing the first statement in the following theorem (cf. the sketch of the proof given in Section 3.3).

**Theorem 3.7.** *For all  $\epsilon > 0$ , if  $d \geq \frac{80000000}{\epsilon^3}$  then for the integrality gap instance described above (Section 3.3),*

1. *There is a solution to the SDP that gives a value of at least  $1 - \epsilon$  to each constraint that is not discarded.*
2. *At most  $6e^{-\frac{10}{\epsilon}}$  of the constraints that were sampled are discarded.*
3. *All assignments violate at least  $\frac{1}{10^{60} \lceil \log(\frac{1}{\epsilon}) \rceil}$  of the sampled NAE-SAT constraints (this includes constraints that were discarded).*

We shall use some technical lemmas.

**Lemma A.1.** *If  $\mathbf{v}_1, \mathbf{v}_2$ , and  $\mathbf{v}_3$  are unit vectors such that  $\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 = -\mathbf{v}_0$  then there is a probability distribution  $D$  supported on the assignments  $\{(-1, -1, 1), (-1, 1, -1), (1, -1, -1)\}$  such that*

1. *For all  $i \in [3]$ ,  $\mathbb{E}_D[x_i] = \mathbf{v}_i \cdot \mathbf{v}_0$ .*
2. *For all distinct  $i, j \in [3]$ ,  $\mathbb{E}_D[x_i x_j] = \mathbf{v}_i \cdot \mathbf{v}_j$ .*

*Proof.* Let  $D$  be the probability distribution where we take the assignment  $(-1, -1, 1)$  with probability  $\frac{1+\mathbf{v}_3 \cdot \mathbf{v}_0}{2}$ , we take the assignment  $(-1, 1, -1)$  with probability  $\frac{1+\mathbf{v}_2 \cdot \mathbf{v}_0}{2}$ , and we take the assignment  $(1, -1, -1)$  with probability  $\frac{1+\mathbf{v}_1 \cdot \mathbf{v}_0}{2}$ . We now make the following observations:

1.  $\mathbb{E}_D[x_1] = \frac{1+\mathbf{v}_1 \cdot \mathbf{v}_0}{2} - \frac{1+\mathbf{v}_2 \cdot \mathbf{v}_0}{2} - \frac{1+\mathbf{v}_3 \cdot \mathbf{v}_0}{2} = \mathbf{v}_1 \cdot \mathbf{v}_0 - \frac{1+(\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}_3) \cdot \mathbf{v}_0}{2} = \mathbf{v}_1 \cdot \mathbf{v}_0$ .
- 2.

$$\begin{aligned} \mathbb{E}_D[x_1 x_2] &= -\frac{1 + \mathbf{v}_1 \cdot \mathbf{v}_0}{2} - \frac{1 + \mathbf{v}_2 \cdot \mathbf{v}_0}{2} + \frac{1 + \mathbf{v}_3 \cdot \mathbf{v}_0}{2} = \frac{-1 + (\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3) \cdot (\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3)}{2} \\ &= \frac{-1 + \mathbf{v}_1 \cdot \mathbf{v}_1 + 2\mathbf{v}_1 \cdot \mathbf{v}_2 + \mathbf{v}_2 \cdot \mathbf{v}_2 - \mathbf{v}_3 \cdot \mathbf{v}_3}{2} = \mathbf{v}_1 \cdot \mathbf{v}_2 \end{aligned}$$

The remaining cases follow by symmetry. □

**Lemma A.2.** *Given unit vectors  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}'_1, \mathbf{v}'_2,$  and  $\mathbf{v}'_3$  such that*

1. *For all  $i \in [3], \mathbf{v}'_i \cdot \mathbf{v}_0 = \mathbf{v}_i \cdot \mathbf{v}_0 + \Delta_i.$*
2. *For all distinct  $i, j \in [3], \mathbf{v}'_i \cdot \mathbf{v}'_j = \mathbf{v}_i \cdot \mathbf{v}_j + \Delta_{ij}.$*

*for some  $\Delta_1, \Delta_2, \Delta_3, \Delta_{12}, \Delta_{13}, \Delta_{23} \in \mathbb{R},$  if there is a probability distribution  $D$  such that*

1. *For all  $i \in [3], \mathbb{E}_D[x_i] = \mathbf{v}_i \cdot \mathbf{v}_0.$*
2. *For all distinct  $i, j \in [3], \mathbb{E}_D[x_i x_j] = \mathbf{v}_i \cdot \mathbf{v}_j.$*

*then there is a pseudo-distribution  $D'$  (i.e.,  $D'$  may give negative probabilities to some events but the probabilities should still sum to 1) such that*

1. *For all  $i \in [3], \mathbb{E}_{D'}[x_i] = \mathbf{v}'_i \cdot \mathbf{v}_0.$*
2. *For all distinct  $i, j \in [3], \mathbb{E}_{D'}[x_i x_j] = \mathbf{v}'_i \cdot \mathbf{v}'_j.$*
3. *For each assignment  $(x_1, x_2, x_3) \in \{-1, 1\}^3,$  the (pseudo-)probabilities of that assignment under  $D'$  and  $D$  differ by at most  $\frac{|\Delta_1|+|\Delta_2|+|\Delta_3|+|\Delta_{12}|+|\Delta_{13}|+|\Delta_{23}|}{8}.$*

This follows from the following proposition.

**Proposition A.3.** *If  $\mathbf{v}_1, \mathbf{v}_2,$  and  $\mathbf{v}_3$  are unit vectors and  $p_{-1,-1,-1}, p_{-1,-1,1}, p_{-1,1,-1}, p_{-1,1,1}, p_{1,-1,-1}, p_{1,-1,1}, p_{1,1,-1},$  and  $p_{1,1,1}$  are real numbers such that*

1.  $p_{1,-1,-1} + p_{1,-1,1} + p_{1,1,-1} + p_{1,1,1} + p_{-1,-1,-1} + p_{-1,-1,1} + p_{-1,1,-1} + p_{-1,1,1} = 1$
2.  $\mathbf{v}_1 \cdot \mathbf{v}_0 = p_{1,-1,-1} + p_{1,-1,1} + p_{1,1,-1} + p_{1,1,1} - p_{-1,-1,-1} - p_{-1,-1,1} - p_{-1,1,-1} - p_{-1,1,1}.$
3.  $\mathbf{v}_2 \cdot \mathbf{v}_0 = p_{-1,1,-1} + p_{-1,1,1} + p_{1,1,-1} + p_{1,1,1} - p_{-1,-1,-1} - p_{-1,-1,1} - p_{1,-1,-1} - p_{1,-1,1}.$
4.  $\mathbf{v}_3 \cdot \mathbf{v}_0 = p_{-1,-1,1} + p_{-1,1,1} + p_{1,-1,1} + p_{1,1,1} - p_{-1,-1,-1} - p_{-1,1,-1} - p_{1,-1,-1} - p_{1,1,-1}.$
5.  $\mathbf{v}_1 \cdot \mathbf{v}_2 = p_{-1,-1,-1} + p_{-1,-1,1} + p_{1,1,-1} + p_{1,1,1} - p_{-1,1,-1} - p_{-1,1,1} - p_{1,-1,-1} - p_{1,-1,1}.$
6.  $\mathbf{v}_1 \cdot \mathbf{v}_3 = p_{-1,-1,-1} + p_{-1,1,-1} + p_{1,-1,1} + p_{1,1,1} - p_{-1,-1,1} - p_{-1,1,1} - p_{1,-1,-1} - p_{1,1,-1}.$
7.  $\mathbf{v}_2 \cdot \mathbf{v}_3 = p_{-1,-1,-1} + p_{1,-1,-1} + p_{-1,1,1} + p_{1,1,1} - p_{-1,-1,1} - p_{1,-1,1} - p_{-1,1,-1} - p_{1,1,-1}.$

*and  $\mathbf{v}'_1, \mathbf{v}'_2,$  and  $\mathbf{v}'_3$  are unit vectors such that*

1. *For all  $i \in [3], \mathbf{v}'_i \cdot \mathbf{v}_0 = \mathbf{v}_i \cdot \mathbf{v}_0 + \Delta_i.$*
2. *For all distinct  $i, j \in [3], \mathbf{v}'_i \cdot \mathbf{v}'_j = \mathbf{v}_i \cdot \mathbf{v}_j + \Delta_{ij}.$*

*for some  $\Delta_1, \Delta_2, \Delta_3, \Delta_{12}, \Delta_{13}, \Delta_{23} \in \mathbb{R}$  then if we take*

1.  $p'_{-1,-1,-1} = p_{-1,-1,-1} + \frac{-\Delta_1 - \Delta_2 - \Delta_3 + \Delta_{12} + \Delta_{13} + \Delta_{23}}{8}$
2.  $p'_{-1,-1,1} = p_{-1,-1,1} + \frac{-\Delta_1 - \Delta_2 + \Delta_3 + \Delta_{12} - \Delta_{13} - \Delta_{23}}{8}$

3.  $p'_{-1,1,-1} = p_{-1,1,-1} + \frac{-\Delta_1 + \Delta_2 - \Delta_3 - \Delta_{12} + \Delta_{13} - \Delta_{23}}{8}$
4.  $p'_{1,-1,-1} = p_{1,-1,-1} + \frac{\Delta_1 - \Delta_2 - \Delta_3 - \Delta_{12} - \Delta_{13} + \Delta_{23}}{8}$
5.  $p'_{-1,1,1} = p_{-1,1,1} + \frac{-\Delta_1 + \Delta_2 + \Delta_3 - \Delta_{12} - \Delta_{13} + \Delta_{23}}{8}$
6.  $p'_{1,-1,1} = p_{1,-1,1} + \frac{\Delta_1 - \Delta_2 + \Delta_3 - \Delta_{12} + \Delta_{13} - \Delta_{23}}{8}$
7.  $p'_{1,1,-1} = p_{1,1,-1} + \frac{\Delta_1 + \Delta_2 - \Delta_3 + \Delta_{12} - \Delta_{13} - \Delta_{23}}{8}$
8.  $p'_{1,1,1} = p_{1,1,1} + \frac{\Delta_1 + \Delta_2 + \Delta_3 + \Delta_{12} + \Delta_{13} + \Delta_{23}}{8}$

we will have that

1.  $p'_{1,-1,-1} + p'_{-1,1,1} + p'_{1,1,-1} + p'_{1,1,1} + p'_{-1,-1,-1} + p'_{-1,-1,1} + p'_{-1,1,-1} + p'_{-1,1,1} = 1$
2.  $\mathbf{v}'_1 \cdot \mathbf{v}_0 = p'_{1,-1,-1} + p'_{-1,1,1} + p'_{1,1,-1} + p'_{1,1,1} - p'_{-1,-1,-1} - p'_{-1,-1,1} - p'_{-1,1,-1} - p'_{-1,1,1}$
3.  $\mathbf{v}'_2 \cdot \mathbf{v}_0 = p'_{-1,1,-1} + p'_{-1,1,1} + p'_{1,1,-1} + p'_{1,1,1} - p'_{-1,-1,-1} - p'_{-1,-1,1} - p'_{1,-1,-1} - p'_{1,-1,1}$
4.  $\mathbf{v}'_3 \cdot \mathbf{v}_0 = p'_{-1,-1,1} + p'_{-1,1,1} + p'_{1,-1,1} + p'_{1,1,1} - p'_{-1,-1,-1} - p'_{-1,1,-1} - p'_{1,-1,-1} - p'_{1,1,-1}$
5.  $\mathbf{v}'_1 \cdot \mathbf{v}'_2 = p'_{-1,-1,-1} + p'_{-1,-1,1} + p'_{1,1,-1} + p'_{1,1,1} - p'_{-1,1,-1} - p'_{-1,1,1} - p'_{1,-1,-1} - p'_{1,-1,1}$
6.  $\mathbf{v}'_1 \cdot \mathbf{v}'_3 = p'_{-1,-1,-1} + p'_{-1,1,-1} + p'_{1,-1,1} + p'_{1,1,1} - p'_{-1,-1,1} - p'_{-1,1,1} - p'_{1,-1,-1} - p'_{1,1,-1}$
7.  $\mathbf{v}'_2 \cdot \mathbf{v}'_3 = p'_{-1,-1,-1} + p'_{1,-1,-1} + p'_{-1,1,1} + p'_{1,1,1} - p'_{-1,-1,1} - p'_{1,-1,1} - p'_{-1,1,-1} - p'_{1,1,-1}$

*Proof.* This follows from the following observations:

1. The  $\pm \frac{\Delta_i}{8}$  terms contribute  $\Delta_i$  to the sum for  $\mathbf{v}'_i \cdot \mathbf{v}_0$  and cancel out for the other sums.
2. Similarly, the  $\pm \frac{\Delta_{ij}}{8}$  terms contribute  $\Delta_{ij}$  to the sum for  $\mathbf{v}'_i \cdot \mathbf{v}'_j$  and cancel out for the other sums.  $\square$

**Lemma A.4.** *If  $\mathbf{v}'_1$ ,  $\mathbf{v}'_2$ , and  $\mathbf{v}'_3$  are unit vectors and  $D'$  is a pseudo-distribution such that*

1. *For all  $i \in [3]$ ,  $\mathbb{E}_{D'}[x_i] = \mathbf{v}'_i \cdot \mathbf{v}_0$ .*
2. *For all distinct  $i, j \in [3]$ ,  $\mathbb{E}_{D'}[x_i x_j] = \mathbf{v}'_i \cdot \mathbf{v}'_j$ .*

*then if we take  $\mathbf{v}''_1 = \sqrt{1 - \epsilon'^2} \mathbf{v}'_1 + \epsilon' \mathbf{z}_1$ ,  $\mathbf{v}''_2 = \sqrt{1 - \epsilon'^2} \mathbf{v}'_2 + \epsilon' \mathbf{z}_2$ , and  $\mathbf{v}''_3 = \sqrt{1 - \epsilon'^2} \mathbf{v}'_3 + \epsilon' \mathbf{z}_3$  where  $\mathbf{z}_1$ ,  $\mathbf{z}_2$ , and  $\mathbf{z}_3$  are orthogonal to  $\mathbf{v}'_1$ ,  $\mathbf{v}'_2$ ,  $\mathbf{v}'_3$ , and each other for some  $\epsilon' > 0$  and take  $D''$  to be the pseudo-distribution where the probability that each assignment occurs under  $D''$  is  $(1 - \epsilon')$  times the probability that the assignment occurs under  $D'$  plus  $\frac{\epsilon'}{8}$  then*

1. *For all  $i \in [3]$ ,  $\mathbb{E}_{D''}[x_i] = \mathbf{v}''_i \cdot \mathbf{v}_0$ .*
2. *For all distinct  $i, j \in [3]$ ,  $\mathbb{E}_{D''}[x_i x_j] = \mathbf{v}''_i \cdot \mathbf{v}''_j$ .*

This follows from the following proposition.

**Proposition A.5.** If  $\mathbf{v}'_1, \mathbf{v}'_2,$  and  $\mathbf{v}'_3$  are unit vectors and  $p'_{-1,-1,-1}, p'_{-1,-1,1}, p'_{-1,1,-1}, p'_{-1,1,1}, p'_{1,-1,-1}, p'_{1,-1,1}, p'_{1,1,-1},$  and  $p'_{1,1,1}$  are real numbers (not necessarily non-negative) such that

1.  $p'_{-1,-1,-1} + p'_{-1,-1,1} + p'_{-1,1,-1} + p'_{-1,1,1} + p'_{1,-1,-1} + p'_{1,-1,1} + p'_{1,1,-1} + p'_{1,1,1} = 1$
2.  $\mathbf{v}'_1 \cdot \mathbf{v}_0 = p'_{-1,-1,-1} + p'_{-1,-1,1} + p'_{-1,1,-1} + p'_{-1,1,1} - p'_{1,-1,-1} - p'_{1,-1,1} - p'_{1,1,-1} - p'_{1,1,1}$ .
3.  $\mathbf{v}'_2 \cdot \mathbf{v}_0 = p'_{-1,1,-1} + p'_{-1,1,1} + p'_{1,1,-1} + p'_{1,1,1} - p'_{-1,-1,-1} - p'_{-1,-1,1} - p'_{1,-1,-1} - p'_{1,-1,1}$ .
4.  $\mathbf{v}'_3 \cdot \mathbf{v}_0 = p'_{-1,-1,1} + p'_{-1,1,1} + p'_{1,-1,1} + p'_{1,1,1} - p'_{-1,-1,-1} - p'_{-1,1,-1} - p'_{1,-1,-1} - p'_{1,1,-1}$ .
5.  $\mathbf{v}'_1 \cdot \mathbf{v}'_2 = p'_{-1,-1,-1} + p'_{-1,-1,1} + p'_{-1,1,-1} + p'_{-1,1,1} - p'_{1,1,-1} - p'_{1,1,1} - p'_{1,-1,-1} - p'_{1,-1,1}$ .
6.  $\mathbf{v}'_1 \cdot \mathbf{v}'_3 = p'_{-1,-1,-1} + p'_{-1,1,-1} + p'_{1,-1,1} + p'_{1,1,1} - p'_{-1,-1,1} - p'_{-1,1,1} - p'_{1,-1,-1} - p'_{1,1,-1}$ .
7.  $\mathbf{v}'_2 \cdot \mathbf{v}'_3 = p'_{-1,-1,-1} + p'_{1,-1,-1} + p'_{-1,1,1} + p'_{1,1,1} - p'_{-1,-1,1} - p'_{1,-1,1} - p'_{-1,1,-1} - p'_{1,1,-1}$ .

then if we take  $\mathbf{v}''_1 = \sqrt{1 - \epsilon'^2} \mathbf{v}'_1 + \epsilon' \mathbf{z}_1,$   $\mathbf{v}''_2 = \sqrt{1 - \epsilon'^2} \mathbf{v}'_2 + \epsilon' \mathbf{z}_2,$  and  $\mathbf{v}''_3 = \sqrt{1 - \epsilon'^2} \mathbf{v}'_3 + \epsilon' \mathbf{z}_3$  where  $\mathbf{z}_1, \mathbf{z}_2,$  and  $\mathbf{z}_3$  are orthogonal to  $\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3,$  and each other then if we take  $p''_{-1,-1,-1} = (1 - \epsilon')p'_{-1,-1,-1} + \frac{\epsilon'}{8},$   $p''_{-1,-1,1} = (1 - \epsilon')p'_{-1,-1,1} + \frac{\epsilon'}{8},$   $p''_{-1,1,-1} = (1 - \epsilon')p'_{-1,1,-1} + \frac{\epsilon'}{8},$   $p''_{-1,1,1} = (1 - \epsilon')p'_{-1,1,1} + \frac{\epsilon'}{8},$   $p''_{1,-1,-1} = (1 - \epsilon')p'_{1,-1,-1} + \frac{\epsilon'}{8},$   $p''_{1,-1,1} = (1 - \epsilon')p'_{1,-1,1} + \frac{\epsilon'}{8},$   $p''_{1,1,-1} = (1 - \epsilon')p'_{1,1,-1} + \frac{\epsilon'}{8},$  and  $p''_{1,1,1} = (1 - \epsilon')p'_{1,1,1} + \frac{\epsilon'}{8}$  then

1.  $p''_{-1,-1,-1} + p''_{-1,-1,1} + p''_{-1,1,-1} + p''_{-1,1,1} + p''_{1,-1,-1} + p''_{1,-1,1} + p''_{1,1,-1} + p''_{1,1,1} = 1$
2.  $\mathbf{v}''_1 \cdot \mathbf{v}_0 = p''_{-1,-1,-1} + p''_{-1,-1,1} + p''_{-1,1,-1} + p''_{-1,1,1} - p''_{1,-1,-1} - p''_{1,-1,1} - p''_{1,1,-1} - p''_{1,1,1}$ .
3.  $\mathbf{v}''_2 \cdot \mathbf{v}_0 = p''_{-1,1,-1} + p''_{-1,1,1} + p''_{1,1,-1} + p''_{1,1,1} - p''_{-1,-1,-1} - p''_{-1,-1,1} - p''_{1,-1,-1} - p''_{1,-1,1}$ .
4.  $\mathbf{v}''_3 \cdot \mathbf{v}_0 = p''_{-1,-1,1} + p''_{-1,1,1} + p''_{1,-1,1} + p''_{1,1,1} - p''_{-1,-1,-1} - p''_{-1,1,-1} - p''_{1,-1,-1} - p''_{1,1,-1}$ .
5.  $\mathbf{v}''_1 \cdot \mathbf{v}''_2 = p''_{-1,-1,-1} + p''_{-1,-1,1} + p''_{-1,1,-1} + p''_{-1,1,1} - p''_{1,1,-1} - p''_{1,1,1} - p''_{1,-1,-1} - p''_{1,-1,1}$ .
6.  $\mathbf{v}''_1 \cdot \mathbf{v}''_3 = p''_{-1,-1,-1} + p''_{-1,1,-1} + p''_{1,-1,1} + p''_{1,1,1} - p''_{-1,-1,1} - p''_{-1,1,1} - p''_{1,-1,-1} - p''_{1,1,-1}$ .
7.  $\mathbf{v}''_2 \cdot \mathbf{v}''_3 = p''_{-1,-1,-1} + p''_{1,-1,-1} + p''_{-1,1,1} + p''_{1,1,1} - p''_{-1,-1,1} - p''_{1,-1,1} - p''_{-1,1,-1} - p''_{1,1,-1}$ .

*Proof.* This follows from the following observations:

1. For all  $i \in [3], \mathbf{v}''_i \cdot \mathbf{v}_0 = (1 - \epsilon') \mathbf{v}'_i \cdot \mathbf{v}_0.$
2. For all distinct  $i, j \in [3], \mathbf{v}''_i \cdot \mathbf{v}''_j = (1 - \epsilon') \mathbf{v}'_i \cdot \mathbf{v}'_j.$
3. In each of the above sums, the  $\frac{\epsilon'}{8}$  terms cancel out as assigning probability  $\frac{\epsilon'}{8}$  to each of the possible assignments does not contribute anything to the biases or pairwise biases.  $\square$

*Proof of the first statement of Theorem 3.7:* To prove the first statement of Theorem 3.7, we observe that for each triple of vectors  $\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3$  sampled from  $T'_x$  or  $T'_{-x}$  that is not discarded, there is a triple of unit vectors  $\mathbf{v}_1, \mathbf{v}_2,$  and  $\mathbf{v}_3$  such that  $\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 = \pm \mathbf{v}_0$  and for all  $i \in [3], \|\frac{\mathbf{v}'_i}{\|\mathbf{v}'_i\|} - \mathbf{v}_i\| \leq \frac{\epsilon}{100}.$

To see this, observe that when we sample  $\mathbf{w}$  and  $\mathbf{u}$  from  $\mathcal{N}(0, 1/d)^d,$  we can take

1.  $\mathbf{v}_1 = bx\mathbf{v}_0 + \sqrt{1-x^2} \left( -\frac{\sqrt{1-(2|x|-1)^2}}{2\sqrt{1-x^2}} \frac{\mathbf{w}}{\|\mathbf{w}\|} + \sqrt{1 - \frac{1-(2|x|-1)^2}{4(1-x^2)}} \frac{\mathbf{u}}{\|\mathbf{u}\|} \right),$
2.  $\mathbf{v}_2 = bx\mathbf{v}_0 + \sqrt{1-x^2} \left( -\frac{\sqrt{1-(2|x|-1)^2}}{2\sqrt{1-x^2}} \frac{\mathbf{w}}{\|\mathbf{w}\|} - \sqrt{1 - \frac{1-(2|x|-1)^2}{4(1-x^2)}} \frac{\mathbf{u}}{\|\mathbf{u}\|} \right),$
3.  $\mathbf{v}_3 = b(1-2x)\mathbf{v}_0 + \sqrt{1-(2|x|-1)^2} \frac{\mathbf{w}}{\|\mathbf{w}\|},$

where  $b \in \{-1, 1\}$ . The adjustments we need to shift from  $\mathbf{v}_i$  to  $\frac{\mathbf{v}'_i}{\|\mathbf{v}'_i\|}$  are as follows:

1.  $\|w\|$  and  $\|u\|$  may differ from 1 by up to  $\frac{\epsilon}{1000}$ .
2. We need to shift from the vectors we obtain based on  $\mathbf{w}$  and  $\mathbf{u}$  to the centers  $\mathbf{v}'_1$ ,  $\mathbf{v}'_2$ , and  $\mathbf{v}'_3$  of the regions they are contained in.
3. We need to rescale  $\mathbf{v}'_1$ ,  $\mathbf{v}'_2$ , and  $\mathbf{v}'_3$ .

It is not hard to show that these adjustments have a total length of at most  $\frac{\epsilon}{100}$ .

A similar argument applies for the triples of vectors  $(\mathbf{v}_0, \mathbf{v}_0, \mathbf{v}')$  where  $\mathbf{v}'$  is sampled from  $A'_{-x_0}$  and the triples of vectors  $(-\mathbf{v}_0, -\mathbf{v}_0, \mathbf{v}')$  where  $\mathbf{v}'$  is sampled from  $A'_{x_0}$ . For these triples, we use  $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = (\mathbf{v}_0, \mathbf{v}_0, -\mathbf{v}_0)$  and  $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = (-\mathbf{v}_0, -\mathbf{v}_0, \mathbf{v}_0)$  respectively.

The first statement now follows by applying Lemmas A.1, A.2, and A.4. For each triple that is not discarded, letting  $\mathbf{v}_1$ ,  $\mathbf{v}_2$ , and  $\mathbf{v}_3$  be the vectors from the argument above, by Lemma A.1, there is a distribution  $D$  supported on satisfying assignments such that

1. For all  $i \in [3]$ ,  $\mathbb{E}_D[x_i] = \mathbf{v}_i \cdot \mathbf{v}_0$ .
2. For all distinct  $i, j \in [3]$ ,  $\mathbb{E}_D[x_i x_j] = \mathbf{v}_i \cdot \mathbf{v}_j$ .

By Lemma A.2, there is a pseudo-distribution  $D'$  such that

1. For all  $i \in [3]$ ,  $\mathbb{E}_{D'}[x_i] = \frac{\mathbf{v}'_i}{\|\mathbf{v}'_i\|} \cdot \mathbf{v}_0$ .
2. For all distinct  $i, j \in [3]$ ,  $\mathbb{E}_{D'}[x_i x_j] = \frac{\mathbf{v}'_i}{\|\mathbf{v}'_i\|} \cdot \frac{\mathbf{v}'_j}{\|\mathbf{v}'_j\|}$ .
3. For each assignment  $(x_1, x_2, x_3) \in \{-1, 1\}^3$ , the (pseudo-)probabilities of this assignment under  $D$  and  $D'$  differ by at most  $\frac{\epsilon}{50}$ .

By Lemma A.4, Letting  $D''$  be the probability distribution where if an assignment has pseudo-probability  $p$  under  $D'$  then it has probability  $(1 - \frac{\epsilon}{4})p + \frac{\epsilon}{32}$  under  $D''$ , we have that

1. For all  $i \in [3]$ ,  $\mathbb{E}_{D''}[x_i] = \left( \sqrt{1 - \frac{\epsilon}{4}} \frac{\mathbf{v}'_i}{\|\mathbf{v}'_i\|} + \frac{\sqrt{\epsilon}}{2} \mathbf{z}_{\mathbf{v}'_i} \right) \cdot \mathbf{v}_0$ .
2. For all distinct  $i, j \in [3]$ ,  $\mathbb{E}_{D''}[x_i x_j] = \left( \sqrt{1 - \frac{\epsilon}{4}} \frac{\mathbf{v}'_i}{\|\mathbf{v}'_i\|} + \frac{\sqrt{\epsilon}}{2} \mathbf{z}_{\mathbf{v}'_i} \right) \cdot \left( \sqrt{1 - \frac{\epsilon}{4}} \frac{\mathbf{v}'_j}{\|\mathbf{v}'_j\|} + \frac{\sqrt{\epsilon}}{2} \mathbf{z}_{\mathbf{v}'_j} \right)$ .  $\square$

**Lemma 3.9.** *The following tail bounds hold:*

1. For all  $t \in [0, 1]$ , if  $\mathbf{w} \sim \mathcal{N}(0, 1/d)^d$  then  $P(|(\|\mathbf{w}\|^2 - 1)| \geq t) \leq 2e^{-\frac{dt^2}{8}}$ .

2. Given a vector  $\mathbf{w}$ , for all  $t \geq 0$ , if  $\mathbf{u} \sim \mathcal{N}(0, 1/d)^d$  is independent of  $\mathbf{w}$  then  $P(|\mathbf{w} \cdot \mathbf{u}| \geq t) \leq 2e^{-\frac{dt^2}{2\|\mathbf{w}\|^2}}$ .

*Proof.* For the first statement, writing  $\|\mathbf{w}\|^2 - 1 = \sum_{j=1}^d (w_j^2 - \frac{1}{d})$  where each  $w_j$  is drawn independently from  $N(0, \frac{1}{d})$ , since

$$\mathbb{E} \left[ e^{\beta(w_j^2 - \frac{1}{d})} \right] = \int_{-\infty}^{\infty} \frac{\sqrt{d}}{\sqrt{2\pi}} e^{-\frac{dw^2}{2}} e^{\beta(w^2 - \frac{1}{d})} dw = \sqrt{\frac{d}{d-2\beta}} e^{-\frac{\beta}{d}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \sqrt{\frac{d}{d-2\beta}} e^{-\frac{\beta}{d}}$$

whenever  $|\beta| < \frac{d}{2}$ , if we let  $\beta' = \frac{\beta}{d}$  then we have that

$$P \left( \sum_{j=1}^d \left( w_j^2 - \frac{1}{d} \right) \geq t \right) \leq \frac{\mathbb{E} \left[ e^{\beta \sum_{j=1}^d (w_j^2 - \frac{1}{d})} \right]}{e^{\beta t}} \leq e^{-d(\beta' t + \beta' + \frac{1}{2} \ln(1-2\beta'))}$$

and

$$P \left( \sum_{j=1}^d \left( w_j^2 - \frac{1}{d} \right) \leq -t \right) \leq \frac{\mathbb{E} \left[ e^{\beta \sum_{j=1}^d (w_j^2 - \frac{1}{d})} \right]}{e^{-\beta t}} \leq e^{-d(-\beta' t + \beta' + \frac{1}{2} \ln(1-2\beta'))}.$$

whenever  $|\beta'| < \frac{1}{2}$ . Since  $\ln(1-x) = -\sum_{j=1}^{\infty} \frac{x^j}{j}$ , if  $|x| \leq \frac{1}{2}$  then  $\ln(1-x) \geq -x - x^2$  so setting  $\beta' = \frac{t}{4}$ , we have that  $P \left( \sum_{j=1}^d (w_j^2 - \frac{1}{d}) \geq t \right) \leq e^{-\frac{dt^2}{8}}$ . Similarly, setting  $\beta' = -\frac{t}{4}$ , we have that  $P \left( \sum_{j=1}^d (w_j^2 - \frac{1}{d}) \leq -t \right) \leq e^{-\frac{dt^2}{8}}$ .

For the second statement, observe that for all  $\beta \geq 0$ , since  $\mathbf{w} \cdot \mathbf{u} = \sum_{j=1}^d w_j u_j$  and

$$\mathbb{E} \left[ e^{\beta w_j u_j} \right] = \int_{-\infty}^{\infty} \frac{\sqrt{d}}{\sqrt{2\pi}} e^{-\frac{dw^2}{2}} e^{\beta w_j u} du = e^{\frac{\beta^2 w_j^2}{2d}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z - \frac{\beta w_j}{\sqrt{d}})^2}{2}} dz = e^{\frac{\beta^2 w_j^2}{2d}}$$

we have that for all  $t \geq 0$ ,  $P(\mathbf{w} \cdot \mathbf{u} \geq t) \leq \frac{\mathbb{E} \left[ e^{\beta \sum_{j=1}^d w_j u_j} \right]}{e^{\beta t}} = e^{\frac{\beta^2 \|\mathbf{w}\|^2}{2d} - \beta t}$ . Plugging in  $\beta = \frac{td}{\|\mathbf{w}\|^2}$ , we obtain that  $P(\mathbf{w} \cdot \mathbf{u} \geq t) \leq 2e^{-\frac{dt^2}{2\|\mathbf{w}\|^2}}$ . By symmetry,  $P(\mathbf{w} \cdot \mathbf{u} \leq -t) \leq 2e^{-\frac{dt^2}{2\|\mathbf{w}\|^2}}$ .  $\square$

## A.2 Omitted proofs from Section 6

We shall prove the following.

**Lemma 6.17.** *For any  $\mathcal{V}, \mathcal{W} \in (\mathbb{S}_A^N)^Y$ , we have that*

(1)  $\text{Mat}(\mathcal{V}) \in K_{Y,A}$ .

(2)  $\text{Mat}(\mathcal{V}) - \text{Mat}(\mathcal{W}) \in K_{Y,A}^0$ .

*Proof.* First we prove that (1) implies (2). To see why, by (1), we have that for all  $\mathbf{a} \in A^Y$ , there exists  $\lambda_{\mathbf{a}} \in \mathbb{R}$  such that

$$\text{Mat}(\mathcal{V}) - \text{Mat}(\mathcal{W}) = \sum_{\mathbf{a} \in A^Y} \lambda_{\mathbf{a}} \text{Mat}(\mathbf{a}). \quad (41)$$

Observe, using (28), that  $\text{Tr}(\text{Mat}(\mathcal{V})) = \text{Tr}(\text{Mat}(\mathcal{W})) = \text{Tr}(\text{Mat}(\mathbf{a})) = |Y|$  for all  $\mathbf{a} \in A^Y$ . Thus, the trace of (41) implies that  $\sum_{\mathbf{a} \in A^Y} \lambda_{\mathbf{a}} = 0$ . Hence,  $\text{Mat}(\mathcal{V}) - \text{Mat}(\mathcal{W}) \in K_{Y,A}^0$ , as desired.

We next focus on proving (1). We start by giving a self-contained proof, and then we show how the same result can be derived from a result on the projection of hypermatrices.

Let  $K_{Y,A}^* : \mathbb{R}^{(Y \times A)^2} \rightarrow \mathbb{R}$  be the dual linear space of  $K_{Y,A}$ . It suffices to show for every  $M \in K_{Y,A}^*$  that  $M(\text{Mat}(\mathcal{V})) = 0$ . We can parameterize each element  $M \in K_{Y,A}^*$  as a list of  $|Y|^2$  matrices  $(M^{y,y'} \in \mathbb{R}^{A \times A} : y, y' \in Y)$  such that, for all  $\mathbf{a} \in A$ , we have that the condition  $M(\text{Mat}(\mathbf{a}))$  is equivalent to

$$\sum_{y,y' \in Y} M_{a_y, a_{y'}}^{y,y'} = 0. \quad (42)$$

We now see to generalize (42) as follows.

**Claim A.6.** *For any list of vectors  $(\mathbf{u}_y \in \mathbb{R}^A : y \in Y)$  for which there exists  $c \in \mathbb{R}$  such that for all  $y \in Y$ ,  $\sum_{a \in A} u_{y,a} = c$ , then*

$$\sum_{\substack{y \in Y \\ a \in A}} M_{a_y, a}^{y,y'} u_{y,a} + \sum_{y \neq y' \in Y} \mathbf{u}_y^T M^{y,y'} \mathbf{u}_{y'} = 0. \quad (43)$$

*Proof.* Assume for now that  $c = 1$ . For each  $a \in A$ , let  $\mathbf{e}_a \in \mathbb{R}^A$  be the vector such that  $\mathbf{e}_{a,a'} = \mathbf{1}[a = a']$  for all  $a' \in A$ . Then, for all  $\mathbf{a} \in A^Y$ , (42) can be reinterpreted as

$$\sum_{y \in Y} M_{a_y, a_y}^{y,y} + \sum_{y \neq y' \in Y} \mathbf{e}_{a_y}^T M^{y,y'} \mathbf{e}_{a_{y'}} = 0.$$

For each  $y \in Y$ , observe that  $\mathbf{u}_y = \sum_{a \in A} u_{y,a} \mathbf{e}_a$ . Now, for all  $\mathbf{a} \in A^Y$ , let  $\lambda_{\mathbf{a}} = \prod_{y \in Y} u_{y, a_y}$ . Then, we can see that

$$\begin{aligned} 0 &= \sum_{\mathbf{a} \in A^Y} \prod_{z \in Y} u_{z, a_z} \left( \sum_{y \in Y} M_{a_y, a_y}^{y,y} + \sum_{y \neq y' \in Y} \mathbf{e}_{a_y}^T M^{y,y'} \mathbf{e}_{a_{y'}} \right) \\ &= \sum_{\substack{y \in Y \\ a \in A}} M_{a_y, a}^{y,y} \sum_{\substack{\mathbf{a} \in A^Y \\ a_y = a}} \prod_{z \in Y} u_{z, a_z} + \sum_{\substack{y \neq y' \in Y \\ a, a' \in A}} \mathbf{e}_a^T M^{y,y'} \mathbf{e}_{a'} \sum_{\substack{\mathbf{a} \in A^Y \\ a_y = a \\ a_{y'} = a'}} \prod_{z \in Y} u_{z, a_z} \\ &= \sum_{\substack{y \in Y \\ a \in A}} M_{a_y, a}^{y,y} u_{y,a} \cdot \prod_{z \in Y \setminus \{y\}} \sum_{a_z \in A} u_{z, a_z} + \sum_{\substack{y \neq y' \in Y \\ a, a' \in A}} \mathbf{e}_a^T M^{y,y'} \mathbf{e}_{a'} \cdot u_{y,a} u_{y', a'} \cdot \prod_{z \in Y \setminus \{y, y'\}} \sum_{a_z \in A} u_{z, a_z} \\ &= \sum_{\substack{y \in Y \\ a \in A}} M_{a_y, a}^{y,y} u_{y,a} + \sum_{\substack{y \neq y' \in Y \\ a, a' \in A}} \mathbf{e}_a^T M^{y,y'} \mathbf{e}_{a'} \cdot u_{y,a} u_{y', a'} \quad (c = 1) \\ &= \sum_{\substack{y \in Y \\ a \in A}} M_{a_y, a}^{y,y'} u_{y,a} + \sum_{y \neq y' \in Y} \mathbf{u}_y^T M^{y,y'} \mathbf{u}_{y'}, \end{aligned}$$

as desired. Now consider  $c \neq 1$ . If  $c \neq 0$  we can rescale  $(\mathbf{u}_y \in \mathbb{R}^A : y \in Y)$  so that  $c = 1$ . If  $c = 0$ , we can construct a sequence of sequences  $\{(\mathbf{u}_y^{(i)} \in \mathbb{R}^A : y \in Y) : i \in \mathbb{N}\}$  approaching  $(\mathbf{u}_y : y \in Y)$  in the  $\ell_2$  norm, but each  $c_i := \sum_{a \in A} u_{y,a}^{(i)}$  is nonzero. Then, we get (43) in the limit.  $\square$

For all  $i \in [N]$  and  $y \in Y$ , let  $\mathbf{z}^{i,y} \in \mathbb{R}^A$  satisfying  $z_a^{i,y} = v_{y,a,i}$ , where we recall that  $\mathbf{v}_{y,a}$  is the  $a$ -th vector defining  $\mathbf{V}_y \in \mathbb{S}_A^N$ . Since  $\sum_{a \in A} \mathbf{v}_{y,a} = \mathbf{v}_0$  for all  $y \in Y$ , we have for all  $y \in Y$  and  $i \in [N]$  that  $\sum_{a \in A} z_a^{i,y} = v_{0,i}$ . Assume that  $v_{0,i} \neq 0$  for all  $i \in [N]$ . Then, apply Claim A.6 with  $(z^{i,y}/v_{0,i} : y \in Y)$  for all  $i \in [N]$  to get that

$$\begin{aligned}
0 &= \sum_{i=1}^N v_{i,0}^2 \left( \sum_{\substack{y \in Y \\ a \in A}} M_{a,a}^{y,y'} \frac{z_a^{i,y}}{v_{0,i}} + \sum_{y \neq y' \in Y} \frac{1}{v_{i,0}^2} (\mathbf{z}^{i,y})^T M^{y,y'} \mathbf{z}^{i,y'} \right) \\
&= \sum_{\substack{y \in Y \\ a \in A}} M_{a,a}^{y,y'} \sum_{i=1}^N v_{y,a,i} v_{0,i} + \sum_{y \neq y' \in Y} \sum_{i=1}^N (\mathbf{z}^{i,y})^T M^{y,y'} \mathbf{z}^{i,y'} \\
&= \sum_{\substack{y \in Y \\ a \in A}} M_{a,a}^{y,y'} \langle \mathbf{v}_{y,a}, \mathbf{v}_0 \rangle + \sum_{y \neq y' \in Y} \sum_{a, a' \in A} M_{a,a'}^{y,y'} \sum_{i=1}^N v_{y,a,i} v_{y',a',i} \\
&= \sum_{\substack{y \in Y \\ a \in A}} M_{a,a}^{y,y'} \langle \mathbf{v}_{y,a}, \mathbf{v}_{y,a} \rangle + \sum_{y \neq y' \in Y} \sum_{a, a' \in A} M_{a,a'}^{y,y'} \langle \mathbf{v}_{y,a}, \mathbf{v}_{y',a'} \rangle \\
&= \sum_{y, y' \in Y} \sum_{a, a' \in A} M_{a,a'}^{y,y'} \langle \mathbf{v}_{y,a}, \mathbf{v}_{y',a'} \rangle \\
&= M(\text{Mat}(\mathcal{V})),
\end{aligned}$$

as desired, so  $\mathcal{V} \in K_{Y,A}$ . □

*Alternative proof of Lemma 6.17 (1).* We shall use a result on the projection of hypermatrices proved in [CŽ23a]. For  $n, q \in \mathbb{N}$ , consider a set  $\mathcal{M} = \{M^{y,y'} : 1 \leq y < y' \leq q\}$  of  $n \times n$  matrices with integral entries. We say that  $\mathcal{M}$  is a *realisable* system if the matrices of  $\mathcal{M}$  are the 2-dimensional projections of a unique hypermatrix; formally, if there exists a hypermatrix  $\Lambda \in \mathbb{Z}^{\underbrace{n \times n \times \dots \times n}_q}$  such that

$$M_{a,b}^{y,y'} = \sum_{\substack{\mathbf{z} \in [n]^q \\ z_y = a, z_{y'} = b}} \Lambda_{\mathbf{z}}$$

for each  $M^{y,y'} \in \mathcal{M}$  and each  $a, b \in [n]$ . We say that  $\mathcal{M}$  is a *realistic* system if its matrices are locally compatible in the following sense: For each  $1 \leq y_1 < y_2 \leq q$ , each  $1 \leq w_1 < w_2 \leq q$ , and each  $r, s \in [2]$  such that  $y_r = w_r$ , it holds that

$$(M^{y_1, y_2})^{or} \mathbf{1} = (M^{w_1, w_2})^{os} \mathbf{1},$$

where we use the notation  $M^{\circ 1} := M$  and  $M^{\circ 2} := M^\top$ . We will use the following result.

**Theorem A.7** ([CŽ23a]). *A system  $\mathcal{M}$  is realistic if and only if it is realisable.*<sup>18</sup>

<sup>18</sup>This result was proved in [CŽ23a] in the case of arbitrarily dimensional projections and hypermatrices that are not necessarily cubic (i.e., whose modes have possibly distinct lengths). Here we only need the restricted case described above.

Take now some global configuration  $\mathcal{V} \in (\mathbb{S}_A^N)^Y$ , and let  $\text{Mat}(\mathcal{V}) \in \mathbb{R}^{(Y \times A)^2}$  be the corresponding Gram matrix. Suppose first that  $\text{Mat}(\mathcal{V})$  has rational entries, and let  $N$  be a common denominator. Denote by  $M^{y,y'}$  the  $(y, y')$ -th block of  $N \cdot \text{Mat}(\mathcal{V})$  (of size  $A \times A$ ), and observe that the quantity

$$(M^{y,y'} \mathbf{1})_a = N \sum_{b \in A} \langle \mathbf{v}_{y,a}, \mathbf{v}_{y',b} \rangle = N \langle \mathbf{v}_{y,a}, \mathbf{v}_0 \rangle$$

is independent on  $y'$  for each  $a \in A$ . Hence, the system  $\mathcal{M} = \{M^{y,y'} : 1 \leq y < y' \leq |Y|\}$  is realistic. By Theorem A.7, it follows that  $\mathcal{M}$  is realisable. Let  $\Lambda$  be its realisation hypermatrix. We claim that  $\frac{1}{N}\Lambda$  witnesses that  $\text{Mat}(\mathcal{V})$  lies in  $K_{Y,A}$ . I.e., we claim that  $\text{Mat}(\mathcal{V}) = \sum_{\mathbf{z} \in A^Y} \frac{1}{N} \Lambda_{\mathbf{z}} \text{Mat}(\mathbf{z})$ . Indeed, for each  $1 \leq y < y' \leq |Y|$  and each  $a, b \in A$  we have

$$N \cdot \text{Mat}(\mathcal{V})_{(y,a),(y',b)} = M_{a,b}^{y,y'} = \sum_{\substack{\mathbf{z} \in A^{|Y|} \\ z_y = a, z_{y'} = b}} \Lambda_{\mathbf{z}} = \sum_{\mathbf{z} \in A^{|Y|}} \Lambda_{\mathbf{z}} \text{Mat}(\mathbf{z})_{(y,a),(y',b)} = \left( \sum_{\mathbf{z} \in A^{|Y|}} \Lambda_{\mathbf{z}} \text{Mat}(\mathbf{z}) \right)_{(y,a),(y',b)}$$

as needed. Since  $M^{y',y} = (M^{y,y'})^\top$  and since the diagonal blocks of the Gram matrix of any global configuration are completely determined by its off-diagonal blocks, the result holds for each  $y, y' \in Y$ , thus proving the claim. Finally, the rational subset of Gram matrices of global configurations in  $(\mathbb{S}_A^N)^Y$  is dense, so the same result must hold for non-rational  $\text{Mat}(\mathcal{V})$  as well, thus concluding the proof.  $\square$