A new point of NP-hardness for Unique Games

Ryan O'Donnell* John Wright†

September 30, 2012

Abstract

We show that distinguishing $\frac{1}{2}$-satisfiable Unique-Games instances from $(\frac{3}{8} + \epsilon)$-satisfiable instances is NP-hard (for all $\epsilon > 0$). A consequence is that we match or improve the best known c vs. s NP-hardness result for Unique-Games for all values of c (except for c very close to 0). For these c, ours is the first hardness result showing that it helps to take the alphabet size larger than 2. Our NP-hardness reductions are quasilinear-size and thus show nearly full exponential time is required, assuming the ETH.

*Department of Computer Science, Carnegie Mellon University. Supported by NSF grants CCF-0747250 and CCF-0915893, and by a Sloan fellowship. Some of this research performed while the author was a von Neumann Fellow at the Institute for Advanced Study, supported by NSF grants DMS-083537 and CCF-0832797.

†Department of Computer Science, Carnegie Mellon University.

1 Introduction

Thanks largely to the groundbreaking work of Håstad [Hås01], we have optimal NP-hardness of approximation results for several constraint satisfaction problems (CSPs), including 3Lin(Z_2) and 3Sat. But for many others — including most interesting CSPs with 2-variable constraints — we lack matching algorithmic and NP-hardness results. Take the 2Lin(Z_2) problem for example, in which there are Boolean variables with constraints of the form "$x_i = x_j$" and "$x_i \neq x_j$". The largest approximation ratio known to be achievable in polynomial time is roughly .878 [GW95], whereas it is only known that achieving ratios above $\frac{11}{12} \approx .917$ is NP-hard [Hås01, TSSW00]. Which of these two results can be improved?

In the early 2000s there was some sentiment that the .878-ratio approximation algorithm could be improved [Fei99, Fei02]. However this began to change after Khot's introduction of the Unique-Games (UG) Conjecture [Kho02]. In that paper, Khot showed that $(1-\epsilon, 1-\epsilon^{1/2+o(1)})$-approximating 2Lin($Z_2$) is "UG-hard" for all $\epsilon > 0$. Assuming the UG Conjecture this would be essentially optimal because the Goemans–Williamson algorithm [GW95] efficiently $(1-\epsilon, 1-O(\epsilon^{1/2}))$-approximates 2Lin($Z_2$). Here we are using the following terminology:

Definition 1.1. A (c, s)-*approximation algorithm* for a maximization CSP is one which satisfies at least an s-fraction of the constraints on any instance where the optimal solution satisfies at least a c-fraction of the constraints.

This was subsequently extended [KKMO07, KO09, OW08] to give matching $(c, s(c))$-approximation algorithms and UG-hardness results for 2Lin(Z_2) for every $c \in [\frac{1}{2}, 1]$, including UG-hardness of achieving $\frac{s}{c} > .878$. In light of such sharp UG-hardness results — shown for *every* CSP by Raghavendra [Rag09] — it seemed in the late '00s that the boundary of efficient approximability had been identified.

However the pendulum swung again in 2010 with the Arora–Barak–Steurer subexponential-time algorithm for Unique-Games [ABS10] (see also the related works [Kol10, Ste10, BRS11, GS11]). For example, we now know there is a universal constant $\epsilon_0 > 0$ such that in time $2^{O(n^{.001})}$ one can both $(1-\epsilon_0, \frac{1}{2})$-approximate Unique-Games ([ABS10]) and $(\frac{1}{2}, \epsilon_0)$-approximate Unique-Games (this follows [Ste11] from [Ste10, BRS11]). There is also now speculation that approximating Max-Cut or 2Lin(Z_2) to factor .879 may be possible in subexponential time. Yet all of these problems are predicted to NP-hard by the UG Conjecture.

This raises the question of what meaning an NP-hardness result actually has. For example, approximating Max-Clique to factor $1/n^{.999}$ is known to be NP-hard [Hås99], yet it's trivially solvable in $n^{n^{.001}+.002}$ time. Such a running time is completely practical for any plausible value of n. If one could make the above-mentioned $2^{O(n^{.001})}$-time algorithms for Unique-Games similarly practical, one might argue that this "morally" disproves the conjecture that $(1-\epsilon, \epsilon)$-approximating UG is hard for every $\epsilon > 0$. In any case, these theoretical results correspond well with the observation that "in practice", decently approximating Unique-Games does not seem to be very hard. In particular, there is no known family of very "hard-seeming" instances for the UG Conjecture, as there is for, say, the 3Sat decision problem.

On the other hand, we do have evidence of extreme hardness for at least *some* gapped approximation version of Unique-Games. For example, Håstad's 1997 work [Hås97] implies that $(\frac{1}{2}, .459)$-approximating Unique-Games is NP-hard (here $.459 \approx \frac{11}{12} \cdot \frac{1}{2}$). The proof is a local gadget reduction from his result on NP-hardness of $(1-\epsilon, 1/2+\epsilon)$-approximating 3Lin($Z_2$). Furthermore, Moshkovitz and Raz [MR10] have shown that the 3Lin(Z_2) result holds under a *quasilinear-size* reduction from 3Sat. Thus assuming the Exponential Time Hypothesis (ETH) [IP01] that deciding 3Sat requires $2^{\Omega(n)}$ time, it follows that $(\frac{1}{2}, .459)$-approximating Unique-Games is truly hard,

requiring essentially full exponential time $2^{n^{1-o(1)}}$. Even if we don't assume the ETH, $(1-\epsilon, \frac{1}{2}+\epsilon)$-approximating $3\mathsf{Lin}(Z_2)$ is a problem for which we can easily generate very hard-in-practice instances (as cryptographic research on the Learning Parity with Noise problem has shown). Applying the local gadget reduction to these instances shows that the same is true of $(\frac{1}{2}, .459)$-approximating Unique-Games.

1.1 Our main result

To recap, for instances of the Unique-Games problem in which the optimal solution satisfies $\frac{1}{2}$ of the constraints, we know that satisfying a certain constant ϵ_0 fraction of the constraints is relatively "easy" (time $2^{O(n^{.001})}$), whereas satisfying the larger constant .459 fraction is "hard" (time $2^{n^{1-o(1)}}$ assuming ETH). Thus, as is often the case in the field of approximation algorithms, we are faced with the task of pinning down the truth between two constants. The main theorem of this paper is some progress on the hardness side:

Main Theorem. *For any constant label size q, $(\frac{1}{2}, \frac{3}{8}+\frac{1}{q^{\Theta(1)}})$-approximating Unique-Games is* NP*-hard under a quasilinear-size reduction; hence the problem requires time $2^{n^{1-o(1)}}$ under the ETH.*

(In fact, our theorem holds for $2\mathsf{Lin}(Z_q)$, $q \leq \mathrm{poly}(\log\log\log n)$.)

Although we would certainly not want to conjecture it, we have to at least raise the possibility that the optimal subexponential-time algorithm for $\frac{1}{2}$-satisfiable Unique-Games instances satisfies $\frac{3}{8}$ of the constraints.

Considerations of subexponential time vs. full exponential time aside, our result is just the second improved NP-hardness result for Unique-Games since 1997. Via trivial reductions, our Main Theorem extends to give NP-hardness of $(c, \frac{3}{4}c + o(1))$-approximating Unique-Games (for $c \leq \frac{1}{2}$) and also of $(c, 1-\frac{5}{4}(1-c)+o(1))$-approximating Unique-Games (for $c \geq \frac{1}{2}$). For all but very small $c \in (0,1)$ this subsumes or improves the best previous result, due to Håstad in 1997 [Hås97]. This best previous result involved taking $q = 2$; our result shows that hardness increases as q increases.

For $c \leq \kappa$, where κ is a small (inexplicit) positive constant, Feige–Reichman 2004 [FR04] has the best (c, s)-inapproximability result for Unique-Games. See Section 2.1 for more detailed comparison with prior work.

1.2 Our approach, and other contributions

More broadly, this paper focuses on trying to obtain unconditional NP-hardness of approximation results for 2-variable CSPs such as Unique-Games. With a few exceptions, the best such results known are derived by gadget reductions from Håstad's $(1-\epsilon, \frac{1}{2}+\epsilon)$-approximation NP-hardness for $3\mathsf{Lin}(Z_2)$. Indeed, for the well-known Boolean 2-CSPs Max-Cut, $2\mathsf{Lin}(Z_2)$, 2Sat, and 2And, the best gadgets were found via computer solution of large linear programs [TSSW00]. This state of affairs is unsatisfactory for a few reasons. First, there is almost no intuition for these computer-generated gadgets. Second, the doubly-exponential size of the linear programs involved makes it infeasible to use the computer-search approach even for 2-CSPs over a ternary domain. Third, since the $(1-\epsilon, \frac{1}{2}+\epsilon)$-hardness for $3\mathsf{Lin}(Z_2)$ is itself a (highly sophisticated) gadget reduction from Label-Cover, these kinds of results are arguably using an artificial "middleman" that could be cut out.

It makes sense then to seek direct reductions from Label-Cover to approximation of 2-CSPs. This has never been done in the "Håstad style" (Long Codes and Fourier analysis) with 2-CSPs

before since it's unclear how to "get a square into the Fourier analysis". In fact we managed to reproduce the $(\frac{3}{4}, \frac{11}{16}+\epsilon)$-NP-hardness result for 2Lin(Z_2) via a Håstad-style analysis (see Appendix F), but for the Unique-Games problem we required a more conceptual approach.

This new conceptual approach for reducing Label-Cover to 2-CSPs has three components:

1. Given a Label-Cover instance (V, E), the usual Håstad reduction methodology introduces a "prover" f_u for each vertex $u \in V$ (also known as a table, function, or collection of auxiliary variables), and replaces each constraint on $(u, v) \in E$ with a distribution on "questions" (constraints) for f_u and f_v. In our approach we also introduce a prover h_{uv} for each constraint $(u, v) \in E$.[1] From this constraint we propose generating 2-CSP questions as follows: First, generate question pairs (x, y) from a product distribution. Next, "corrupt" x to $\widetilde{x}$ and y to $\widetilde{y}$ in some *correlated* random way. Finally, send prover h_{uv} the pair $(\widetilde{x}, \widetilde{y})$, and test its answer against a random choice of $f_u(x)$ or $f_v(y)$.

2. We next develop the Invariance Principle technology [MOO10, DMR09, Mos10, Rag09] to show that if the "Håstad decoding procedure" applied to f_u and f_v fails, then the analysis surrounding f_u, f_v, and h_{uv} can be performed as though the corruptions $x \to \widetilde{x}$ and $y \to \widetilde{y}$ were *independent*. This seems to be the first published example of using Invariance to analyze reductions from Label-Cover, as opposed to from Unique-Games or the d-to-1 Conjecture.[2]

3. Given the Invariance result, all concerns involving Fourier analysis and computational complexity are eliminated. Analyzing any proposed "test" reduces to analyzing a purely information-theoretic problem of *non-interactive correlation distillation* (NICD) type (see, e.g., [Yan07]). Such a task can still be difficult; e.g., to obtain the best known NP-hardness even for 2Lin(Z_2) we needed to resolve a 2004 NICD conjecture of Yang [Yan04]. Still, the fact that we are reduced to information-theoretic problems makes things clean enough that we can obtain the Main Theorem for Unique-Games.

Our new approach lets us recover in a conceptual way the best known NP-hardness results [Hås97, TSSW00] for Max-Cut, 2Lin(Z_2), 2Sat, and 2And.

2 Preliminaries

We consider weighted CSPs. An instance $\mathcal{I}$ consists of a set of variables over a finite domain (of "labels"), along with a weighted list of constraints on these variables. We assume the weights are nonnegative rationals summing to 1, so we can also think of the instance as a probability distribution on constraints. We usually write n for the number of variables. The *size* of an instance is the total number of bits needed to specify the constraint relations and the weights; we will only consider instances which have size $n^{1+o(1)}$.[3] Given an assignment F to the variables we write $\mathrm{Val}_{\mathcal{I}}(F)$ for the total weight of the constraints it satisfies; we also write $\mathrm{Opt}(\mathcal{I}) = \max_F\{\mathrm{Val}_{\mathcal{I}}(F)\}$. The hardness of approximation results we prove in this paper will hold even for the problem of (c, s)-*deciding* the CSP; i.e., outputting 'YES' when $\mathrm{Opt}(\mathcal{I}) \geq c$ and outputting 'NO' when $\mathrm{Opt}(\mathcal{I}) < s$.

[1] This idea is certainly not new. Indeed, for Long Code-based reductions from Label-Cover with projection constraints it is known that such provers never *need* to be introduced, though conceptually it may help to do so. Our main reduction from Label-Cover does not actually assume projection constraints, so we *do* need to add these provers.

[2] It was known to some experts [WZ10] that Invariance techniques can be used to analyze Håstad's Label-Cover to 3Lin(Z_q) reduction, though the analysis essentially degenerates to the Håstad style used in [OWZ11].

[3] In this paper we will not concern ourselves with the bit-representation size of the rational numbers giving the weights; the reader may check that this can be accounted for without changing the statements of our theorems.

CSPs are distinguished by the domain of the variables and the kinds of constraints allowed. We mostly consider the case when the domain is Z_q, the additive group of integers modulo q, for some q.

Definition 2.1. Let ϕ be a predicate on Z_q^k. Then Max-ϕ denotes the CSP where the variables have domain Z_q and the constraints are ϕ applied to various k-tuples of variables. Max-ϕ^+ denotes the generalization where ϕ is applied to k-tuples of *literals*; by a literal we mean $x + c \pmod q$ where x is a variable and c is a constant. We also extend the notation to allow collections Φ of predicates.

For $q = 2$, the familiar CSPs 2Lin(Z_2), 2Sat, 2And, Max-Cut are equivalent to Max-$=^+$, Max-$\vee^+$, Max-$\wedge^+$, Max-$\neq$, respectively. 3Lin(Z_2) is Max-ϕ^+ for $\phi(x, y, z) = x+y+z$. For general q, 2Lin(Z_q) is Max-$=^+$. A related problem is 2Lin($\mathbb{F}_q$) for q a prime power, in which general 2-variable linear equations over $\mathbb{F}_q$ are allowed. The "Unique-Games" 2-CSP UG$_q$ has variable domain Z_q, with any *bijective* constraint being allowed. We remark that 2Lin(Z_q) is a special case of UG$_q$, and for $q = 2$ the problems are in fact identical. The *Unique-Games Conjecture* of Khot [Kho02] states that for all $\epsilon > 0$ there exists q such that $(1 - \epsilon, \epsilon)$-deciding UG$_q$ is NP-hard. In [KKMO07] it is shown that the UG Conjecture implies the stronger statement that $(1 - \epsilon, q^{-\epsilon/(2-\epsilon)})$-deciding 2Lin($Z_q$) is NP-hard for all $\epsilon > 0$ and sufficiently large q.

Finally, we define a (generalization) of the "Label Cover" CSP LC$_{d_1K,d_2K}$:

Definition 2.2. The input for LC$_{d_1K,d_2K}$ is a biregular bipartite graph $((U, V), E)$; the vertices of U are variables with domain $[d_1K]$; the vertices of V are variables with domain $[d_2K]$. Also, for each edge $e = (u, v)$ there is given a d_1-to-1 map $\pi_{e,u} : [d_1K] \to [K]$ and a d_2-to-1 map $\pi_{e,v} : [d_2K] \to [K]$. The associated constraints (of equal weight) are that $\pi_{e,u}(u) = \pi_{e,v}(v)$.

The more usual definition of Label Cover is the special case of LC$_{K,dK}$ (i.e., $d_1 = 1$). Hardness results for LC$_{K,dK}$ immediately extend to LC$_{d_1K,dK}$ by duplication of labels for the U vertices.

Feige–Kilian [FK94] and Raz [Raz95] first proved that for all $\epsilon > 0$ there exists K and d such that $(1, \epsilon)$-deciding LC$_{K,dK}$ is NP-hard. We will use the following strong form of this result from [MR10] (see also [DH09]):

Moshkovitz–Raz Theorem. *For any $\epsilon = \epsilon(n) \geq n^{-o(1)}$ there exists $K, d \leq 2^{\mathrm{poly}(1/\epsilon)}$ such that the problem of deciding a* 3Sat *instance of size n can be Karp-reduced in* poly(n) *time to the problem of $(1, \epsilon)$-deciding a* LC$_{K,dK}$ *instance of size $n^{1+o(1)}$.*

For brevity, we say that $(1, \epsilon)$-deciding LC$_{K,dK}$ is NP*-hard under quasilinear-size reductions.*

2.1 Comparison with prior work

To state inapproximability results for various CSPs, let us make a definition (essentially from [OW08]):

Definition 2.3. For a given CSP we say that (c, s) is a *point of* NP*-hardness* if $(c, s + \epsilon)$-deciding the CSP is NP-hard for all $\epsilon > 0$. We may also qualify this definition by insisting on quasilinear-size reductions.

We now state some best known inapproximability results for Boolean 2-CSPs, from [Hås01, TSSW00]:

Theorem 2.4. *We have the following points of* NP*-hardness, even under quasilinear-size reductions [MR10]:* $(\frac{3}{4}, \frac{11}{16})$ *for* 2Lin(Z_2), $(\frac{11}{12}, \frac{7}{8})$ *for* 2Sat, $(\frac{5}{12}, \frac{3}{8})$ *for* 2And, *and* $(\frac{17}{21}, \frac{16}{21})$ *for* Max-Cut.

Chlebík and Chlebíková [CC04] have also shown that $(1-\delta, 1-2.889\delta)$-deciding 2Sat (and even "Non-Mixed-2Sat") is NP-hard for all $0 < \delta < \delta_0$, where δ_0 is an unspecified positive constant. Their reduction is *not* quasilinear-size, relying as it does on the alternative PCP constructions of Dinur and Safra [DS05].

Padding. Given a point of NP-hardness (c, s) for some CSP, one can trivially obtain some other points of NP-hardness by what we'll call "padding" (see Appendix D). Specifically, one can obtain any point (c', s') on the line segments (in $\mathbb{R}^2$) joining (c, s) to $(1, 1)$ and to (c_0, c_0). Here c_0 is a CSP-specific constant equal to the infimum of $\mathrm{Opt}(\mathcal{I})$ over all instances $\mathcal{I}$. E.g., $c_0 = \frac{1}{q}$ for $2\mathsf{Lin}(Z_q)$ because the "least satisfiable instance" (namely $\{x - y = 0, x - y = 1, \ldots, x - y = q - 1\}$) has $\mathrm{Opt} = \frac{1}{q}$. So for $2\mathsf{Lin}(Z_2)$ we obtain the points of NP-hardness $(1-\delta, 1-\frac{5}{4}\delta)$ for each $0 < \delta < \frac{1}{4}$ and also $(\frac{1}{2} + \frac{1}{4}\lambda, \frac{1}{2} + \frac{3}{16}\lambda)$ for each $0 < \lambda < 1$. Padding preserves quasilinearity of reduction size.

Unique-Games. We now discuss the previous best NP-hardness of approximation for the Unique-Games problem UG_q. For $q = 2$ we have the $(\frac{3}{4}, \frac{11}{16})$ point of NP-hardness from Theorem 2.4, since $2\mathsf{Lin}(Z_2)$ and UG_2 are identical. For larger q we still have the point $(\frac{3}{4}, \frac{11}{16})$ for UG_q (at least if q is even), since $2\mathsf{Lin}(Z_2)$ can be considered a subproblem of UG_q by duplicating labels. By padding this result we get the points of NP-hardness $(1-\delta, 1-\frac{5}{4}\delta)$ for $0 < \delta < \frac{1}{4}$ and also $(\lambda \cdot \frac{3}{4} + \frac{1-\lambda}{q}, \lambda \cdot \frac{11}{16} + \frac{1-\lambda}{q}) \xrightarrow{q\to\infty} (\lambda \cdot \frac{3}{4}, \lambda \cdot \frac{11}{16})$ for $0 < \lambda < 1$. Rather surprisingly, no stronger result was previously known (except for tiny c, as we will describe shortly); in particular, the explicit hardness results in [AEH01, FR04] for $2\mathsf{Lin}(\mathbb{F}_q)$ for small q are inferior. In other words, the best known NP-hardness for UG_q involved taking $q = 2$!

We now state our Main Theorem precisely:

Theorem 2.5 (Main Theorem, precise statement.)**.** *There exists* $\epsilon = \epsilon(n) = \widetilde{\Theta}(1/\log\log\log n)$ *such that for each integer* $2 \le q \le (\log\log\log n)^3$*, there is a quasilinear-size reduction from size-*n *instances of* 3Sat *to the problem of* $(\frac{1}{2} + \frac{1}{2q}, s(q) + \epsilon)$*-deciding* $2\mathsf{Lin}(Z_q)$*, where* $s(q) = \frac{3}{8} + \frac{5}{8q}$ *for* $q < 7$ *and in general* $s(q) = \frac{3}{8} + O(\frac{1}{q^{1/3}})$*.*

(We believe that one can take $s(q) = \frac{3}{8} + \frac{5}{8q}$ for all q but we haven't proved this; see Section 6.) For $q = 2$ our theorem matches the previous result; we improve upon the prior work by taking $q \to \infty$, getting a point of NP-hardness for UG_q tending to $(\frac{1}{2}, \frac{3}{8})$. By padding, this extends the previous $(1-\delta, 1-\frac{5}{4}\delta)$ result and also gives the points of hardness $(\lambda \cdot \frac{1}{2}, \lambda \cdot \frac{3}{8})$ for $0 < \lambda < 1$.

As mentioned, there is one more known NP-hardness result for UG_q, due to Feige and Reichman [FR04]. For any sufficiently large prime power q they establish that $(q^{-1+\eta}, \Theta(q^{-1}))$ is a point of NP-hardness for $2\mathsf{Lin}(\mathbb{F}_q)$; here $\eta > 0$ is an unspecified universal constant. The hardness holds under quasilinear-size reductions, using [MR10].

We illustrate the new state of (in)approximability for UG_q in Figure 1 (with q fixed slightly superconstant, for simplicity).

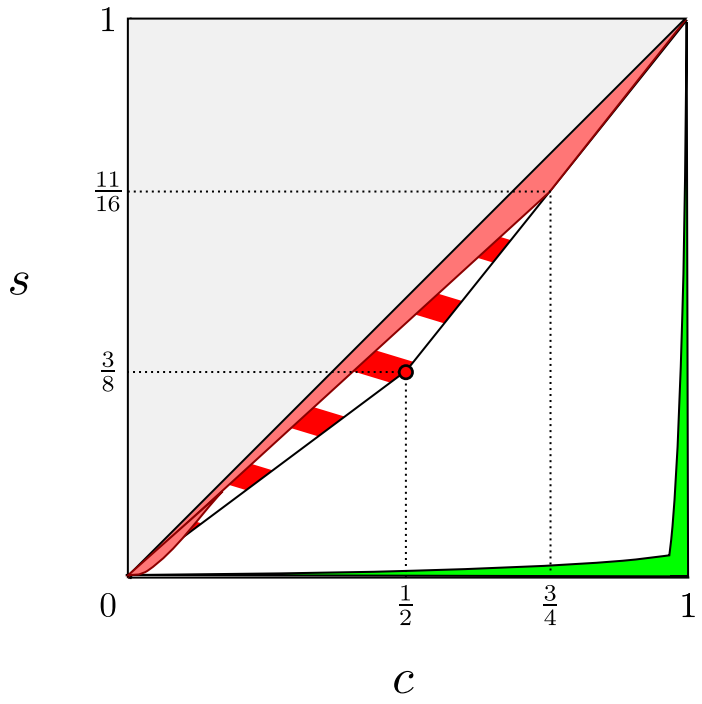

Figure 1: Approximability results for Unique-Games with label size $\omega(1) \leq q \leq \log\log n$

The solid pink regions in Figure 1 are the best previously known points (c, s) of NP-hardness. The triangle with vertex $(\frac{3}{4}, \frac{11}{16})$ (from [Hås01]) is drawn to scale; the bump near $c = 0$ (from [FR04]) is an "artistic impression". Our new hardness result is the red-striped region implied by the point of hardness at $(\frac{1}{2}, \frac{3}{8})$. These NP-hardness results hold under quasilinear-size reductions. In green we have illustrated (an artistic impression of) points (c, s) for which there is a subexponential-time (c, s)-time approximation algorithm for Unique-Games ([ABS10, Ste10, BRS11]). The true boundary between what is achievable in subexponential time for Unique-Games and what is not (assuming ETH) lies somewhere in the white region of the figure.

3 Overview of our method

In this section we introduce our method by showing how we can use it to recover Håstad's hardness result for $2\mathsf{Lin}(Z_2)$ [Hås01]. At the highest level, our method is standard: to prove inapproximability for $2\mathsf{Lin}(Z_2)$, we design a suitable matching dictators test and compose it with a Label Cover instance. However, the test we design is somewhat nonstandard; in particular, it bears more than a passing resemblance to a purely information-theoretic problem in *non-interactive correlation distillation* (NICD). In performing the soundness analysis of the test, we make this resemblance explicit: to upper-bound the soundness value of the test, it turns out to be sufficient to solve the NICD problem. Thus, we prefer to think of our method as reducing inapproximability to problems in NICD.

3.1 An example NICD problem

The NICD problem which our $2\mathsf{Lin}(Z_2)$ test resembles most is actually the "basic" NICD problem, which we introduce here. Let f be a party who receives a uniformly random q-ary string $\boldsymbol{x} \in Z_q^n$. The string $\boldsymbol{x}$ is sent to a "middleman" h along an *erasure channel* which erases each symbol

independently with probability 1/2; thus h receives some $\widetilde{\boldsymbol{x}} \in (Z_q \cup \{*\})^n$. Now f and h have some correlated information, but they are not allowed to interact further. Nevertheless, they would like to agree on a common symbol $\ell \in Z_q$. (This setup explains the name "non-interactive correlation distillation".) The "strategy" of f is just a function $f : Z_q^n \to Z_q$, and similarly the strategy of h is a function $h : (Z_q \cup \{*\})^n \to Z_q$. Thus the party and middleman together succeed with probability $\mathbf{Pr}_{\boldsymbol{x},\widetilde{\boldsymbol{x}}}[f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}})]$.

The NICD problem here is to find functions f and h which maximize this success probability. Of course, if f and h are (matching) constant functions then the success probability is 1. However such trivial solutions are disallowed by insisting the function f be (at least) *balanced*, meaning $\mathbf{Pr}_{\boldsymbol{x}}[f(\boldsymbol{x}) = \ell] = \frac{1}{q}$ for all $\ell \in Z_q$.

This erasure-channel NICD problem was first proposed by Yang [Yan04] in the boolean ($q = 2$) case. He showed that the success probability is at most $\frac{1}{2} + \frac{1}{2\sqrt{2}} \approx .85$; he also conjectured that the best upper bound is in fact $\frac{3}{4}$, the success probability achieved when f is a *dictator* (i.e. $f(x) = x_i$ for some i) and h plays optimally for this dictator. More generally, for the binary erasure channel with erasure probability $1 - \rho$, Yang bounded the success probability by $\frac{1}{2} + \frac{1}{2}\sqrt{\rho}$ but conjectured that the dictator's success probability $\frac{1}{2} + \frac{1}{2}\rho$ is optimal (at least for $\rho \geq \frac{1}{2}$). In Section 5 we prove this conjecture. Furthermore, in Section 6 we prove an analogous conjecture for the q-ary erasure channel with erasure probability $\frac{1}{2}$.

3.2 $2\mathsf{Lin}(Z_2)$ proof outline

Now we describe how the intuition behind our $2\mathsf{Lin}(Z_2)$ test. As a starting point, consider the hard instances of $2\mathsf{Lin}(Z_2)$ produced by Håstad: they begin by taking a d-to-1 Label Cover instance and replacing each vertex with its corresponding Long Code. Then, $3\mathsf{Lin}(Z_2)$ tests (equivalently, constraints) are placed between appropriately chosen Long Code vertices to produce a hard instance of the $3\mathsf{Lin}(Z_2)$ problem. Finally, the gadget from [TSSW00] is applied to the $3\mathsf{Lin}(Z_2)$ instance, resulting in a $2\mathsf{Lin}(Z_2)$ instance. The gadget works locally: given a $3\mathsf{Lin}(Z_2)$ test, it adds new vertices which are not part of any Long Code and puts in place a set of $2\mathsf{Lin}(Z_2)$ tests, each of which is performed exclusively between one vertex in the original $3\mathsf{Lin}(Z_2)$ test—a Long Code vertex—and one of the newly added vertices. We stress that the newly added vertices are unique to each $3\mathsf{Lin}(Z_2)$ test in the original $3\mathsf{Lin}(Z_2)$ instance. The result is that the final $2\mathsf{Lin}(Z_2)$ instance has groups of Long Code vertices along with clouds of vertices which sit between pairs of Long Codes and to which these Long Code pairs are compared.

Thus, it is sensible for a direct reduction from Label Cover to $2\mathsf{Lin}(Z_2)$ to take the following form: given u and v which are adjacent in the original Label Cover instance and whose Long Codes are f and g, respectively, there is a "cloud" of vertices which sits between u and v to which f and g may be compared. We think of this cloud of vertices as being labeled by some function h and indexed into by strings z from some set. A typical test will select an x for the u side, a y for the v side, and then to use the chosen x and y to select an appropriate z from the cloud. At this point, either the test $f(x) = h(z)$ is performed, or the test $g(y) = h(z)$ is performed, each with some probability. Thus, given an x and a y, the test will select a z and ask that $h(z)$ "predict" the value of $f(x)$ and $g(y)$. To help it do this, we will choose a z that contains some information about the strings x and y.

The test: Now, we give a (mostly) complete description of our hard $2\mathsf{Lin}(Z_2)$ instance. We begin with a d-to-1 Label Cover instance over the graph $G = (U \cup V, E)$. For each $u \in U$ and each $v \in V$, introduce the Long Codes $f_u : \{-1,1\}^K \to \{-1,1\}$ and $g_v : \{-1,1\}^{dK} \to \{-1,1\}$, respectively. For each edge $(u,v) \in E$, introduce the function $h_{uv} : \{-1,1,*\}^K \times \{-1,1,*\}^{dK} \to \{-1,1\}$. We

think of strings $y \in \{-1,1\}^{dK}$ as being formed of K "blocks" of size d each, so that y_1 through y_d is the first block, y_{d+1} through y_{2d} is the second block, and so forth. Denote the ith length-d block of y by $y[i] = (y_{d(i-1)+1}, \ldots, y_{d(i-1)+d})$. Now, pick an edge $(u,v) \in E$ uniformly at random, and perform the following test on f_u, g_v, and h_{uv}:

$\boxed{2\mathsf{Lin}(Z_2)\text{-}\textsc{Test}}$

- Given functions $f : \{-1,1\}^K \to \{-1,1\}$, $g : \{-1,1\}^{dK} \to \{-1,1\}$, and $h : \{-1,1,*\}^K \times \{-1,1,*\}^{dK} \to \{-1,1\}$:
- Draw $\boldsymbol{x} \in \{-1,1\}^K$ and $\boldsymbol{y} \in \{-1,1\}^{dK}$ independently and uniformly at random.
- Form "corrupted" versions of $\boldsymbol{x}$ and $\boldsymbol{y}$ as follows: for each block $i \in [K]$, with probability $1/2$ replace $\boldsymbol{x}_i$ with a $*$ and keep $\boldsymbol{y}[i]$ the same, and with probability $1/2$ replace $\boldsymbol{y}[i]$ with $*^d$ and keep $\boldsymbol{x}_i$ the same. Call the corrupted versions $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$, respectively.
- Test either $f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ or $g(\boldsymbol{y}) = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$, each with equal probability.

Note that this actually produces a $2\mathsf{Lin}(Z_2)$ hardness instance, as equality tests are $2\mathsf{Lin}(Z_2)$ tests. We think of the $*$'s as the gaps in knowledge h has about $\boldsymbol{x}$ and $\boldsymbol{y}$. The string $\widetilde{\boldsymbol{x}}$ contains about half of $\boldsymbol{x}$, and the string $\widetilde{\boldsymbol{y}}$ contains about half of $\boldsymbol{y}$. Moreover this (lack of) information is correlated: the part of $\boldsymbol{x}$ missing from $\widetilde{\boldsymbol{x}}$ is exactly the part of $\boldsymbol{y}$ *not* missing from $\widetilde{\boldsymbol{y}}$, and vice versa. This test looks like a "two-party" version of the NICD problem given in Section 3.1. One might hope that the analysis we used on that NICD problem to solve Yang's conjecture would help with analyzing this test; at first blush, unfortunately, this doesn't work because these correlations are difficult to reason about. Thus, we will need to find a way to "break" the correlations.

It is useful to talk about the probability that f and g pass the test without reference to h. Since in the $2\mathsf{Lin}(Z_2)$ hardness instance, h_{uv} is only ever referenced when tests corresponding to the edge (u,v) are performed, we may as well assume that h_{uv} is selected to perform optimally with f_u and g_v. Thus, by "the probability that f and g pass the test" (and similar phrases) we mean the probability that f, g, and an optimal h pass the test.

Analyzing the test: The correlation in what information is present or missing is extremely helpful in the "matching dictators" case, i.e. when $f(x) = x_i$ and $g(y) = (y[i])_j$, for $i \in [K], j \in [d]$. In this case, h is always given full information about either $f(\boldsymbol{x})$ or $g(\boldsymbol{y})$, because exactly one of $\boldsymbol{x}_i$ or $\boldsymbol{y}[i]$ is always kept when forming $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$. As a result, h can always predict one of $f(\boldsymbol{x})$ or $g(\boldsymbol{y})$ with perfect accuracy, whereas against the other one its success is just an unbiased coin flip. Thus, f and g pass the test with $(1 + 1/2)/2 = 3/4$ probability. Interestingly, when f and g are dictators, even if h were to be given complete information about $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ (i.e., nothing was erased when forming $\widetilde{\boldsymbol{x}}$ or $\widetilde{\boldsymbol{y}}$), the test could still only be passed with probability at most $3/4$. This is because f and g are balanced (+1 and -1 with equal probability), and thus are opposites of each other with probability exactly $1/2$. When f and g are opposites, no matter what $h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ outputs, it will equal only one of $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$. Since this happens half of the time, $1/4$ of the tests must be failed, upper bounding the success probability by $3/4$.

The other important case is when f and g are "nonmatching dictators", i.e. when $f(x) = x_i$ and $g(y) = (y[i'])_j$, for $i \neq i' \in [K], j \in [d]$. This case is useless when decoding to a satisfying assignment to the original Label Cover instance, and so we hope that they succeed less than $3/4$ of the time. In this case, whether h receives full information of $f(\boldsymbol{x})$ is completely independent of whether it receives full information of $g(\boldsymbol{y})$, as the indices the two depend on sit in entirely different blocks.

Indeed, the correlation in the erasures has no effect; the corresponding blocks of $\boldsymbol{x}$ and $\boldsymbol{y}$ could be erased with half probability *independently* of each other, and this would not help nor hinder h. A large fraction of the time (1/4, to be exact), h receives no information about either $f(\boldsymbol{x})$ or $g(\boldsymbol{y})$, as both $\boldsymbol{x}_i$ and $\boldsymbol{y}[i]$ are erased when forming $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$ with probability 1/4. Thus, h can do nothing but guess randomly when this happens, which is again a coin flip. When this does not happen, then as argued earlier, f and g can pass the test with probability no more than 3/4 because they are balanced. The result is that the success probability in this case is $3/4 * 3/4 + 1/4 * 1/2 = 11/16$, which is less than $12/16 = 3/4$.

The intuition behind the performance of matching dictators carries over to more general functions: if f and g are functions which do not share matching "influential coordinates", then the fact that the erasures are correlated in the $2\mathsf{Lin}(Z_2)$-Test doesn't matter. Indeed, we develop an Invariance Principle in Appendices A and B that shows that when f and g do not share matching influential coordinates, then the probability they pass the $2\mathsf{Lin}(Z_2)$-Test is the same as the probability they pass a modified, uncorrelated version of the test where the erasures of $\boldsymbol{x}$ and $\boldsymbol{y}$ are performed independently of each other. In fact, this Invariance Principle can be applied to *any* "function-in-the-middle" test in the above mold, meaning any test in which f and g are only every compared to an intermediary function h.

The next step is to upper bound the success probability of any pair of functions in this uncorrelated version of the $2\mathsf{Lin}(Z_2)$-Test. The uncorrelated version of the $2\mathsf{Lin}(Z_2)$-Test is exactly the NICD problem stated in Section 3.1, except now the "middleman" function h is playing the game simultaneously with two functions, both f and g. As a result, we are able to adapt the analysis we use to prove Yang's conjecture; the result is Theorem 5.11, which shows that the uncorrelated version of the test cannot be passed with probability more than 11/16. Thus, nonmatching dictators are basically optimal among functions f and g which do not share matching influential coordinates.

Encoding and decoding: It remains to translate these results on the performance of the test to results on the hardness of the $2\mathsf{Lin}(Z_2)$ instance we've generated. This part is entirely standard and follows the basic methodology of Håstad [Hås01]. We have shown above that matching dictators pass the test with probability 3/4. This means that if the starting Label Cover instance was fully satisfiable, then the $2\mathsf{Lin}(Z_2)$ instance is 3/4-satisfiable. On the other hand, if on a significant fraction of the edges (u, v), the Long Codes f_u and g_v pass the $2\mathsf{Lin}(Z_2)$-Test with probability greater than 11/16, then our Invariance Principle tells us that they must share matching influential coordinates. Thus, we can decode the Long Codes to an assignment which satisfies a constant fraction of the Label Cover edges, which concludes our soundness proof. This final part of the hardness result we present for general function-in-the-middle tests in Section 4. Therein we also show that the resulting $2\mathsf{Lin}(Z_2)$ instance is of quasilinear size, implying that under the Exponential Time Hypothesis, nearly exponential time is needed to $(3/4, 11/16 + \epsilon)$-approximate $2\mathsf{Lin}(Z_2)$. One question remains: why does this exactly match Håstad's hardness result? We answer this in Appendix G.

Hardness for other CSPs: A general definition of a function-in-the-middle test is given in Definition 4.7. If such a test T performs its checks using predicates from the set Φ, then Theorem 4.19 automatically implies (c, s)-approximating the Max-Φ problem under quasilinear reductions, where c is the probability matching dictators pass T and s is the highest probability with which any functions can pass test T', the version of test T for which erasures are uncorrelated. The remainder of the paper constructs tests T using various sets of predicates Φ. Our main result, for $2\mathsf{Lin}(Z_q)$, is presented first in Section 6. (The description of the $2\mathsf{Lin}(Z_q)$ test is simple: take the above

2Lin(Z_2)-Test and replace all instances of the set $\{-1, 1\}$ with Z_q.) Next, in Section 5, we analyze the 2Lin(Z_2)-Test (an alternative analysis, using the Fourier transform method, is presented in Appendix F). Finally, in Section 7, we give our Max-Cut, 2And, and 2Sat tests.

Fixing the 2Lin(Z_2)-Test: There are a couple of technical details which we omitted in the above description of the 2Lin(Z_2)-Test for sake of exposition. As it turns out, we can assume that the initial Label Cover instance is d-to-d rather than d-to-1, and we need not assume it is necessarily bipartite. This is the version we choose to present in full in Section 5.3. In addition, when constructing the actual hardness instance, for technical reasons we need to slightly perturb $\boldsymbol{x}$ and $\boldsymbol{y}$ after forming $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$ and use these perturbed versions as the inputs to f and g. These noisy versions of $\boldsymbol{x}$ and $\boldsymbol{y}$ are referred to as $\dot{\boldsymbol{x}}$ and $\dot{\boldsymbol{y}}$, respectively. Details about this are presented in Section 4. Finally, we will often reverse the order the strings are selected, as the following is equivalent method of selecting $\boldsymbol{x}$, $\boldsymbol{y}$, $\widetilde{\boldsymbol{x}}$, and $\widetilde{\boldsymbol{y}}$: first select $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$ so that they are distributed as in the 2Lin(Z_2)-Test. Then, "fill in" their $*$'s with uniformly random elements of $\{-1, 1\}$ to form $\boldsymbol{x}$ and $\boldsymbol{y}$.

4 Inapproximability from NICD

In this section we describe our method for proving 2-CSP inapproximability results by reduction from Label Cover. It is somewhat similar the standard LC- and UG-based approaches ("Long Codes", "dictator tests", etc.) but has some twists. We prefer to think of it as reducing inapproximability to problems in *non-interactive correlation distillation* (NICD).

4.1 NICD tests

It is natural to allow the functions in NICD problems to be "randomized". We formalize this as follows:

Definition 4.1. Let $\triangle_q$ denote the set of probability distributions over Z_q. Equivalently, $\triangle_q = \{(p_0, \dots, p_{q-1}) \in \mathbb{R}^q_{\geq 0} \mid p_0 + \cdots + p_{q-1} = 1\}$. We often identify an element $\ell \in Z_q$ with the constant probability distribution $e_\ell = (0, \dots, 0, 1, 0, \dots, 0) \in \triangle_q$ (with the 1 in coordinate ℓ). We may think of a function f with range $\triangle_q$ as being a "randomized function" with range Z_q.

We will not use the notion of a "randomized function" until Section 6, but it is convenient to point out that "non-randomized functions"–those whose range is Z_q–may be equivalently written as having range $\triangle_q$. Such non-randomized functions only ever map to e_ℓ, for $\ell \in Z_q$. We now give the definitions which let us rule out "trivial" solutions to NICD problems:

Definition 4.2. The function $f : Z_q^m \to \triangle_q$ is *balanced* if $\mathbf{E}_{\boldsymbol{x} \sim Z_q^m}[f(\boldsymbol{x})] = (\frac{1}{q}, \dots, \frac{1}{q})$. Here and throughout, $\boldsymbol{x} \sim Z_q^m$ means that $\boldsymbol{x}$ is uniformly distributed.

Definition 4.3. The function $f : Z_q^m \to \triangle_q$ is *folded* (a stronger condition than being balanced) if $f(x + \ell) = \mathrm{rot}_\ell(f(x))$ for all $x \in Z_q^m$ and $\ell \in Z_q$. Here $x + \ell$ is shorthand for $x + (\ell, \ell, \dots, \ell)$, and $\mathrm{rot}_\ell(\nu) \in \triangle_q$ is the "cyclic rotation of ν by ℓ places"; i.e., $\mathrm{rot}_\ell(\nu)_i = \nu_{(i-\ell \bmod q)}$. When $f : Z_q^m \to Z_q$ is "non-randomized", this simply means that $f(x + \ell) = f(x) + \ell$.

We require a few more definitions:

Definition 4.4. We write $\mu^d_{q,\rho}$ for the probability distribution on $Z_q^d \cup \{*^d\}$ which is uniform on Z_q^d with probability ρ and $*^d$ with probability $1 - \rho$. If $d = 1$ or $q = 2$ or $\rho = \frac{1}{2}$ we omit writing it. Note that $\mu^d_{q,\rho}$ is invariant under permutations of the d coordinates.

Definition 4.5. Given a string $x \in Z_q^m$, an *η-noisy copy* is defined to be the randomly chosen string $\dot{\boldsymbol{x}}$ in which for each $i \in [m]$ independently, $\dot{\boldsymbol{x}}_i = x_i$ with probability $1 - \eta$ and $\dot{\boldsymbol{x}}_i$ is set uniformly at random with probability η. We also define the operator T_η on functions $f : Z_q^m \to \triangle_q$ by $\mathrm{T}_\eta f(x) = \mathbf{E}_{\dot{\boldsymbol{x}}}[f(\dot{\boldsymbol{x}})]$. We may denote this by $\dot{\boldsymbol{x}} \sim_\rho x$.

Our main result will concern "correlated information distributions" π on $(Z_q^{d_1} \cup \{*^{d_1}\}) \times (Z_q^{d_2} \cup \{*^{d_2}\})$. One property we will require of our correlated information distributions is that the only correlation present between the two coordinates concerns the presence or lack of information. For example, π is allowed to always put $*^d$ in exactly one of the two coordinates, so that information is always available for one of the two sides, but we forbid π to always output two matching strings. This is formalized in the next definition.

Definition 4.6. Given a correlated information distribution π, consider the "filled in" version of π:

1. Sample $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ from π.
2. Form $\boldsymbol{x}$ and $\boldsymbol{y}$ by replacing each $*$ in $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$ by an independent and uniformly random element of Z_q.

Then π is *pseudo-independent* if $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent.

We are now able to define the general class of NICD "tests" which will be useful for our inapproximability results. Because we are working with Label Cover our tests need to operate on "blocked" functions; i.e., functions over domains $\Sigma^{dK} \cong (\Sigma^d)^K$. Aside from this generality, we have chosen to make our definition rather narrow for the sake of simplicity. To attack inapproximability for more CSPs, one might wish to generalize this definition in several directions; however, such generalizations would require proving more complicated "Invariance" theorems.

Definition 4.7. A *2-party, q-ary, Φ-based, (d_1, d_2)-blocked, η-noise correlated test* Test consists of two probability distributions:

- A correlated information distribution π on $(Z_q^{d_1} \cup \{*^{d_1}\}) \times (Z_q^{d_2} \cup \{*^{d_2}\})$. Each marginal π_j on $Z_q^{d_j} \cup \{*^{d_j}\}$ (for $j = 1, 2$) must be equal to $\mu_{q,\rho_j}^{d_j}$ for some $\rho_j \in [0, 1]$. Further, π must be pseudo-independent.
- A "test distribution" $\mathcal{T}$ on $\Phi \times \{1, 2\}$, where Φ is a fixed collection of predicates $\phi : Z_q \times Z_q \to \{0, 1\}$.

Given $K \in \mathbb{N}^+$, Test operates on blocked functions $f : Z_q^{d_1 K} \to Z_q$, $g : Z_q^{d_2 K} \to Z_q$, and $h : (Z_q^{d_1} \cup \{*^{d_1}\})^K \times (Z_q^{d_2} \cup \{*^{d_2}\})^K \to Z_q$ as follows:

1. A pair of strings $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \in (Z_q^{d_1} \cup \{*^{d_1}\})^K \times (Z_q^{d_2} \cup \{*^{d_2}\})^K$ is chosen randomly from the product distribution $\pi^{\otimes K}$.
2. String $\boldsymbol{x} \in Z_q^{d_1 K}$ is formed by randomly "filling in" $\widetilde{\boldsymbol{x}}$; i.e., replacing each $*$ by an independent uniformly random symbol from Z_q. Similarly $\boldsymbol{y} \in Z_q^{d_2 K}$ is (independently) formed by randomly filling in $\widetilde{\boldsymbol{y}}$.
3. String $\dot{\boldsymbol{x}} \in Z_q^{d_1 K}$ is set to be an η-noisy copy of $\boldsymbol{x}$, and similarly for $\dot{\boldsymbol{y}}$.
4. Finally, $(\boldsymbol{\phi}, \boldsymbol{j})$ is chosen from $\mathcal{T}$. If $\boldsymbol{j} = 1$ then one "tests" $\boldsymbol{\phi}(f(\dot{\boldsymbol{x}}), h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}))$; if $\boldsymbol{j} = 2$ then one "tests" $\boldsymbol{\phi}(g(\dot{\boldsymbol{y}}), h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}))$.

Definition 4.8. The *success probability* or *value* of f, g, h on TEST, denoted $\mathrm{Val}_{\mathrm{TEST}}(f, g, h)$, is simply the probability that the test is satisfied.

Remark 4.9. Given TEST and functions f, g, the optimal choice (or choices) for h is determined. We will often just consider varying f and g, and choosing the best h to go along with them. We therefore introduce the notation $\mathrm{Val}_{\mathrm{TEST}}(f, g) = \max_h\{\mathrm{Val}_{\mathrm{TEST}}(f, g, h)\}$.

Definition 4.10. The *optimal success probability* or *optimal value* of TEST, denoted $\mathrm{Opt}(\mathrm{TEST})$, is $\max_{f,g}\{\mathrm{Val}_{\mathrm{TEST}}(f, g)\}$. We also define $\mathrm{Opt}_{\mathrm{folded}}(\mathrm{TEST})$ for when the maximization is only over folded f and g, and also $\mathrm{Opt}_{\mathrm{balanced}}(\mathrm{TEST})$ for when the maximization is only over balanced f and g.

Fact 4.11. *Let* TEST *be a 2-party 0-noise test and let* TEST_η *be its η-noisy version. Then* $\mathrm{Opt}(\mathrm{TEST}) \geq \mathrm{Opt}(\mathrm{TEST}_\eta)$ *and* $\mathrm{Opt}_{\mathrm{folded}}(\mathrm{TEST}) \geq \mathrm{Opt}_{\mathrm{folded}}(\mathrm{TEST}_\eta)$.

Proof. This is because $\mathrm{Val}_{\mathrm{TEST}_\eta}(f, g) = \mathrm{Val}_{\mathrm{TEST}}(\mathrm{T}_\eta f, \mathrm{T}_\eta g)$, and $\mathrm{T}_\eta f$, $\mathrm{T}_\eta g$ are folded if f, g are. However, we must account for the fact that $\mathrm{T}_\eta f$ and $\mathrm{T}_\eta g$ do not necessarily have the domain Z_q, even if f and g do. Consider selecting $\boldsymbol{z}$ (respectively, $\boldsymbol{w}$) from a distribution on strings in $Z_q^{d_1 K}$ (respectively, $Z_q^{d_2 K}$) for which each coordinate is independently 0 with probability η and uniform on Z_q otherwise. Then given $\boldsymbol{x}$, $\boldsymbol{x} + \boldsymbol{z}$ is distributed as $\dot{\boldsymbol{x}}$, and given $\boldsymbol{y}$, $\boldsymbol{y} + \boldsymbol{w}$ is distributed as $\dot{\boldsymbol{y}}$. Thus, in TEST_η, we may use $\boldsymbol{x}+\boldsymbol{z}$ and $\boldsymbol{y}+\boldsymbol{w}$ in place of $\dot{\boldsymbol{x}}$ and $\dot{\boldsymbol{y}}$, respectively. By the probabilistic method, there is a setting to z and w for which, conditioned on $\boldsymbol{z} = z$ and $\boldsymbol{w} = w$, the test is passed with probability at least $\mathrm{Val}_{\mathrm{TEST}_\eta}(f, g)$. Then define the functions $f' = f(\cdot + z)$ and $g' = g(\cdot + w)$. Clearly, $\mathrm{Val}_{\mathrm{TEST}}(f', g') \geq \mathrm{Val}_{\mathrm{TEST}_\eta}(f, g)$. Furthermore, f' and g' are folded if f and g are. □

4.2 Blocks, influences, Invariance, and uncorrelated tests

As mentioned, our 2-party NICD tests operate on "blocked" functions; e.g., $f : Z_q^{d_1 K} \to Z_q$, $g : Z_q^{d_2 K} \to Z_q$. In fact, for reductions from $\mathsf{LC}_{d_1 K, d_2 K}$ the "blocks" will not necessarily be contiguous. Rather, the "blocks" for f will be $\pi_u^{-1}(1), \ldots, \pi_u^{-1}(K)$ for some d_1-to-1 map $\pi_u : [d_1 K] \to [K]$, and similarly for g with a d_2-to-1 map π_v.

Definition 4.12. Given a 2-party (d_1, d_2)-blocked test TEST, we introduce the natural notion of applying it to f, g, h *under the "blocking maps" π_u and π_v*. Notice that for this to make sense, it is crucial that our definition of 2-party tests is insensitive to permutations of coordinates within blocks and also to permutations of blocks. (Briefly this is because the distributions μ^d_{q,ρ_j} are permutation-invariant, the test uses product distributions across the K blocks, and because making η-noisy copies acts independently on coordinates.)

Often the optimal choice of f and g for a given 2-party NICD test is "matching dictators" (note that dictators are folded):

Definition 4.13. Suppose that $f : Z_q^{d_1 K} \to Z_q$ and $g : Z_q^{d_2 K} \to Z_q$ are *dictator functions*, meaning that $f(x) = x_i$ for some $i \in [d_1 K]$ and $g(y) = y_j$ for some $j \in [d_2 K]$. We say they are *matching dictators* (under the blocking maps π_u, π_v) if $\pi_u(i) = \pi_v(j)$; otherwise we say they are *non-matching dictators*.

As is common in inapproximability, we will be concerned with functions which are "far from" being dictators in the sense of having "small noisy-influences". To make this notion precise we recall the *Hoeffding orthogonal decomposition* (or *Efron–Stein decomposition*) for functions $f : \Sigma^m \to \mathbb{R}^q$:

Definition 4.14. For $f : \Sigma^m \to \mathbb{R}^q$ we define $\|f\|_2^2 = \mathbf{E}_{\boldsymbol{x}\sim\Sigma^m}[|f(\boldsymbol{x})|^2]$. Here $|\cdot|$ denotes Euclidean length in $\mathbb{R}^q$. Note that if $f : \Sigma^m \to \triangle_q$ then $\|f\|_2^2 \le 1$.

Fact 4.15. *Every function $f : \Sigma^m \to \mathbb{R}^q$ can be written as $f = \sum_{S\subseteq[m]} f^S$, where the functions $f^S : \Sigma^m \to \mathbb{R}^q$ have the following properties: $f^S(x)$ depends only on $(x_i)_{i\in S}$; and, $\mathbf{E}[\langle f^S(\boldsymbol{x}), f^{S'}(\boldsymbol{x})\rangle] = 0$ for any $S \neq S'$, where $\boldsymbol{x}$ is uniformly distributed on Σ^m and $\langle\cdot,\cdot\rangle$ denotes the usual inner product on $\mathbb{R}^q$. As a consequence,*

$$\|f\|_2^2 = \sum_{S\subseteq[m]} \|f^S\|_2^2.$$

In addition, if $S' \subseteq [m]$ does not contain S, then for any assignment $x_{S'}$ to the variables in S', $\mathbf{E}[f^S(\boldsymbol{x})] = 0$, where $\boldsymbol{x}$ is a uniformly random element of Σ^m conditioned on $\boldsymbol{x}_i = (x_{S'})_i$ for $i \in S'$.

Definition 4.16. The *η-noisy influence of $B \subseteq [m]$ on $f : \Sigma^m \to \mathbb{R}^q$* is defined to be

$$\mathbf{Inf}_B^{(1-\eta)}[f] = \sum_{S:S\cap B\neq\emptyset} (1-\eta)^{|S|}\|f^S\|_2^2.$$

Suppose that $f : Z_q^{d_1K} \to Z_q$ and $g : Z_q^{d_2K} \to Z_q$ have (roughly speaking) "no influential blocks in common" (under blocks induced by π_u and π_v). For example, f and g may be non-matching dictators. In this case, we might expect that f and g cannot "take advantage" of the correlation between h's inputs $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$ in test Test. The key technical result we need for our hardness reduction is an "Invariance" theorem which formalizes this idea:

Definition 4.17. Given a 2-party correlated test Test, its *uncorrelated version* Test$'$ is the same test but with the pair of strings $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ drawn *independently* from π's marginal distributions; i.e., $\widetilde{\boldsymbol{x}} \sim \pi_1$ and $\widetilde{\boldsymbol{y}} \sim \pi_2$ independently.

Theorem 4.18. *There is a function $\kappa(\eta, \epsilon, q) > 0$ with $\kappa(\eta, \epsilon, q) = (\eta\epsilon/q)^{O((\log q)/\eta)}$ such that the following holds:*

Let Test *be a 2-party, q-ary, (d_1, d_2)-blocked, η-noise correlated test. Let* Test$'$ *denote its uncorrelated version. Assume we are applying these tests under the blocking maps π_u and π_v. Let $f : Z_q^{d_1K} \to \triangle_q$ and $g : Z_q^{d_2K} \to \triangle_q$ satisfy*

$$\min(\mathbf{Inf}^{(1-\eta)}_{\pi_u^{-1}(t)}[f], \mathbf{Inf}^{(1-\eta)}_{\pi_v^{-1}(t)}[g]) \le \kappa(\eta, \epsilon, q) \quad \forall t \in [K].$$

Then $\mathrm{Val}_{\text{Test}}(f, g) \le \mathrm{Opt}(\text{Test}') + \epsilon,$

and $\mathrm{Val}_{\text{Test}}(f, g) \le \mathrm{Opt}_{\text{folded}}(\text{Test}') + \epsilon$ *if f, g folded.*

We prove Theorem 4.18 in Appendix B. Mostly, our proof follows the general outline of Mossel's Invariance theorem [Mos10], with a few differences. In the terminology of [Mos10], we have a sequence of K orthonormal ensembles, each of which corresponds to a certain block of $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$. Mossel's Invariance theorem uses as a parameter for an application of hypercontractivity the least nonzero probability of an assignment, which in our case is $\frac{1}{2}q^{-\max\{d_1,d_2\}}$. This is too small for us. Instead, we are able to maneuver it so that we only need to worry about the hypercontractivity of Z_q.

4.3 The hardness of approximation reduction

The following theorem shows how to convert a 2-party correlation test to a hardness of approximation result. The proof is quite standard, and largely follows of Håstad's methods [Hås01] .

Theorem 4.19.
Let Test *be a 2-party, q-ary, Φ-based, (d_1, d_2)-blocked, 0-noise correlated test.*
Let Test$'$ *be its uncorrelated version.*
Suppose that matching dictators achieve success probability at least c for Test.
Suppose also that $\mathrm{Opt}(\text{Test}') \le s$.
Let $\eta, \epsilon \in \mathbb{Q}^+$ and assume $\delta \le \epsilon \cdot \eta^2 \cdot \kappa(\eta, \epsilon, q)^2$, where κ is as in Theorem 4.18.
Then there is a reduction from $(1, \delta)$-deciding LC_{d_1K, d_2K} to $(c - \eta, s + 2\epsilon)$-deciding Max-$\Phi$.
If instead we assume $\mathrm{Opt}_{\text{folded}}(\text{Test}') \le s$ *then we get the result for* Max-Φ^+.
The reduction maps size-n instances to size-$q^{O(d_1K+d_2K)}n$ instances and runs in time polynomial in the size of its output.

Proof. Let $((U, V), E, (\pi_{e,u}, \pi_{e,v}))$ be a given size-n instance of LC_{d_1K,d_2K}. For each $u \in U$ (respectively, $v \in V$) the reduction introduces a collection of Max-Φ variables identified with $Z_q^{d_1K}$ (respectively, $Z_q^{d_2K}$); we think of an assignment to these variables as a function $f_u : Z_q^{d_1K} \to Z_q$ (respectively, $g_v : Z_q^{d_2K} \to Z_q$). Furthermore, for each edge $(u, v) \in E$ the reduction introduces a collection of Max-Φ variables identified with $(Z_q^{d_1} \cup \{*^{d_1}\})^K \times (Z_q^{d_2} \cup \{*^{d_2}\})^K$; we think of an assignment to these variables as a function h_{uv} with range Z_q.

Let Test_η denote the η-noisy version of the test Test (and Test'_η its uncorrelated version). For each edge $e = (u, v) \in E$ the reduction introduces a collection of Φ-constraints on the assignments f_u, g_v, and h_{uv}. These constraints are precisely those given by applying Test_η under the blocking maps $\pi_{e,u}$, $\pi_{e,v}$ (with weights/probabilities scaled down by a factor of $|E|$). This completes the description of the reduction.

Completeness. Assume that $F : U \to [d_1K], V \to [d_2K]$ is an assignment satisfying all constraints of the LC_{d_1K,d_2K} instance. Consider the assignment to the Max-Φ instance in which for each $u \in U$ and $v \in V$ we take $f_u(x) = x_{F(u)}$, $g_v(y) = y_{F(v)}$. For $e = (u, v) \in E$, these are matching dictators with respect to $\pi_{e,u}$ and $\pi_{e,v}$, since F satisfies the constraint on e. Therefore there exists a choice for h_{uv} such that $\mathrm{Val}_{\text{Test}}(f_u, g_v, h_{uv}) \ge c$. Since f_u and g_v are dictators it is easy to see that $\mathrm{Val}_{\text{Test}_\eta}(f_u, g_v, h_{uv})$ is still at least $c - \eta$. Since this holds for each edge $(u, v) \in E$, it follows that our assignment achieves value at least $c - \eta$ on the Max-Φ instance.

Soundness. We prove the contrapositive. Suppose there are assignments $(f_u)_{u \in U}$, $(g_v)_{v \in V}$, $(h_{uv})_{(u,v) \in E}$ for the Max-Φ which collectively achieve value exceeding $s + 2\epsilon$. Then for at least an ϵ fraction of edges $e = (u, v) \in E$ — call them "good" edges — we have $\mathrm{Val}_{\text{Test}_\eta}(f_u, g_v, h_{uv}) > s + \epsilon \ge \mathrm{Opt}(\text{Test}'_\eta) + \epsilon$, the second inequality being Fact 4.11. We may therefore apply Theorem 4.18 to deduce that for each good (u, v),

$$\exists t_{uv} \in [K] \text{ s.t. } \mathbf{Inf}^{(1-\eta)}_{\pi_{e,u}^{-1}(t_{uv})}[f_u],\ \mathbf{Inf}^{(1-\eta)}_{\pi_{e,v}^{-1}(t_{uv})}[g_v] \ge \kappa = \kappa(\eta, \epsilon, q). \tag{1}$$

Consider now the following randomized procedure for generating an assignment $\boldsymbol{F}$ for the LC instance. For each $u \in U$, the procedure first chooses $S \subseteq [d_1K]$ with probability $\|f_u^S\|_2^2$. (From Definition 4.14 these numbers sum to at most 1; for any remaining probability, S can be chosen arbitrarily.) Then $\boldsymbol{F}(u)$ is set to be a uniformly random element of S (or an arbitrary label if

$S = \emptyset$). An identical procedure is used to assign $\boldsymbol{F}(v)$ for $v \in V$, based on g_v. Observe that for any set $B \subseteq [d_1 K]$ and $u \in U$,

$$\begin{aligned}\mathbf{Pr}[\boldsymbol{F}(u) \in B] &\geq \sum_{S : S \cap B \neq \emptyset} \|f_u^S\|_2^2 \cdot (|S \cap B|/|S|) \\ &\geq \sum_{S : S \cap B \neq \emptyset} \|f_u^S\|_2^2 \cdot \eta(1-\eta)^{|S|/|S \cap B|} \quad (\text{since } r \geq \eta(1-\eta)^{1/r},\ \forall r > 0, \eta \in [0,1]) \\ &\geq \eta \sum_{S : S \cap B \neq \emptyset} \|f_u^S\|_2^2 \cdot (1-\eta)^{|S|} = \eta \mathbf{Inf}_B^{(1-\eta)}[f_u], \end{aligned} \tag{2}$$

and similarly for $v \in V$.

Now for a good edge $e = (u, v)$, let us estimate the probability (over $\boldsymbol{F}$) that $\boldsymbol{F}(u)$ and $\boldsymbol{F}(v)$ satisfy the constraint associated with e. It is

$$\begin{aligned}\mathop{\mathbf{Pr}}_{\boldsymbol{F}}[\pi_{e,u}(\boldsymbol{F}(u)) = \pi_{e,v}(\boldsymbol{F}(v))] &\geq \mathbf{Pr}[\boldsymbol{F}(u) \in \pi_{e,u}^{-1}(t_{uv}) \ \&\ \boldsymbol{F}(v) \in \pi_{e,v}^{-1}(t_{uv})] \\ &= \mathbf{Pr}[\boldsymbol{F}(u) \in \pi_{e,u}^{-1}(t_{uv})] \cdot \mathbf{Pr}[\boldsymbol{F}(v) \in \pi_{e,v}^{-1}(t_{uv})] \\ &\geq \eta^2 \cdot \mathbf{Inf}^{(1-\eta)}_{\pi_{e,u}^{-1}(t_{uv})}[f_u] \cdot \mathbf{Inf}^{(1-\eta)}_{\pi_{e,v}^{-1}(t_{uv})}[g_v] \geq \eta^2 \kappa^2, \end{aligned}$$

where we used (2) and then (1). It follows that expected fraction of constraints in the LC instance that $\boldsymbol{F}$ satisfies is at least $\epsilon \cdot \eta^2 \kappa^2 \geq \delta$; hence the optimal value of the LC instance is at least δ.

This completes the proof except for the statement about $\mathrm{Opt}_{\mathrm{folded}}(\textsc{Test}')$ and $\mathsf{Max}\text{-}\Phi^+$. For this we the standard folding trick: Instead of having the reduction introduce a collection of variables corresponding to $Z_q^{d_1 K}$ for each $u \in U$, we only introduce variables for $Z_q^{d_1 K - 1}$. We think of an assignment for these variables as the restriction of a function $f_u : Z_q^{d_1 K} \to Z_q$ to inputs of the form $(0, x')$, $x' \in Z_q^{d_1 K - 1}$. We extend f_u to a folded function via $f_u(\ell, x') = f_u(0, x' - (\ell, \dots, \ell)) + \ell$. Is this way, any $\mathsf{Max}\text{-}\Phi$ constraint involving $f_u(x)$ can be replaced with a $\mathsf{Max}\text{-}\Phi^+$ constraint involving $f_u(0, x')$. We similarly arrange for folded functions g_v. We now proceed through the above proof: for the completeness we use the fact that dictators are folded; for the soundness we use the folded versions of Fact 4.11 and Theorem 4.18. □

Corollary 4.20. *Let $2 \leq q = q(n) \leq (\log\log\log n)^3$. Assume that for all $d_2, K \in \mathbb{N}^+$ there is a $q^{O(d_2 K)}$-time algorithm for generating (the description of) a (d_1, d_2)-blocked* Test *satisfying the hypotheses of Theorem 4.19, for some $d_1 \leq d_2$. Assume that $c < 1$. Then for a certain $\epsilon = \epsilon(n) = \widetilde{\Theta}(1/\log\log\log n)$, there is a quasilinear-size reduction from size-n instances of* 3Sat *to the problem of $(c, s+\epsilon)$-deciding* $\mathsf{Max}\text{-}\Phi$ *(or* $\mathsf{Max}\text{-}\Phi^+$*, under the assumption* $\mathrm{Opt}_{\mathrm{folded}}(\textsc{Test}') \leq s$*).*

Proof. (Sketch.) This follows by combining the Moshkovitz–Raz Theorem with Theorem 4.19. One takes $\delta = (\log\log n)^{-c}$ for a sufficiently small constant c and $\eta = \epsilon$. With these choices the conditions of Theorem 4.19 are satisfied and the overall size of the $\mathsf{Max}\text{-}\Phi$ instance produced is still $n^{1+o(1)}$. Using $c < 1$, we can convert the resulting $(c-\epsilon, s+2\epsilon)$-hardness into $(c, s+O(\epsilon))$-hardness using padding. □

5 A test for $2\mathsf{Lin}(Z_2)$

In this section we introduce the model of non-interactive correlation distillation (NICD), with the goal of proving Yang's conjecture [Yan07]. We then use this proof technique to analyze our $2\mathsf{Lin}(Z_2)$ function test, which yields the $q = 2$ case of Theorem 2.5 and matches the gadget-based

hardness result of [Hås01, TSSW00]. The approach to solving NICD problems we develop here is useful not only for our $2\mathsf{Lin}(Z_2)$ result, but also for the hardness of approximation results for the remaining binary CSPs presented in Section 7. In some sense, the binary NICD model and our $2\mathsf{Lin}(Z_2)$ test form the foundation of all our binary correlation tests. Finally, these have natural q-ary generalizations, and these form the basis of our Unique Games hardness result in Section 6.

5.1 Binary NICD

We begin with the model of NICD *over the erasure channel*, which considers the following test, for $0 < \rho < 1$:

NICD-Test(ρ)

- Given functions $f : \{-1,1\}^n \to \{-1,1\}$ and $h : \{-1,1,*\}^n \to \{-1,1\}$:
- Draw $\widetilde{\boldsymbol{x}} \in \{-1,1,*\}^n$ so that each coordinate is independently a $*$ with probability $1-\rho$ and a uniformly random element of $\{-1,1\}$ otherwise.
- Form $\boldsymbol{x} \in \{-1,1\}^n$ by replacing each $*$ in $\widetilde{\boldsymbol{x}}$ with an independent uniformly random element from $\{-1,1\}$.
- Test $f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}})$.

We call the pair (f,h) a *strategy*. This is only a 1-party test, but many of the notions we used for 2-party tests transfer over naturally. In particular, if T is a 1-party test, we may write $\mathrm{Val}_T(f,h)$ for the probability the test is satisfied, $\mathrm{Val}_T(f)$ for $\max_h\{\mathrm{Val}_T(f,h)\}$, and $\mathrm{Opt}(T)$ for $\max_f\{\mathrm{Val}_T(f)\}$. Furthermore, $\mathrm{Opt}_{\mathrm{folded}}(T)$ and $\mathrm{Opt}_{\mathrm{balanced}}(T)$ take the optimum over strategies where f (not necessarily h) is folded or balanced, respectively.

It is easy to see that given f, an optimal h is

$$h(\widetilde{\boldsymbol{x}}) = \mathrm{sign}(\mathbf{E}[f(\boldsymbol{x}) \mid \widetilde{\boldsymbol{x}}]),$$

where we define $\mathrm{sign}(0) = 1$. It is also easy to see that this test can be passed always if f and h are constant; henceforth, we will only consider balanced f.

Perhaps the most obvious strategy is for f to be a dictator. In this case, the success probability of f is $1/2 + \rho/2$. Ke Yang [Yan07] conjectured that this was optimal (we assume his conjecture was for $\rho \geq 1/2$ only), and indeed our Theorem 5.1 confirms this.

Theorem 5.1. *For $\rho \geq 1/2$, $\mathrm{Opt}_{\mathrm{balanced}}(\text{NICD-Test}(\rho)) = 1/2 + \rho/2$. When $\rho > 1/2$, dictators and negated dictators uniquely achieve this value.*

Which function is best for $\rho < 1/2$ is less clear, however, but what is clear is that dictators are no longer optimal. In fact, what is perhaps the only other obvious strategy for NICD—majority—outperforms dictators for this range of ρ. For n odd, we define the majority function to be

$$\mathsf{MAJ}_n(x) = \mathrm{sign}\left(\sum_{i=1}^{n} x_i\right).$$

It is clear that if f is $\mathsf{MAJ}_n(x)$, then the optimal strategy for h is

$$h(z) = \mathrm{sign}\left(\sum_{i:z_i \neq *}^{n} z_i\right),$$

which is simply the majority over the unerased bits. The exact success probability of MAJ_n for a specific n is tedious to compute and rather unilluminating, but fortunately one can compute the limiting success probability as n approaches ∞. Using the 2-dimensional Central Limit Theorem, it can be shown that the success probability is $1/2 + \arcsin(\sqrt{\rho})/\pi$ (see, for example, the analysis in [O'D02]). For $\rho > 1/2$, this is strictly worse than the dictator strategy, but it is better for $\rho < 1/2$ and is exactly equal for $\rho = 1/2$. In fact, the case of $\rho = 1/2$ is especially nice, and the entire class of linear threshold functions succeeds here with probability exactly $3/4$.

In contrast to $\rho \geq 1/2$, what we know about NICD over the erasure channel when $\rho < 1/2$ is fairly limited. What we do know comes mainly from the following theorem of [O'D02, Yan07] (see also [Mos10]):

Theorem 5.2. *For $0 < \rho < 1$, $\mathrm{Opt}_{\mathrm{balanced}}(\text{NICD-}\textsc{Test}(\rho)) \leq 1/2 + \sqrt{\rho}/2$.*

The success probability of the best strategy is clearly an increasing function of ρ, so combining Theorem 5.1 with Theorem 5.2 gives the best upper bound for $\mathrm{Opt}_{\mathrm{balanced}}(\text{NICD-}\textsc{Test}(\rho))$ as $\min(3/4, 1/2 + \sqrt{\rho}/2)$, when $\rho < 1/2$. This bound is tight to within a constant as ρ approaches zero, because the success probability of majority is $1/2 + \sqrt{\rho}/\pi + \Theta(\rho^{3/2})$. Indeed, we believe that majority is the best strategy for $\rho < 1/2$ but are unable to prove it. One piece of evidence for this is the "Majority is Most Predictable" theorem of Mossel [Mos10], which states, roughly, that among "low-influence" functions, the success probability of majority is optimal. It seems reasonable that the best strategy for $\rho < 1/2$ would be a low-influence function, as any function which relies heavily on specific coordinates would be difficult to predict when the values of those coordinates are erased.

We now begin proving our results. First, we give a simple proof of Theorem 5.1 in the case where $\rho = 1/2$. The technique we use here will be reused in several of the later NICD proofs.

Proof. Let (f, h) be a strategy where f is balanced. Consider selecting two strings $\boldsymbol{y}, \boldsymbol{y}' \in \{-1, 1\}^n$ independently and uniformly at random, and forming $\widetilde{\boldsymbol{y}}$ as follows:

$$\widetilde{\boldsymbol{y}}_i = \begin{cases} \boldsymbol{y}_i & \text{if } \boldsymbol{y}_i = \boldsymbol{y}'_i, \\ * & \text{if } \boldsymbol{y}_i \neq \boldsymbol{y}'_i. \end{cases}$$

Then clearly $\widetilde{\boldsymbol{y}}$ is distributed as $\widetilde{\boldsymbol{x}}$ is in the test, and $\boldsymbol{y}$ and $\boldsymbol{y}'$ are both randomly "filled-in" versions of it. Thus,

$$\begin{aligned} \mathbf{Pr}[f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}})] &= \mathrm{avg}\{\mathbf{Pr}[f(\boldsymbol{y}) = h(\widetilde{\boldsymbol{y}})], \mathbf{Pr}[f(\boldsymbol{y}') = h(\widetilde{\boldsymbol{y}})]\} \\ &\leq \frac{1}{2}\,\mathbf{E}[M(f(\boldsymbol{y}), f(\boldsymbol{y}'))], \end{aligned}$$

where M outputs the number of input bits in the majority. This is because whatever $f(\boldsymbol{y})$ and $f(\boldsymbol{y}')$ turn out to be, $h(\widetilde{\boldsymbol{y}})$ can only agree with at most $M(f(\boldsymbol{y}), f(\boldsymbol{y}'))$ of them.

Because $\boldsymbol{y}$ and $\boldsymbol{y}'$ are independent and f is balanced, $f(\boldsymbol{y})$ and $f(\boldsymbol{y}')$ are distributed as independent, uniformly-random ± 1 bits, so $M(f(\boldsymbol{y}), f(\boldsymbol{y}'))$ is either 1 or 2, each with probability $1/2$. Thus, $\mathbf{E}[M(f(\boldsymbol{y}), f(\boldsymbol{y}'))] = 3/2$, and the success probability of f and h is at most $3/4$. □

We will extend this proof technique to prove Theorem 5.1 in its entirety. First, we need the following well-known fact about the noise stability of balanced functions.

Proposition 5.3. *For $0 < \eta < 1$ and a function $f : \{-1, 1\}^n \to \{-1, 1\}$, the* noise stability *of f at η is $\mathbb{S}_\eta(f) = \mathbf{E}_x[f(x)(T_\eta f)(x)]$. If f is balanced, then $\mathbb{S}_\eta(f) \leq \eta$. Equality holds if and only if f is a dictator or a negated dictator.*

Proof. We assume familiarity with Fourier analysis on the Boolean hypercube. Define $\lambda \in \{-1,1\}^n$ to be distributed so that each coordinate λ_i is independently random subject to $\mathbf{E}[\lambda_i] = \eta$. Then $y \sim_\eta x$ is identically distributed to $\lambda \cdot x$, using coordinate-wise multiplication. Thus,

$$\begin{aligned}\mathbb{S}_\eta(f) = \mathop{\mathbf{E}}_x[f(x)(T_\eta f)(x)] &= \sum_{S,T\subseteq[n]} \hat{f}(S)\hat{f}(T) \mathop{\mathbf{E}}_{x,y\sim_\eta x}[\chi_S(x)\chi_T(y)] \\ &= \sum_{S,T\subseteq[n]} \hat{f}(S)\hat{f}(T) \mathop{\mathbf{E}}_{x,\lambda}[\chi_S(x)\chi_T(x\cdot\lambda)]. \end{aligned} \tag{3}$$

As $\chi_T(x\cdot\lambda) = \chi_T(x)\chi_T(\lambda)$, we have that

$$\mathop{\mathbf{E}}_{x,\lambda}[\chi_S(x)\chi_T(x\cdot\lambda)] = \mathop{\mathbf{E}}_x[\chi_S(x)\chi_T(x)] \mathop{\mathbf{E}}_\lambda[\chi_T(\lambda)] = \mathbf{1}_{S=T}\cdot\eta^{|T|}.$$

Substituting this into Equation (3) yields $\mathbb{S}_\eta(f) = \sum_{S\subseteq[n]} \hat{f}(S)^2\eta^{|S|}$. It is well-known (Parseval's theorem) that $\sum_{S\subseteq[n]} \hat{f}(S)^2 = 1$ when f is $\{-1,1\}$-valued. Furthermore, it is easy to see that if f is balanced, then $\hat{f}(\emptyset) = 0$. Thus, for a balanced f, $\mathbb{S}_\eta(f) \le \eta$, with equality in the case that all of f's Fourier mass is on sets of size one. The only such functions are dictators and antidictators, which concludes the proof. □

We can now use this to prove Theorem 5.1.

Proof of Theorem 5.1. Let (f,h) be a strategy where f is balanced. Consider selecting two strings $\boldsymbol{y}, \boldsymbol{y}' \in \{-1,1\}^n$, where $\boldsymbol{y}$ is uniformly random and $\boldsymbol{y}'$ is a $(2\rho-1)$-correlated copy of $\boldsymbol{y}$. Form $\widetilde{\boldsymbol{y}}$ as above:

$$\widetilde{\boldsymbol{y}}_i = \begin{cases} \boldsymbol{y}_i & \text{if } \boldsymbol{y}_i = \boldsymbol{y}'_i, \\ * & \text{if } \boldsymbol{y}_i \neq \boldsymbol{y}'_i. \end{cases}$$

Clearly, $\widetilde{\boldsymbol{y}}$ is distributed as $\widetilde{\boldsymbol{x}}$ is in the test, and $\boldsymbol{y}$ and $\boldsymbol{y}'$ are distributed like randomly "filled-in" versions of it. Thus,

$$\begin{aligned}\mathbf{Pr}[f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}})] &= \mathrm{avg}\{\mathbf{Pr}[f(\boldsymbol{y}) = h(\widetilde{\boldsymbol{y}})], \mathbf{Pr}[f(\boldsymbol{y}') = h(\widetilde{\boldsymbol{y}})]\} \\ &\le \frac{1}{2}\mathbf{E}[M(f(\boldsymbol{y}), f(\boldsymbol{y}'))],\end{aligned}$$

where M outputs the number of input bits in the majority. This is because whatever $f(\boldsymbol{y})$ and $f(\boldsymbol{y}')$ turn out to be, $h(\widetilde{\boldsymbol{y}})$ can only agree with at most $M(f(\boldsymbol{y}), f(\boldsymbol{y}'))$ of them.

Unfortunately, $\boldsymbol{y}$ and $\boldsymbol{y}'$ are not independent, so we need a different way of analyzing the expected size of this majority. We can write out M as $M(u_1,u_2) = 3/2 + u_1u_2/2$. Then

$$\begin{aligned}\frac{1}{2}\mathbf{E}[M(f(\boldsymbol{y}), f(\boldsymbol{y}'))] &= \frac{1}{2}\mathbf{E}[3/2 + f(\boldsymbol{y})f(\boldsymbol{y}')/2] \\ &= \frac{3}{4} + \frac{1}{4}\mathbf{E}[f(\boldsymbol{y})f(\boldsymbol{y}')] \\ &= \frac{3}{4} + \frac{1}{4}\mathbb{S}_{(2\rho-1)}(f).\end{aligned} \tag{4}$$

The final equality holds because $\boldsymbol{y}'$ is a $(2\rho-1)$-correlated version of $\boldsymbol{y}$. By Proposition 5.3, so long as $2\rho - 1 \ge 0$, the noise stability term in equation (4) is maximized when f is a dictator, in which case its noise stability is $(2\rho-1)$. This means that $\mathbf{Pr}[f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}})] \le 1/2 + \rho/2$, which is the upper bound we wanted. Furthermore, by Proposition 5.3, when $1/2 < \rho < 1$, this value is attained only by dictators and negated dictators, because they uniquely maximize the noise stability term. □

5.2 NICD with blocks

In the last section, we talked about the natural NICD test in which the functions do not have blocks. Unfortunately, due to the hardness of approximation setting, when constructing tests, we often have no choice but to ensure that the functions and strings used are blocked. This is somewhat of a pain for the analysis, and it would be nice if we could just ignore the blocks and have our results go through anyway. In this section, we show that the results from Section 5.1 mostly still hold if NICD-$\textsc{Test}(\rho)$ had incorporated blocks, with one surprising exception.

Recall that a string $z \in (Z_q^d \cup \{*^d\})^K$ is thought of as containing K "blocks" of size d apiece, where the t-th block corresponds to those indices in the set $\{(t-1)d+1, \dots, td\}$.

Notation 5.4. For $t \in [K]$ we will write $z[t]$ for the t-th block of string $z \in (Z_q^d \cup \{*^d\})^K$.

Consider, for $0 < \rho < 1$, the NICD test NICD-$\textsc{Test}(\rho)_d$, which is a blocked version of NICD-$\textsc{Test}(\rho)$. In NICD-$\textsc{Test}(\rho)_d$, the input strings have K blocks of size d apiece. The test selects $\widetilde{\boldsymbol{x}} \in (\{-1,1\}^d \cup \{*^d\})^K$ such that each block $\widetilde{\boldsymbol{x}}[t]$ of $\widetilde{\boldsymbol{x}}$ is independently selected to be $*^d$ with probability $1-\rho$ and a uniformly random element of $\{-1,1\}^d$ otherwise. Then the $*$'s of $\widetilde{\boldsymbol{x}}$ are randomly "filled in" to form $\boldsymbol{x}$, and the test checks that $f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}})$. The following lemma shows that so long as we consider folded strategies, the two tests are essentially equivalent.

Lemma 5.5. $\mathrm{Opt}_{\mathrm{folded}}(\text{NICD-}\textsc{Test}(\rho)_d) \le \mathrm{Opt}_{\mathrm{folded}}(\text{NICD-}\textsc{Test}(\rho))$.

Proof. To avoid confusion, we will refer to the blocked strings selected in the NICD-$\textsc{Test}(\rho)_d$ as $\boldsymbol{x}$ and $\widetilde{\boldsymbol{x}}$, and the strings selected in the NICD-$\textsc{Test}(\rho)$ as $\boldsymbol{y}$ and $\widetilde{\boldsymbol{y}}$. For this proof we will use the set Z_2 in place of $\{-1,1\}$. For $z \in (Z_2^d \cup \{*^d\})^K$ and $c \in (Z_2 \cup \{*\})^K$, let $z + c$ denote the string in $(Z_2^d \cup \{*\}^d)^K$ whose t-th block equals $z[t] + (c_t, c_t, \dots, c_t)$. (In this proof we interpret $l + * = *$ for any $l \in Z_q$.)

Let (f,h) be an optimal strategy for the test NICD-$\textsc{Test}(\rho)_d$ in which f is folded. Draw $\widetilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ as in NICD-$\textsc{Test}(\rho)_d$. Let $\boldsymbol{w}$ be a uniformly random element of Z_2^K. Consider the strings $\boldsymbol{x} + \boldsymbol{w}$ and $\widetilde{\boldsymbol{x}} + \boldsymbol{w}$. It is clear that $(\boldsymbol{x} + \boldsymbol{w}, \widetilde{\boldsymbol{x}} + \boldsymbol{w})$ has the same distribution as $(\boldsymbol{x}, \widetilde{\boldsymbol{x}})$. Thus

$$\mathrm{Val}_{\text{NICD-}\textsc{Test}(\rho)_d}(f,h) = \mathbf{Pr}[f(\boldsymbol{x}+\boldsymbol{w}) = h(\widetilde{\boldsymbol{x}}+\boldsymbol{w})].$$

By the probabilistic method, there must exist a setting x to $\boldsymbol{x}$ for which

$$\mathbf{Pr}[f(\boldsymbol{x}+\boldsymbol{w}) = h(\widetilde{\boldsymbol{x}}+\boldsymbol{w}) \mid \boldsymbol{x} = x] \ge \mathrm{Val}_{\text{NICD-}\textsc{Test}(\rho)_d}(f,h).$$

Let us fix this x. We now define the strategy (f_x, h_x) for NICD-$\textsc{Test}(\rho)$ by setting $f_x(y) = f(x+y)$ and $h_x(\widetilde{y}) = h(x+\widetilde{y})$. Note that h_x fixes x even though x has no $*$'s. When $\boldsymbol{y}$ and $\widetilde{\boldsymbol{y}}$ are selected as in NICD-$\textsc{Test}(\rho)_d$, $(x+\boldsymbol{y}, x+\widetilde{\boldsymbol{y}})$ is distributed exactly as $(\boldsymbol{x}+\boldsymbol{w}, \widetilde{\boldsymbol{x}}+\boldsymbol{w})$ conditioned on $\boldsymbol{x} = x$. Thus, f_x and h_x pass the NICD-$\textsc{Test}(\rho)$ with probability

$$\mathbf{Pr}[f_x(\boldsymbol{y}) = h_x(\widetilde{\boldsymbol{y}})] = \mathbf{Pr}[f(\boldsymbol{x}+\boldsymbol{w}) = h(\widetilde{\boldsymbol{x}}+\boldsymbol{w}) \mid \boldsymbol{x} = x] \ge \mathrm{Val}_{\text{NICD-}\textsc{Test}(\rho)_d}(f,h),$$

which is equal to $\mathrm{Opt}_{\mathrm{balanced}}(\text{NICD-}\textsc{Test}(\rho)_d)$. It remains to be checked is that f' is folded, and this follows from f being folded. □

The balanced case: Crucially, when we end up with $f(x+\cdot)$ in the proof of Lemma 5.5, we know that it is folded since f is folded. However, this fails when trying to prove a similar result for $\mathrm{Opt}_{\mathrm{balanced}}$, because $f(x+\cdot)$ is not necessarily balanced, even if f is. Somewhat unexpectedly, this fails for a more fundamental reason, which is that the statement

$$\mathrm{Opt}_{\mathrm{balanced}}(\text{NICD-}\textsc{Test}(\rho)_d) \le \mathrm{Opt}_{\mathrm{balanced}}(\text{NICD-}\textsc{Test}(\rho))$$

is simply false. There are balanced protocols which pass the NICD-$\textsc{Test}(\rho)_d$ test with much higher probability than even the best balanced protocol passes the NICD-$\textsc{Test}(\rho)$ with.

The example we have of this is the "tribes" function $f : \{0,1\}^{dK} \to \{0,1\}$ of Ben-Or and Linial [BOL90]. For a given block size of d, let K be the nearest integer to $(\ln 2)2^d$, so that the expectation of tribes is approximately $1/2$. The tribes function is defined so that $f(x)$ is 1 whenever at least one of x's blocks is all 1's, and otherwise $f(x)$ is 0. The optimal predictor $h : \{0,1,*\}^{dK} \to \{0,1\}$ does the same thing: on input z, if one of z's blocks is all 1's, then $f(x)$ is certainly 1, so $h(z)$ outputs 1 as well. Otherwise, it outputs 0. (Strictly speaking, the tribes function as defined here is not balanced, but it can be made so by changing its values on some $o_d(1)$ fraction of inputs; this only changes the success probability by at most $o_d(1)$.)

The analysis of the tribes protocol is relatively simple: when $f(\boldsymbol{x}) = 0$, which happens with probability $1/2$, there are no blocks of 1's for h to receive, so $h(\widetilde{\boldsymbol{x}})$ will be 0 as well, and the two will equal. On the other hand, when $f(\boldsymbol{x}) = 1$, there are some blocks in $\boldsymbol{x}$ which are all 1's, and h receives each one independently with probability ρ. When there's only one, it outputs $h(\boldsymbol{x}) = 1$ with probability ρ, but when there are more than one, which happens with nonzero probability, it outputs 1 with probability strictly greater than ρ. Thus, the total success probability is strictly greater than $1/2 + \rho/2$, which is the success probability of the dictator strategy, which is optimal for NICD-$\textsc{Test}(\rho)$. Indeed, it is easy to calculate the exact success probability of tribes.

Proposition 5.6. *The tribes strategy passes the* NICD-$\textsc{Test}(\rho)_d$ *with probability*

$$1 - \left(1 - \frac{\rho}{2^d}\right)^K + \left(1 - \frac{1}{2^d}\right)^K.$$

As $K \approx (\ln 2)2^d$, the success probability of tribes in the limit as d approaches infinity is $3/2 - 1/2^\rho$.

At $\rho = 1/2$, the success probability of tribes is about 79.3%. In contrast, dictators succeed at $\rho = 1/2$ with probability exactly 75%. It is interesting to consider whether tribes is optimal in this setting, or if there is a protocol which does better.

5.3 $2\mathsf{Lin}(Z_2)$ hardness

In this section, we state and analyze the test which yields our hardness of approximation result for $2\mathsf{Lin}(Z_2)$. We will be interested only in folded functions, and thus we can get a hardness result for $2\mathsf{Lin}(Z_2)$ by designing a binary $=$-based test. In fact, our test is just a 2-party blocked version of the NICD-$\textsc{Test}(\frac{1}{2})$ in which the deletions are correlated across the parties.

Let π be the following distribution:

$$(\widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{b}}) \sim \pi \text{ means } (\widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{b}}) = \begin{cases} \text{uniformly random from } \{-1,1\}^d \times \{*^d\}, & \text{with probability } \frac{1}{2}; \\ \text{uniformly random from } \{*^d\} \times \{-1,1\}^d, & \text{with probability } \frac{1}{2}. \end{cases} \tag{5}$$

Recall from Definition 4.4 the distribution μ^d on $\{-1,1\}^d \times \{*\}^d$ which is $*^d$ with probability $1/2$ and a uniformly random element of $\{-1,1\}^d$ otherwise. Then π's marginals are both μ^d, however correlated. The test is:

$\boxed{2\mathsf{Lin}(Z_2)\text{-}\textsc{Test}}$

- Given functions $f : \{-1,1\}^{dK} \to \{-1,1\}$, $g : \{-1,1\}^{dK} \to \{-1,1\}$, and $h : \{-1,1,*\}^{dK} \times \{-1,1,*\}^{dK} \to \{-1,1\}$:

- Draw $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \sim \pi^{\otimes K}$.
- Form $\boldsymbol{x} \in \{-1,1\}^{dK}$ by replacing each $*$ of $\widetilde{\boldsymbol{x}}$ with a uniformly random element of $\{-1,1\}$. Form $\boldsymbol{y} \in \{-1,1\}^{dK}$ from $\widetilde{\boldsymbol{y}}$ similarly.
- Test either $f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ or $g(\boldsymbol{y}) = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$, each with equal probability.

Remark 5.7. For this test, the block sizes are unimportant so we have chosen $d_1 = d_2 = d$ for simplicity. Thus our NP-hardness result for $2\mathsf{Lin}(Z_2)$ does not even need to reduce from Label-Cover with "projection constraints"; any $\mathsf{LC}_{d_1 K, d_2 K}$ would suffice.

Observation 5.8. It is helpful to also think about the strings in $2\mathsf{Lin}(Z_2)$-Test being generated in the opposite order, as described in Section 1.2. By this we mean the following viewpoint: First, $\boldsymbol{x}, \boldsymbol{y} \sim \{-1,1\}^{dK}$ are chosen *independently* and uniformly at random. Then $\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}} \in (\{-1,1\}^d \cup \{*^d\})^K$ are formed as follows: independently for each block $t \in [K]$, exactly one of $\boldsymbol{x}[t]$, $\boldsymbol{y}[t]$ is replaced with $*^d$, with probability $\frac{1}{2}$ each. For each block t we think of h as "knowing" either $\boldsymbol{x}[t]$ or $\boldsymbol{y}[t]$, and "knowing that it doesn't know" the other one.

For the "completeness" part of our hardness result, we compute the success probability of matching dictators:

Proposition 5.9. *Matching dictators achieve success probability $\frac{3}{4}$ in $2\mathsf{Lin}(Z_2)$-Test.*

Proof. Suppose $f, g : \{-1,1\}^{dK} \to \{-1,1\}$ are of the form $f(x) = x_i$, $g(y) = y_j$, where i, j are both in the t-th block (i.e., $(t-1)d < i, j \le td$). Let $h(\widetilde{x}, \widetilde{y})$ be the following (optimal) function: if $\widetilde{x}[t] \neq *^d$ then h outputs $\widetilde{x}_i$; if $\widetilde{y}[t] \neq *^d$ then h outputs $\widetilde{y}_j$ (one of these two always holds). Half of the time h is tested against the function (f or g) whose output it "knows"; then it succeeds with probability 1. The other half of the time h is tested against the function whose output it doesn't know; in this case it succeeds with probability $\mathbf{Pr}[\boldsymbol{x}_i = \boldsymbol{y}_j] = \frac{1}{2}$. Thus the overall success probability is $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$. □

Let's now move on to the "soundness" part of our hardness result. For this Theorem 4.19 tells us we need to analyze the optimal success probability in the uncorrelated version of $2\mathsf{Lin}(Z_2)$-Test. But before we do this, let's informally think about how well functions f and g with "no influential coordinates in common" can pass $2\mathsf{Lin}(Z_2)$-Test. The usual candidates to consider are non-matching dictators and $f = g =$ Majority. In the latter case, it can be shown that the success probability is exactly $\frac{2}{3}$ by a noise stability calculation. As for the former case:

Fact 5.10. *Non-matching dictators achieve success probability $\frac{11}{16}$ in $2\mathsf{Lin}(Z_2)$-Test.*

Proof. Suppose now that $f(x) = x_i$, $g(y) = y_j$ where i and j are in *different* blocks. In this case the optimal h acts as follows: If it knows *either* x_i or y_j, it outputs that value. (If it knows both, it can output either.) This happens with probability $\frac{3}{4}$, and when it happens the success probability is again $\frac{3}{4}$. But when h knows neither x_i or y_j, h can only guess a label. This happens with probability $\frac{1}{4}$, and when it happens the success probability is only $\frac{1}{2}$. Thus the overall success probability of nonmatching dictators is $\frac{3}{4} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{2} = \frac{11}{16}$. □

These observations suggest that the hardness reduction for $2\mathsf{Lin}(Z_2)$ will achieve soundness $\frac{11}{16}$. Formally, we must proceed using Invariance and Theorem 4.19. To that end, consider $2\mathsf{Lin}(Z_2)$-Test$'$, the uncorrelated version of $2\mathsf{Lin}(Z_2)$-Test. In $2\mathsf{Lin}(Z_2)$-Test$'$, $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ is sampled from $(\mu_2^d \times \mu_2^d)^{\otimes K}$ rather than $\pi^{\otimes K}$. If f and g are dictators, they are accepted by $2\mathsf{Lin}(Z_2)$-Test$'$ with probability $\frac{11}{16}$. Our next theorem shows that this is optimal.

Theorem 5.11. $\mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_2)\text{-}\textsc{Test}') = \frac{11}{16}$.

Proof. By an extension of Lemma 5.5, it suffices to consider the case when $d = 1$. Let (f, g, h) be any strategy where f and g are folded. Consider selecting two strings $\boldsymbol{x}, \boldsymbol{x}' \in \{-1, 1\}^n$ independently and uniformly at random, and forming $\widetilde{\boldsymbol{x}}$ as follows:

$$\widetilde{\boldsymbol{x}}_i = \begin{cases} \boldsymbol{x}_i & \text{if } \boldsymbol{x}_i = \boldsymbol{x}'_i, \\ * & \text{if } \boldsymbol{x}_i \neq \boldsymbol{x}'_i. \end{cases}$$

Then clearly $\widetilde{\boldsymbol{x}}$ is distributed according to $(\mu)^{\otimes n}$ (recall the definition of μ from Definition 4.4), and both $\boldsymbol{x}$ and $\boldsymbol{x}'$ are distributed as random "filled-in" versions of $\widetilde{\boldsymbol{x}}$. In addition, consider selecting two more strings $\boldsymbol{y}, \boldsymbol{y}' \in \{-1, 1\}^n$ independently and uniformly at random, and forming $\widetilde{\boldsymbol{y}}$ analogously to $\widetilde{\boldsymbol{x}}$. It is also clear that $\widetilde{\boldsymbol{y}}$ is distributed according to $(\mu)^{\otimes n}$, and both $\boldsymbol{y}$ and $\boldsymbol{y}'$ are distributed as random "filled-in" versions of $\widetilde{\boldsymbol{w}}$. Thus,

$$\begin{aligned} \mathrm{Val}_{2\mathsf{Lin}(Z_2)\text{-}\textsc{Test}'}(f, g, h) &= \mathrm{avg}\{\mathbf{Pr}[f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \mathbf{Pr}[g(\boldsymbol{y}) = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})]\} \\ &= \mathrm{avg}\left\{ \begin{array}{l} \mathbf{Pr}[f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[f(\boldsymbol{x}') = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[g(\boldsymbol{y}) = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[g(\boldsymbol{y}') = h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})] \end{array} \right\} \leq \frac{1}{4}\,\mathbf{E}[M(f(\boldsymbol{x}), f(\boldsymbol{x}'), g(\boldsymbol{y}), g(\boldsymbol{y}'))], \end{aligned}$$

where M outputs the number of input bits in the majority. This is because whatever $f(\boldsymbol{x})$, $f(\boldsymbol{x}')$, $g(\boldsymbol{y})$, and $g(\boldsymbol{y}')$ turn out to be, $h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ can only agree with at most $M(f(\boldsymbol{x}), f(\boldsymbol{x}'), g(\boldsymbol{y}), g(\boldsymbol{y}'))$ of them.

By construction, $\boldsymbol{x}$, $\boldsymbol{x}'$, $\boldsymbol{y}$, and $\boldsymbol{y}'$ are all independent of each other. We require f and g to be folded, so in particular both are balanced. As a result, $f(\boldsymbol{x})$, $f(\boldsymbol{x}')$, $g(\boldsymbol{y})$, and $g(\boldsymbol{y}')$ are distributed as independent uniformly random ± 1 bits. The question remains to find the expected size of the majority of four random ± 1 bits, and it is easily verified that this is $\frac{11}{4}$. Thus, the probability that (f, g, h) passes the test is no greater than $\frac{11}{16}$. □

We invite the reader to compare this proof with the proof of Theorem 5.1 in the case of $\rho = 1/2$.

By using Proposition 5.9 and Theorem 5.11 together with Corollary 4.20, we obtain the $q = 2$ case of our Main Theorem (Theorem 2.5); in other words, NP-hardness of $(\frac{3}{4}, \frac{11}{16} + \epsilon)$-deciding $2\mathsf{Lin}(Z_2)$.

6 A test for $2\mathsf{Lin}(Z_q)$

In this section we state and analyze the test which yields our Main Theorem on the inapproximability of $2\mathsf{Lin}(Z_q)$. At a high level, this section takes everything from Section 5 and replaces every instance of Z_2 and $\{-1, 1\}$ with Z_q. Many of the intuitions and facts in the binary case—such as the equivalence of blocked and nonblocked tests for folded functions—carry over to the larger alphabet case basically verbatim. On the other hand, upper-bounding the value of the new q-ary tests requires new ideas.

6.1 q-ary NICD

Section 5.1 introduced the binary NICD model, in which the strings have characters in $Z_2 \cup \{*\}$ (equivalently, $\{-1, 1, *\}$) and the functions output values in Z_2. Let q-NICD-$\textsc{Test}(\rho)$ be the natural q-ary version of NICD-$\textsc{Test}(\rho)$. We will mainly analyze q-NICD-$\textsc{Test}(\rho)$ for the value of $\rho = \frac{1}{2}$; in this case we write q-NICD-$\textsc{Test}$ for short. It is easy to see that f passes the q-NICD-$\textsc{Test}$ with probability $\frac{1}{2} + \frac{1}{2q}$ if it is a dictator. The next theorem shows this is optimal.

Theorem 6.1. $\mathrm{Opt}_{\mathrm{folded}}(q\text{-NICD-Test}) \leq \frac{1}{2} + \frac{1}{2q}$.

We will prove Theorem 6.1 shortly; we actually require it later for the *soundness* part of our $2\mathsf{Lin}(Z_q)$ hardness result. As discussed in Sections 3.1 and 5.1, Theorem 6.1 resolves the q-ary (folded) generalization of Yang's conjecture on ρ-erasure NICD for $\rho = \frac{1}{2}$. For $q > 2$ we were unable to generalize Theorem 6.1 to any other value of $\rho \in (0,1)$. A potentially complicating factor is the surprising fact that even for $q = 3$, there are folded functions $f : Z_q^K \to Z_q$ which depend on *more* than one coordinate yet achieve $\mathrm{Val}_{q\text{-NICD-Test}(\rho)}(f) = \frac{1}{q} + \rho(1 - \frac{1}{q})$, the same value achieved by dictators. For example, the function $f : Z_3^2 \to Z_3$ defined by $f(a,a) = a$, $f(a,a-1) = a$, $f(a,a+1) = a+1$ is folded and it is easy to check that it succeeds (using optimal h) with probability $\frac{1}{3} + \frac{2}{3}\rho$. More generally, the function $f : Z_q^2 \to Z_q$ defined by "$f(a,b) = a$ if $a - b \pmod q \in \{0, 1, \dots, \lceil q/2 \rceil - 1\}$ else $f(a,b) = b$" is folded, and it can be checked that it has value $\frac{1}{q} + \rho(1 - \frac{1}{q})$.

We now give the proof of Theorem 6.1:

Proof of Theorem 6.1. We will in fact prove something stronger than Theorem 6.1, namely that $\mathrm{Val}_{q\text{-NICD-Test}}(f) \leq \frac{1}{2} + \frac{1}{2q}$ even in the case when f and h are allowed to be randomized (i.e., map into $\triangle_q$). The proof is actually somewhat simpler if we disallow randomized f and h, but we will need the generalization later.

So let $f : Z_q^K \to \triangle_q$ be folded. It's easy to see that it is optimal for $h : (Z_q \cup \{*\})^K \to \triangle_q$ to work as follows: on input $\widetilde{x}$ it computes

$$\nu = \mathbf{E}[f(\boldsymbol{x}) \mid \widetilde{\boldsymbol{x}} = \widetilde{x}] \in \triangle_q$$

and then outputs a uniformly random coordinate $\ell \in Z_q$ from among those which maximize ν_ℓ. Let us henceforth fix this optimal h, and also observe that it is "folded" in the sense that $h(\widetilde{x} + (c, \dots, c)) = \mathrm{rot}_c(h(\widetilde{x}))$ (where as before $* + c = *$).

Let us make an aside which will be useful for a future proof. Introducing the short-form notation $f(\widetilde{x}) = \mathbf{E}[f(\boldsymbol{x}) \mid \widetilde{\boldsymbol{x}} = \widetilde{x}]$, we have that the success probability conditioned on $\widetilde{\boldsymbol{x}} = \widetilde{x}$ is precisely $\|f(\widetilde{x})\|_\infty$. Thus

$$\mathrm{Val}_{q\text{-NICD-Test}}(f,h) = \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{x}}}[\|f(\widetilde{\boldsymbol{x}})\|_\infty]. \tag{6}$$

This proof shows that the above quantity is at most $\frac{1}{2} + \frac{1}{2q}$.

Let us return to the original formulation of value:

$$\mathrm{Val}_{q\text{-NICD-Test}}(f,h) = \mathop{\mathbf{Pr}}_{\boldsymbol{x},\widetilde{\boldsymbol{x}}}[f(\boldsymbol{x}) = h(\widetilde{\boldsymbol{x}})] = \mathop{\mathbf{E}}_{\boldsymbol{x},\widetilde{\boldsymbol{x}}}[\langle f(\boldsymbol{x}), h(\widetilde{\boldsymbol{x}})], \tag{7}$$

the latter equality because f and h are in fact randomized ($\triangle_q$-valued) functions. We now employ the following trick. Let $(\widetilde{\boldsymbol{z}}, \widetilde{\boldsymbol{z}}') \in (Z_q \cup \{*\})^K \times (Z_q \cup \{*\})^K$ be generated according to $\pi^{\otimes K}$, where π is defined as in (9) (with $d = 1$). I.e., for each coordinate $t \in [K]$, one of $\widetilde{\boldsymbol{z}}_t, \widetilde{\boldsymbol{z}}'_t$ is $*$ and the other is random from Z_q. Define the "composite string" denoted $\boldsymbol{z} = \widetilde{\boldsymbol{z}} \circ \widetilde{\boldsymbol{z}}' \in Z_q^K$, meaning that $\boldsymbol{z}_t$ equals the non-$*$ value among $\widetilde{\boldsymbol{z}}_t, \widetilde{\boldsymbol{z}}'_t$, for each $t \in [K]$. The key observation is that the pair $(\boldsymbol{z}, \widetilde{\boldsymbol{z}})$ has the same distribution as $(\boldsymbol{x}, \widetilde{\boldsymbol{x}})$ and that $(\boldsymbol{z}, \widetilde{\boldsymbol{z}}')$ *also* has the same distribution as $(\boldsymbol{x}, \widetilde{\boldsymbol{x}})$. Thus upon inserting either pair into (7) we get the same number, and hence the average of the two numbers

is again the same:

$$\begin{aligned}\mathrm{Val}_{q\text{-NICD-Test}}(f,h) &= \mathrm{avg}\Big\{\mathop{\mathbf{E}}_{\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{z}}',\boldsymbol{z}}[\langle f(\boldsymbol{z}),h(\widetilde{\boldsymbol{z}})\rangle], \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{z}}',\boldsymbol{z}}[\langle f(\boldsymbol{z}),h(\widetilde{\boldsymbol{z}}')\rangle]\Big\}\\ &= \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{z}}',\boldsymbol{z}}[\langle f(\boldsymbol{z}),\tfrac{h(\widetilde{\boldsymbol{z}})+h(\widetilde{\boldsymbol{z}}')}{2}\rangle]\\ &\le \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{z}}'}\Big[\|\tfrac{h(\widetilde{\boldsymbol{z}})+h(\widetilde{\boldsymbol{z}}')}{2}\|_\infty\Big]. \qquad (8)\end{aligned}$$

Here the last step uses the fact that $\langle \alpha,\beta\rangle \le \|\alpha\|_1\|\beta\|_\infty = \|\beta\|_\infty$ for $\alpha,\beta\in\triangle_q$.

Suppose that in addition to $\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{z}}'$ we also choose $\boldsymbol{\ell}\sim Z_q$ uniformly at random. Even conditioned on $\widetilde{\boldsymbol{z}}'$, the distributions of $\widetilde{\boldsymbol{z}}$ and $\widetilde{\boldsymbol{z}}+(\boldsymbol{\ell},\dots,\boldsymbol{\ell})$ are the same. Hence (8) equals

$$\mathop{\mathbf{E}}_{\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{z}}',\boldsymbol{\ell}}\Big[\|\tfrac{h(\widetilde{\boldsymbol{z}}+(\boldsymbol{\ell},\dots,\boldsymbol{\ell}))+h(\widetilde{\boldsymbol{z}}')}{2}\|_\infty\Big] = \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{z}}',\boldsymbol{\ell}}\Big[\|\tfrac{\mathrm{rot}_{\boldsymbol{\ell}}(h(\widetilde{\boldsymbol{z}}))+h(\widetilde{\boldsymbol{z}}')}{2}\|_\infty\Big],$$

using the foldedness of h. We will bound this by $\frac{1}{2}+\frac{1}{2q}$ for *every* pair of outcomes $h(\widetilde{\boldsymbol{z}})=\gamma$, $h(\widetilde{\boldsymbol{z}}')=\gamma'$ in $\triangle_q$. The quantity to bound, $\mathbf{E}_{\boldsymbol{\ell}}[\|(\mathrm{rot}_{\boldsymbol{\ell}}(\gamma)+\gamma')/2\|_\infty]$, is a (separately) convex function of both γ and γ' (since $\|\cdot\|_\infty$ is a norm). Hence the quantity is maximized when γ and γ' are extreme points of $\triangle_q$. But for $\gamma=e_\ell$, $\gamma=e_{\ell'}$, say, one immediately calculates that $\mathbf{E}_{\boldsymbol{\ell}}[\|(\mathrm{rot}_{\boldsymbol{\ell}}(\gamma)+\gamma')/2\|_\infty]=\frac{1}{2}+\frac{1}{2q}$. This completes the proof. □

6.2 $2\mathsf{Lin}(Z_q)$ hardness

We now state our $2\mathsf{Lin}(Z_q)$ test. As before, our hardness results will ensure that the functions involved are folded, and so we are free to design an $=$-based test. First, we generalize the distribution π from Section 5.3 to the larger alphabet size by defining π_q to be the following distribution:

$$(\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}})\sim\pi_q \text{ means } (\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}}) = \begin{cases}\text{uniformly random from } Z_q^d\times\{*^d\}, & \text{with probability } \frac{1}{2};\\ \text{uniformly random from } \{*^d\}\times Z_q^d, & \text{with probability } \frac{1}{2}.\end{cases} \qquad (9)$$

Recall from Definition 4.4 the distribution μ_q^d on $Z_q^d\times\{*\}^d$ which is $*^d$ with probability $1/2$ and a uniformly random element of Z_q^d otherwise. Then π_q's marginals are both μ_q^d, however correlated. The test is:

$\boxed{2\mathsf{Lin}(Z_q)\text{-Test}}$

- Given functions $f:Z_q^{dK}\to Z_q$, $g:Z_q^{dK}\to Z_q$, and $h:(Z_q\cup\{*\})^{dK}\to Z_q$:
- Draw $(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\sim\pi_q^{\otimes K}$.
- Form $\boldsymbol{x}\in Z_q^{dK}$ by replacing each $*$ of $\widetilde{\boldsymbol{x}}$ with a uniformly random element of Z_q. Form $\boldsymbol{y}\in Z_q^{dK}$ from $\widetilde{\boldsymbol{y}}$ similarly.
- Test either $f(\boldsymbol{x})=h(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})$ or $g(\boldsymbol{y})=h(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})$, each with equal probability.

Many of the statements made in Section 5.3 concerning the $2\mathsf{Lin}(Z_2)$-Test apply here too. See Remark 5.7 and Observation 5.8.

For the completeness part of our hardness result, we compute the probability that matching dictators succeed. This can be seen by a proof analogous to the one used for Proposition 5.9.

Proposition 6.2. *Matching dictators achieve success probability* $\frac{1}{2} + \frac{1}{2q}$ *in* $2\mathsf{Lin}(Z_q)$-Test.

The intuition behind the soundness analysis is similar to that for the $2\mathsf{Lin}(Z_2)$-Test. As before, let's consider the canonical examples of functions f and g with "no influential coordinates in common": non-matching dictators and $f = g =$ Plurality. The latter case actually has very low success probability, roughly $q^{-3+2\sqrt{2}}$. As for the former case, an analysis similar to Fact 5.10 shows:

Fact 6.3. *Non-matching dictators achieve success probability* $\frac{3}{8} + \frac{5}{8q}$ *in* $2\mathsf{Lin}(Z_q)$-Test.

These observations suggest that the hardness reduction for $2\mathsf{Lin}(Z_q)$ will achieve soundness $\frac{3}{8} + \frac{5}{8q}$. Formally, we must proceed using Invariance and Theorem 4.19. To that end, consider $2\mathsf{Lin}(Z_q)$-Test$'$, the uncorrelated version of $2\mathsf{Lin}(Z_q)$-Test. In $2\mathsf{Lin}(Z_q)$-Test$'$, $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ is sampled from $(\mu_q^d \times \mu_q^d)^{\otimes K}$ rather than $\pi_q^{\otimes K}$. The goal for the remainder of this section is to upper-bound $\mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_q)\text{-Test}')$. We believe that dictators are the folded functions with highest success probability for $2\mathsf{Lin}(Z_q)$-Test$'$. An almost identical analysis to Fact 6.3 gives:

Fact 6.4. *Dictators achieve success probability* $\frac{3}{8} + \frac{5}{8q}$ *in* $2\mathsf{Lin}(Z_q)$-Test$'$.

Unfortunately, we are only able to prove that dictators are optimal for $q < 7$. The $q = 2$ case is Theorem 5.11; the cases $3 \le q \le 6$ require computer assistance, see Appendix E. We would not be surprised if there was a short proof giving the result for all q; however for $q \ge 7$ we have only proved an upper bound of $\frac{3}{8} + O(\frac{1}{q^{1/3}})$. We now proceed to obtain this bound:

Theorem 6.5. $\mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_q)\text{-Test}') \le \frac{3}{8} + O(\frac{1}{q^{1/3}})$.

Proof. By an extension of Lemma 5.5, it suffices to consider the case when $d = 1$. Again, we prove this even when f and g are allowed to be randomized. So let $f, g : Z_q^K \to \triangle_q$ be optimal folded functions for $2\mathsf{Lin}(Z_q)$-Test$'$. As before it's easy to see that the optimal $h : (Z_q \cup \{*\})^K \to \triangle_q$ works as follows: on input $(\widetilde{x}, \widetilde{y})$ it computes

$$\nu = \tfrac{1}{2}\,\mathbf{E}[f(\boldsymbol{x}) \mid \widetilde{\boldsymbol{x}} = \widetilde{x}] + \tfrac{1}{2}\,\mathbf{E}[g(\boldsymbol{y}) \mid \widetilde{\boldsymbol{y}} = \widetilde{y}] = \tfrac{1}{2} f(\widetilde{x}) + \tfrac{1}{2} g(\widetilde{y})$$

and then outputs a uniformly random coordinate $\ell \in Z_q$ from among those which maximize ν_ℓ. With this h the success probability conditioned on $\widetilde{\boldsymbol{x}} = \widetilde{x}, \widetilde{\boldsymbol{y}} = \widetilde{y}$ is precisely $\|\nu\|_\infty$. Thus

$$\mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_q)\text{-Test}') = \mathrm{Val}_{2\mathsf{Lin}(Z_q)\text{-Test}'}(f, g) = \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}}[\tfrac{1}{2}\|f(\widetilde{\boldsymbol{x}}) + g(\widetilde{\boldsymbol{y}})\|_\infty].$$

Note that $f(\widetilde{\boldsymbol{x}})$ and $g(\widetilde{\boldsymbol{y}})$ are independent $\triangle_q$-valued random variables. Further, by virtue of the fact that f is folded, $f(\widetilde{\boldsymbol{x}})$ is "cyclically symmetric" in the sense that $\mathrm{rot}_\ell(f(\widetilde{\boldsymbol{x}}))$ has the same distribution for each $\ell \in Z_q$. (The same is true of $g(\widetilde{\boldsymbol{y}})$ but we won't need this.) Thus we may also write

$$\mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_q)\text{-Test}') = \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}}[\operatorname*{avg}_{\ell \in Z_q}\{\tfrac{1}{2}\|\mathrm{rot}_\ell(f(\widetilde{\boldsymbol{x}})) + g(\widetilde{\boldsymbol{y}})\|_\infty\}]. \tag{10}$$

Define $m(s, t) = s + t - st$. The key to our proof is showing that

$$\operatorname*{avg}_{\ell \in Z_q}\{\tfrac{1}{2}\|\mathrm{rot}_\ell(\sigma) + \tau\|_\infty\} \le \tfrac{1}{2} m(\|\sigma\|_\infty, \|\tau\|_\infty) + O(\tfrac{1}{q^{1/3}}) \tag{11}$$

holds for every $\sigma, \tau \in \triangle_q$. From this we can easily complete the proof: applying it to (10) gives

$$\begin{aligned}\mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_q)\text{-Test}') &\le \tfrac{1}{2} \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}}[m(\|f(\widetilde{\boldsymbol{x}})\|_\infty, \|g(\widetilde{\boldsymbol{y}})\|_\infty)] + O(\tfrac{1}{q^{1/3}}) \\ &= \tfrac{1}{2} m\left(\mathop{\mathbf{E}}_{\widetilde{\boldsymbol{x}}}[\|f(\widetilde{\boldsymbol{x}})\|_\infty], \quad \mathop{\mathbf{E}}_{\widetilde{\boldsymbol{y}}}[\|g(\widetilde{\boldsymbol{y}})\|_\infty]\right) + O(\tfrac{1}{q^{1/3}}),\end{aligned}$$

where we were able to pass the expectations inside the $m(\cdot,\cdot)$ using the fact that $f(\widetilde{\boldsymbol{x}})$ and $g(\widetilde{\boldsymbol{y}})$ are independent. But by virtue of (6) in the proof of Theorem 6.1, and by the theorem's result itself, we have $\mathbf{E}_{\widetilde{\boldsymbol{x}}}[\|f(\widetilde{\boldsymbol{x}})\|_\infty], \mathbf{E}_{\widetilde{\boldsymbol{y}}}[\|g(\widetilde{\boldsymbol{y}})\|_\infty] \le \frac{1}{2} + \frac{1}{2q}$. Since $m(s,t)$ is an increasing function of $s,t \in [0,1]$, we deduce that

$$\mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_q)\text{-}\textsc{Test}') \le \tfrac{1}{2} m(\tfrac{1}{2} + \tfrac{1}{2q}, \tfrac{1}{2} + \tfrac{1}{2q}) + O(\tfrac{1}{q^{1/3}}) = \tfrac{3}{8} + O(\tfrac{1}{q^{1/3}}),$$

as needed.

It remains then to prove (11). Let $\triangle_{q,s}$ denote the convex set $\triangle_q \cap \{\sigma : \|\sigma\|_\infty \le s\}$. We will in fact show

$$\operatorname*{avg}_{\ell \in Z_q} \{\tfrac{1}{2}\|\mathrm{rot}_\ell(\sigma) + \tau\|_\infty\} \le \tfrac{1}{2}\max(s,t) + O(\tfrac{1}{q^{1/3}}) \quad \text{for all } \sigma \in \triangle_{q,s},\ \tau \in \triangle_{q,t}; \tag{12}$$

this is stronger since $\max(s,t) \le m(s,t)$ for $s,t \in [0,1]$. Inequality (12) is obvious if either s or t is at most $\frac{1}{q^{1/3}}$; thus we need only be concerned with the case $s,t \ge \frac{1}{q^{1/3}}$.

As in the previous proof we observe that $\operatorname{avg}_{\ell \in Z_q}\{\frac{1}{2}\|\mathrm{rot}_\ell(\sigma)+\tau\|_\infty\}$ is convex in σ and in τ; hence it suffices to prove (12) when σ and τ are extreme points of $\triangle_{q,s}$ and $\triangle_{q,t}$, respectively. Note that an extreme point for $\triangle_{q,s}$ has exactly $\lceil 1/s \rceil$ nonzero coordinates, of which $\lfloor 1/s \rfloor$ equal s; similarly for $\triangle_{q,t}$. For fixed extreme σ and τ, a simple union-bound argument shows that the fraction of $\ell \in Z_q$ for which $\mathrm{rot}_\ell(\sigma)$ and τ have a nonzero coordinate in common is at most $\lceil 1/s \rceil \lceil 1/t \rceil / q$. This is at most $4/(stq) \le O(\frac{1}{q^{1/3}})$ since we are assuming $s,t \ge \frac{1}{q^{1/3}}$. On the other hand, if $\mathrm{rot}_\ell(\sigma)$ and τ do *not* have a nonzero coordinate in common, $\frac{1}{2}\|\mathrm{rot}_\ell(\sigma)+\tau\|_\infty = \frac{1}{2}\max(s,t)$. Together these facts justify (12), completing the proof. □

We may now plug our completeness result (Proposition 6.2) and our soundness result (Theorem 6.5 combined with Lemma 5.5) for $2\mathsf{Lin}(Z_q)$-Test into our hardness reduction (Corollary 4.20) to obtain our Main Theorem on the NP-hardness of Unique-Games (in its precise form, Theorem 2.5).

7 Other binary CSPs

In this section, we design and analyze binary 2-party tests which allow us to recover the best known NP-hardness results for Max-Cut, 2Sat, and 2And.

7.1 Max-Cut hardness

Our Max-Cut-Test is similar to our $2\mathsf{Lin}(Z_2)$-Test, except it can no longer guarantee that the functions involved are folded. This presents a problem, as without this guarantee, f and g could both, for example, be constantly 1, in which case the $2\mathsf{Lin}(Z_2)$-Test could be passed with probability 1. To handle this, our Max-Cut-Test devotes a certain fraction of its tests to ensuring that f and g "look" folded. This certain fraction is chosen so that the constantly 1 protocol is no better than the dictator protocol, and it turns out that this is sufficient for making the dictator protocol optimal. Let π be as in the $2\mathsf{Lin}(Z_2)$-Test. The test is:

Max-Cut-Test

- Given functions $f : \{-1,1\}^{dK} \to \{-1,1\}$, $g : \{-1,1\}^{dK} \to \{-1,1\}$, and $h : \{-1,1,*\}^{dK} \times \{-1,1,*\}^{dK} \to \{-1,1\}$, in which f and g are folded:

- Draw $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \sim \pi^{\otimes K}$.
- Form $\boldsymbol{x} \in \{-1,1\}^{dK}$ by replacing each $*$ of $\widetilde{\boldsymbol{x}}$ with a uniformly random element of $\{-1,1\}$. Form $\boldsymbol{y} \in \{-1,1\}^{dK}$ from $\widetilde{\boldsymbol{y}}$ similarly.
- With probability 16/21, test either $f(\boldsymbol{x}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ or $g(\boldsymbol{y}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$, each with equal probability.
- With the remaining 5/21 probability, test either that $f(\boldsymbol{x}) \neq f(-\boldsymbol{x})$ or $g(\boldsymbol{y}) \neq g(-\boldsymbol{y})$, each with equal probability.

In other words, the Max-Cut-Test runs the 2Lin(Z_2)-Test (with inequalities) with probability 16/21, and tests folding with the remaining probability. The Max-Cut-Test doesn't fit tidily in the 2-party correlated test framework we developed in previous sections, because such tests must only test constraints which involve h. However, a natural modification to this framework, which we omit for simplicity, does cover the Max-Cut-Test. Intuitively, the "folding tests" that the Max-Cut-Test performs with probability 5/21 don't involve h or the correlated distribution π; thus, when decoupling the correlation in π, we may simply ignore these folding tests. As a result, Theorem 4.18 still holds for the Max-Cut-Test.

The following fact is easy to check:

Fact 7.1. *Let f and g be matching dictators. Then* $\mathrm{Val}_{\text{Max-Cut-Test}}(f,g) = 17/21$.

This follows from Proposition 5.9: since dictators always pass the folding test, their overall success probability is $16/21 \cdot 12/16 + 5/21 \cdot 1 = 17/21$.

Consider Max-Cut-Test$'$, the uncorrelated version of the Max-Cut-Test. In Max-Cut-Test$'$, $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ is sampled from $(\mu^d \times \mu^d)^{\otimes K}$ rather than $\pi^{\otimes K}$. If f and g are dictators, they are accepted by T' with probability 16/21. Our next theorem shows that this is optimal.

Theorem 7.2. *Let* Max-Cut-Test$'$ *be the uncorrelated version of the* Max-Cut-Test. *Then we have* Opt(Max-Cut-Test$'$) $= 16/21$.

Proof. By an argument similar to Lemma 5.5, it suffices to consider the case when $d = 1$. Let (f,g,h) be any strategy. Say that $\theta = \mathbf{E}[f(x)]$ and $\nu = \mathbf{E}[g(y)]$. Consider selecting two strings $\boldsymbol{x}, \boldsymbol{x}' \in \{-1,1\}^n$ independently and uniformly at random, and forming $\widetilde{\boldsymbol{x}}$ as follows:

$$\widetilde{\boldsymbol{x}}_i = \begin{cases} \boldsymbol{x}_i & \text{if } \boldsymbol{x}_i = \boldsymbol{x}'_i, \\ * & \text{if } \boldsymbol{x}_i \neq \boldsymbol{x}'_i. \end{cases}$$

Then clearly $\widetilde{\boldsymbol{x}}$ is distributed according to $(\mu)^{\otimes n}$, and both $\boldsymbol{x}$ and $\boldsymbol{x}'$ are distributed as random "filled-in" versions of $\widetilde{\boldsymbol{x}}$. In addition, consider selecting two more strings $\boldsymbol{y}, \boldsymbol{y}' \in \{-1,1\}^n$ independently and uniformly at random, and forming $\widetilde{\boldsymbol{y}}$ analogously to $\widetilde{\boldsymbol{x}}$. It is also clear that $\widetilde{\boldsymbol{y}}$ is distributed according to $(\mu)^{\otimes n}$, and both $\boldsymbol{y}$ and $\boldsymbol{y}'$ are distributed as random "filled-in" versions of $\widetilde{\boldsymbol{w}}$. Thus

$$\begin{aligned} \mathrm{Val}_{\text{Max-Cut-Test}'}(f,g,h) &= \mathrm{avg}\{\mathbf{Pr}[f(\boldsymbol{x}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \mathbf{Pr}[g(\boldsymbol{y}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})]\} \\ &= \mathrm{avg}\left\{ \begin{array}{l} \mathbf{Pr}[f(\boldsymbol{x}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[f(\boldsymbol{x}') \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[g(\boldsymbol{y}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[g(\boldsymbol{y}') \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})] \end{array} \right\} \leq \frac{1}{4}\,\mathbf{E}[M(f(\boldsymbol{x}), f(\boldsymbol{x}'), g(\boldsymbol{y}), g(\boldsymbol{y}'))], \quad (13) \end{aligned}$$

where M outputs the number of input bits in the majority. This is because whatever $f(\boldsymbol{x})$, $f(\boldsymbol{x}')$, $g(\boldsymbol{y})$, and $g(\boldsymbol{y}')$ turn out to be, $h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ can only disagree with at most $M(f(\boldsymbol{x}), f(\boldsymbol{x}'), g(\boldsymbol{y}), g(\boldsymbol{y}'))$ of them.

By construction, $\boldsymbol{x}$, $\boldsymbol{x}'$, $\boldsymbol{y}$, and $\boldsymbol{y}'$ are all independent of each other. Thus, $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$ are distributed as mean-θ ± 1 bits, $g(\boldsymbol{y})$ and $g(\boldsymbol{y}')$ are distributed as mean-ν ± 1 bits, and all four are independent of each other. The function M can be written as

$$M(u, u', v, v') = \frac{11}{16} + \frac{1}{16}(uu' + uv + \ldots + u'v' + vv') - \frac{uu'vv'}{16}.$$

We can therefore calculate the expectation in Equation (13) as

$$\mathbf{E}[M(f(\boldsymbol{x}), f(\boldsymbol{x}'), g(\boldsymbol{y}), g(\boldsymbol{y}'))] = \frac{11}{16} + \frac{\theta^2}{16} + \frac{\theta\nu}{4} + \frac{\nu^2}{16} - \frac{\theta^2\nu^2}{16}.$$

The only term in this expression which depends on the signs of θ and ν is $\theta\nu/4$, and it is easy to see that it is maximized when θ has the same sign as ν. Thus, it can only improve the success probability of a protocol to assume that θ and ν have the same sign. By symmetry, we may assume that they are positive.

Given that the mean of f is θ, the highest chance it can pass the folding test is $1 - |\theta| = 1 - \theta$. Similarly, the highest chance g can pass the folding test is $1 - \nu$. The overall probability of f and g passing the entire test is therefore at most

$$\begin{aligned} P(\theta, \nu) :&= \frac{16}{21}\left(\frac{11}{16} + \frac{\theta^2}{16} + \frac{\theta\nu}{4} + \frac{\nu^2}{16} - \frac{\theta^2\nu^2}{16}\right) + \frac{5}{21}\left(1 - \frac{\theta}{2} - \frac{\nu}{2}\right) \\ &= \frac{16}{21} - \frac{5}{42}(\theta + \nu) + \frac{\theta^2}{21} + \frac{4\theta\nu}{21} + \frac{\nu^2}{21} - \frac{\theta^2\nu^2}{21}. \end{aligned}$$

What remains is to show that on $[0, 1] \times [0, 1]$, P has maximal value $16/21$. As a function of θ,

$$P_\nu(\theta) := P(\theta, \nu) = \left(\frac{1}{21} - \frac{\nu^2}{21}\right) \cdot \theta^2 + \left(\frac{4\nu}{21} - \frac{5}{42}\right) \cdot \theta + \frac{\nu^2}{21} - \frac{5\nu}{42} + \frac{16}{21}.$$

The coefficient on θ^2 is always nonnegative, so $P_\nu(\theta)$ is a convex function of θ, for each fixed ν. Thus, for each fixed ν it's maximized at either $\theta = 0$ or $\theta = 1$. So it suffices to maximize $P(0, \nu)$ and $P(1, \nu)$. The latter is linear (and increasing) in ν, so $P(1, 1)$ is a possible maximizer. The former is a convex parabola, so it's maximized at $\nu = 0$ or $\nu = 1$. It remains to check the corners; indeed, $P(0, 0) = P(1, 1) = 16/21$ are the maximizers. □

Inspired by the test presented in this section, it is interesting to consider what happens in the model of NICD over the erasure channel presented in Section 5.1 if we drop the requirement that f be balanced and add a folding test to compensate. If one sets the probability of performing the original test at $4/5$ and the probability of performing the folding test at $1/5$, then the constant function protocol will no longer be better than the dictator protocol; both succeed with probability $4/5$. And just as in Theorem 7.2, though with different numbers, this sufficient for making the dictator protocol optimal.

By using Fact 7.1 and Theorem 7.2 together with Corollary 4.20, we obtain the Max-Cut case Theorem 2.4; in other words, NP-hardness of $(\frac{17}{21}, \frac{16}{21} + \epsilon)$-deciding Max-Cut.

7.2 2Sat hardness

The next two tests we present are quite different from the 2Lin(Z_2) and Max-Cut NICD tests. For example, both require starting from d-to-1 projective Label Cover. Even then, our analysis suggests that there is still an underlying similarity. Let ν be the following distribution:

$$(\boldsymbol{a}, \widetilde{\boldsymbol{b}}) \sim \nu \text{ means } (\boldsymbol{a}, \widetilde{\boldsymbol{b}}) = \begin{cases} \text{uniformly random from } \{1\} \times \{-1,1\}^d, & \text{with probability } \frac{1}{2}; \\ \text{uniformly random from } \{-1\} \times *^d, & \text{with probability } \frac{1}{2}. \end{cases} \tag{14}$$

Note that we have written $\boldsymbol{a}$ rather than $\widetilde{\boldsymbol{a}}$. This is because $\boldsymbol{a}$ is never $*$. The following is our 2Sat test:

2Sat-Test

- Given functions $f : \{-1,1\}^K \to \{-1,1\}$, $g : \{-1,1\}^{dK} \to \{-1,1\}$, and $h : \{-1,1\}^K \times \{-1,1,*\}^{dK} \to \{-1,1\}$:
- Sample $(\boldsymbol{x}, \widetilde{\boldsymbol{y}}) \sim \nu^{\otimes K}$.
- Form $\boldsymbol{y} \in \{-1,1\}^{dK}$ by replacing each $*$ of $\widetilde{\boldsymbol{y}}$ with a uniformly random element of $\{-1,1\}$.
- With probability 1/3, test $f(\boldsymbol{x}) \vee \neg h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$.
- With probability 2/3, test $g(\boldsymbol{y}) \vee h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$.

We will be concerned with folded protocols. Note that h is given the entire string $\boldsymbol{x}$; using the notation established earlier for 2-party correlation tests, we achieve this by setting $\widetilde{\boldsymbol{x}} = \boldsymbol{x}$.

Now, the following is an easy fact:

Fact 7.3. *Let f and g be matching dictators. Then* $\mathrm{Val}_{\text{2Sat-Test}}(f, g) = 11/12$.

Proof. When $f(\boldsymbol{x}) = -1$, $h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ can always output -1, satisfying all the constraints and passing the test with probability 1. Now, condition on $f(\boldsymbol{x}) = 1$. In this case, because f and g are matching dictators, h is given the value of $g(\boldsymbol{y})$. So if $g(\boldsymbol{y}) = -1$, setting $h(\boldsymbol{x}, \widetilde{\boldsymbol{y}}) = 1$ will satisfy all the constraints and pass the test with probability 1. Finally, when $g(\boldsymbol{y}) = 1$, setting $h(\boldsymbol{x}, \widetilde{\boldsymbol{y}}) = -1$ will satisfy the second constraint only, passing the test with probability 2/3. The result is a total success probability of $3/4 \cdot 1 + 1/4 \cdot 2/3 = 11/12$. □

Let (f, g, h) be a strategy in which f and g are folded. Consider the test 2Sat-Test$'$, the uncorrelated version of 2Sat-Test. In 2Sat-Test$'$, $(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ is sampled so that $\boldsymbol{x}$ is an independent uniformly random $\{-1,1\}^K$ string, and $\widetilde{\boldsymbol{y}}$ is drawn from $(\mu^d)^{\otimes K}$. If f and g are dictators, they pass the 2Sat-Test$'$ with probability 7//8. Our next theorem shows that this is optimal:

Theorem 7.4. $\mathrm{Opt}_{\text{folded}}(\text{2Sat-Test}') = 7/8$.

Proof. Conditioned on $f(\boldsymbol{x}) = -1$, an optimal h will always output -1, and will pass the test with probability 1. On the other hand, when $f(\boldsymbol{x}) = 1$ the test reduces to testing $\neg h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ with probability 1/3 and $g(\boldsymbol{y}) \vee h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ with probability 2/3. The probability of success is therefore

$$\begin{aligned} \frac{1}{3}\,\mathbf{Pr}[h(\boldsymbol{x}, \widetilde{\boldsymbol{y}}) = 1] + \frac{2}{3}\,\mathbf{Pr}[g(\boldsymbol{y}) \vee h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})] &= \frac{1}{3}\,\mathbf{E}\left[\frac{1 + h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2}\right] + \frac{2}{3}\,\mathbf{E}\left[1 - \left(\frac{1 + g(\boldsymbol{y})}{2}\right)\left(\frac{1 + h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2}\right)\right] \\ &= \frac{1}{3}\,\mathbf{E}\left[\frac{1}{2} + \frac{h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2} + 2 - \frac{1}{2} - \frac{g(\boldsymbol{y})}{2} - \frac{h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2} - \frac{g(\boldsymbol{y})h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2}\right] \\ &= \frac{5}{6} - \frac{1}{3}\,\mathbf{E}\left[\frac{1}{2} + \frac{g(\boldsymbol{y})h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2}\right], \end{aligned} \tag{15}$$

where we are able to drop the $-g(\boldsymbol{y})/2$ term because g is balanced. The expectation is equal to the probability that $g(\boldsymbol{y}) = h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$, which Theorem 5.1 says is no less than $1/4$ (it actually says it is no greater than $3/4$, but we may negate h). So when $f(\boldsymbol{x}) = 1$, we can upper bound the success probability of the test by

$$(15) \leq \frac{5}{6} - \frac{1}{3} \cdot \frac{1}{4} = \frac{3}{4}.$$

Since f is balanced, both of these cases happens with the same probability. Thus, we can upper bound the success probability of the f, g, and h by $1 \cdot 1/2 + 3/4 \cdot 1/2 = 7/8$. □

By using Fact 7.3 and Theorem 7.4 together with Corollary 4.20, we obtain the 2Sat case Theorem 2.4; in other words, NP-hardness of $(\frac{11}{12}, \frac{7}{8} + \epsilon)$-deciding 2Sat.

7.3 2And hardness

Recall the definition of ν from Equation (14). The following is our 2And test:

2And-Test

- Given functions $f : \{-1,1\}^K \to \{-1,1\}$, $g : \{-1,1\}^{dK} \to \{-1,1\}$, and $h : \{-1,1\}^K \times \{-1,1,*\}^{dK} \to \{-1,1\}$, in which f and g are folded:
- Sample $(\boldsymbol{x}, \widetilde{\boldsymbol{y}}) \sim \nu^{\otimes K}$.
- Form $\boldsymbol{y} \in \{-1,1\}^{dK}$ by replacing each $*$ of z with a uniformly random element of $\{-1,1\}$.
- With probability $1/3$, test $f(\boldsymbol{x}) \wedge h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$.
- With probability $1/3$, test $g(\boldsymbol{y}) \wedge h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$.
- With probability $1/3$, test $\neg g(\boldsymbol{y}) \wedge \neg h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$.

We will be concerned with folded protocols. Note that h is again given the entire string $\boldsymbol{x}$. Now, the following is an easy fact:

Fact 7.5. *Let f and g be matching dictators. Then* $\mathrm{Val}_{\text{2And-Test}}(f, g) = 10/24$.

Proof. When $f(\boldsymbol{x}) = -1$, setting $h(\boldsymbol{x}, \widetilde{\boldsymbol{y}}) = -1$ will pass the test with probability $1/3 + \mathbf{Pr}[g(\boldsymbol{y}) = -1]/3 = 1/2$ (it can be checked that this is the best of the two settings for $h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ in this case). Otherwise, when $f(\boldsymbol{x}) = 1$, because f and g are matching dictators, h is given the value of $g(\boldsymbol{y})$. No matter what $g(\boldsymbol{y})$ is, $h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ can always be set to satisfy exactly one of the three constraints. This is an overall success probability of $1/2 \cdot 1/2 + 1/2 \cdot 1/3 = 5/12$. □

Consider 2And-Test′, the uncorrelated version of the 2And-Test. In 2And-Test′, $(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ is sampled so that $\boldsymbol{x}$ is an independent uniformly random $\{-1,1\}^K$ string, and $\widetilde{\boldsymbol{y}}$ is drawn from $(\mu^d)^{\otimes K}$. If f and g are dictators, they pass the 2And-Test with probability $9/24$. Our next theorem shows that this is optimal:

Theorem 7.6. $\mathrm{Opt}_{\text{folded}}(\text{2And-Test}') = 9/24$.

Proof. Conditioned on $f(\boldsymbol{x}) = -1$, an optimal h will always output -1. The probability that doing so passes the test is $1/3+1/3\cdot\mathbf{Pr}[g(\boldsymbol{y}) = -1]$. Because g is balanced, this is $1/2$. On the other hand, when $f(\boldsymbol{x}) = 1$ the test reduces to testing $g(\boldsymbol{y}) \wedge h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ with probability $1/3$, $\neg g(\boldsymbol{y}) \wedge \neg h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$ with probability $1/3$, and rejecting outright with the remaining $1/3$ probability. This is

$$\begin{aligned}
&\frac{1}{3}\,\mathbf{Pr}[g(\boldsymbol{y}) \wedge h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})] + \frac{1}{3}\,\mathbf{Pr}[\neg g(\boldsymbol{y}) \wedge \neg h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})] \\
=&\frac{1}{3}\,\mathbf{E}\left[\left(\frac{1-g(\boldsymbol{y})}{2}\right)\left(\frac{1-h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2}\right)\right] + \frac{1}{3}\,\mathbf{E}\left[\left(\frac{1+g(\boldsymbol{y})}{2}\right)\left(\frac{1+h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2}\right)\right] \\
=&\frac{1}{3}\,\mathbf{E}\left[\frac{1}{2} + \frac{g(\boldsymbol{y})h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})}{2}\right]. \qquad (16)
\end{aligned}$$

The expectation is equal to the probability that $g(\boldsymbol{y}) = h(\boldsymbol{x}, \widetilde{\boldsymbol{y}})$, which Theorem 5.1 says is no more than $3/4$. So when $f(\boldsymbol{x}) = 1$, we can upper bound the success probability of the test by

$$(16) \leq \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}.$$

Since f is balanced, each of these cases happens with equal probability. Thus, we can upper bound the success probability of the test by $1/2 \cdot 1/2 + 1/4 \cdot 1/2 = 3/8$. □

By using Fact 7.5 and Theorem 7.6 together with Corollary 4.20, we obtain the 2And case Theorem 2.4; in other words, NP-hardness of $(\frac{5}{12}, \frac{3}{8} + \epsilon)$-deciding 2And.

8 Future directions

It seems reasonable that our hardness reductions could be converted into $n^{1-o(1)}$-round Lasserre SDP integrality gaps, using the methods of Tulsiani [Tul09]; we have not yet investigated this. This would give a new barrier for strongly approximating Unique-Games using Lasserre relaxations, a topic which has seen recent progress [BHK+11]. Relatedly, we believe our work motivates the problem of making the constants in recent subexponential-time algorithms for Unique-Games [ABS10, BRS11, GS11] more explicit.

Our new methodology seems to hold promise for improving the best known NP-hardness results for other notable 2-CSPs. Natural problems to try it on include Max-k-Cut and Khot's "2-to-1 problem". An intriguing further possibility would be to use the framework to improve the best known NP-hardness result for Metric-TSP, Min-Steiner-Tree, or Max-Acyclic-Subgraph. For each of these problems, the current best result is by a somewhat intricate gadget reduction from 3Lin(Z_2) [PV06, CC02, New00]. Such was the case previously for Max-Cut, 2Sat, 2And, and 2Lin(Z_2); perhaps our new approach could lead to an improved direct reduction from Label-Cover.

Acknowledgments. The first-named author would like to thank Nathanaël François and Anupam Gupta for collaboration on some early aspects of this research. The authors also thank Prasad Raghavendra and Siu On Chan for pointing out errors in an earlier version of the paper.

A An Invariance theorem

A.1 Notation and preliminaries

Let $f : Z_q^{d_1 K} \to \triangle_q$ and $g : Z_q^{d_2 K} \to \triangle_q$. Write $B_1(t) = \{(t-1)d_1 + 1, \ldots, td_1\}$ and $B_2(t) = \{(t-1)d_2 + 1, \ldots, td_2\}$ for $t \in [K]$ for the tth "blocks". Let π be a distribution as in Definition 4.7

with marginals π_1 and π_2. Recall that π_j (for $j = 1, 2$) must equal $\mu_{q,\rho_j}^{d_j}$ for some $\rho_j \in [0, 1]$. Let μ be the distribution on $(Z_q \cup \{*\})^{d_1} \times (Z_q \cup \{*\})^{d_2}$ with independent marginals equal to π_1 and π_2. As before, we will consider the product distributions $\mu^{\otimes K}$ and $\pi^{\otimes K}$ as distributions on pairs of strings from $(Z_q \cup \{*\})^{d_1 K} \times (Z_q \cup \{*\})^{d_2 K}$. Note that for $\mu^{\otimes K}$, the two strings are independent.

Denote the random variable distributed as $\pi^{\otimes K}$ by $\boldsymbol{z}$. There are two components to $\boldsymbol{z}$: the one used as an input to f and the one used as an input to g. (Ignore the fact that $\boldsymbol{z}$ will most likely have $*$'s in it even though f and g don't accept $*$'s as inputs. We will address this shortly.) Denote the first as $\boldsymbol{x}$ and the second as $\boldsymbol{y}$, so that $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$. Similarly, denote the random variable distributed as $\mu^{\otimes K}$ by $\boldsymbol{z}' = (\boldsymbol{x}', \boldsymbol{y}')$. These can be written in vector notation:

$$\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_K), \qquad \boldsymbol{y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_K), \qquad \text{etc.}$$

It should be clear that for all t, $\boldsymbol{z}_t = (\boldsymbol{x}_t, \boldsymbol{y}_t)$ is distributed as π, and $\boldsymbol{z}'_t = (\boldsymbol{x}'_t, \boldsymbol{y}'_t)$ is distributed as μ. Our Lindeberg-style proof of the Invariance principle uses distributions which are hybrids of $\boldsymbol{z}$ and $\boldsymbol{z}'$. For each $t = 0, 1, \dots, K$, the tth hybrid distribution between $\boldsymbol{z}$ and $\boldsymbol{z}'$ is $\boldsymbol{z}^{(t)} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_t, \boldsymbol{z}'_{t+1}, \dots, \boldsymbol{z}'_K)$. So, $\boldsymbol{z}^{(K)} = \boldsymbol{z}$, and $\boldsymbol{z}^{(0)} = \boldsymbol{z}'$. This extends naturally to $\boldsymbol{x}^{(t)}$ and $\boldsymbol{y}^{(t)}$, where $\boldsymbol{x}^{(t)} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_t, \boldsymbol{x}'_{t+1}, \dots, \boldsymbol{x}'_K)$ and $\boldsymbol{y}^{(t)} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_t, \boldsymbol{y}'_{t+1}, \dots, \boldsymbol{y}'_K)$. Note finally that we are choosing to write strings generated by $\pi^{\otimes K}$ as $(\boldsymbol{x}, \boldsymbol{y})$, even though we have previously written them as $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$; this is chosen for this section only to reduce notational clutter.

Identify the erasure symbol $*$ with the point $(\frac{1}{q}, \dots, \frac{1}{q}) \in \triangle_q$. (This is sensible since erased symbols will be equally likely to be any $l \in Z_q$.) We may view points $\widetilde{w} \in \Delta_q$ as a probability distribution over elements w of Z_q in the natural way. Then we may extend f and g to have domain $\triangle_q^{d_1 K}$ and $\triangle_q^{d_2 K}$, respectively, by

$$f(\widetilde{w}) := \mathbf{E}[f(\boldsymbol{w}) \mid \widetilde{w}],$$

and similarly for g, recalling that f and g are simplex-valued. We refer to functions extended this way as *multilinearized*.

From the perspective of h, given the input $x \in (Z_q \cup \{*\})^{d_1 K}$, the distribution of the output of f comes from taking the expectation of f's output over a random "filling in" of x, with η noise added. Write

$$F(x) := (\mathrm{T}_{1-\eta} f)(x) = f((1-\eta)x + \eta *) \in \Delta_q,$$

and

$$G(y) := (\mathrm{T}_{1-\eta} g)(y) = g((1-\eta)y + \eta *) \in \Delta_q.$$

Then if h is given inputs x and y, the distributions of the outputs of f and g are $F(x)$ and $G(y)$, respectively.

For a multilinearized function $f : \triangle_q^{d_1 K} \to \triangle_q$, we define the operators L_t and E_t by

$$\mathrm{L}_t f = \sum_{S : S \cap B_1(t) \neq \emptyset} f^S, \quad \text{and} \quad \mathrm{E}_t f = \sum_{S : S \cap B_1(t) = \emptyset} f^S.$$

Note that $\mathrm{E}_t f$ does not depend on the inputs within the tth block, whereas $\mathrm{L}_t f$ does. Intuitively, $\mathrm{E}_t f$ represents f "averaged over" the coordinates in its tth block, and $\mathrm{L}_t f$ is what remains. We formalize this in the following proposition, which follows easily from Fact 4.15:

Proposition A.1. *Given a multilinearized function $f : \triangle_q^{d_1 K} \to \triangle_q$ and an input $x \in Z_q^{d_1 K}$, let x_* denote x with its t-th block replaced with $*^{d_1}$, for some $t \in [K]$. Then $\mathrm{E}_t f(x) = f(x_*)$.*

We have the following relation between f's influences and $\mathrm{L}_t f$:

$$\mathbf{Inf}_{B_1(t)}^{(1-\eta)}[f] = \sum_{S : S \cap B_1(t) \neq \emptyset} (1-\eta)^{|S|} \|f^S\|_2^2 = \|\mathrm{T}_{\sqrt{1-\eta}} \mathrm{L}_t f\|_2^2.$$

A.2 Statement of the theorem

Let $\Psi : \mathbb{R}^{2q} \to \mathbb{R}$ be a smooth function with bounded derivatives. Our Invariance principle shows that if f and g do not share influential coordinates, then the test function Ψ cannot distinguish between the case when their inputs are correlated, as in $\boldsymbol{x}$ and $\boldsymbol{y}$, or uncorrelated, as in $\boldsymbol{x}'$ and $\boldsymbol{y}'$. More formally, we prove the following theorem:

Theorem A.2. *Let Ψ be a C^∞ function satisfying*

$$|\partial^{(\beta)}\Psi| \le C_1 \quad \forall|\beta| = 1, \qquad |\partial^{(\beta)}\Psi| \le C_3 \quad \forall|\beta| = 3,$$

where C_1 and C_3 are constants. Let $q \ge 2$ and $\eta > 0$ be given, and let $f : \triangle_q^{d_1 K} \to \triangle_q$ and $g : \triangle_q^{d_2 K} \to \triangle_q$ be multilinearized functions satisfying

$$\min(\mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f], \mathbf{Inf}^{(1-\eta)}_{B_2(t)}[g]) \le \kappa \quad \forall t \in [K],$$

for κ a constant. Then

$$|\,\mathbf{E}[\Psi(F(\boldsymbol{x}), G(\boldsymbol{y}))] - \mathbf{E}[\Psi(F(\boldsymbol{x}'), G(\boldsymbol{y}'))]| \le 2C_1 c_\eta \sqrt{q}\kappa^{1/4} + 2C_3 c_\eta q^3 \kappa^{c(q,\eta)/4},$$

where $c_\eta = \frac{2}{\eta}\ln\left(\frac{1}{\eta}\right)$ and $c(q,\eta) = \Theta(\eta/\log q)$.

A.3 Invariance

We now begin the proof of Theorem A.2. What follows is highly similar to Mossel's Invariance principle [Mos10]. As usual in Invariance proofs, we replace the K coordinates of $\boldsymbol{z}$ with those of $\boldsymbol{z}'$, one at a time. We mainly investigate the magnitude of the error incurred in the tth step; we add up these K errors at the end. Write F and G as two coordinates of a single multi-dimensional function $H = (F, G)$. (Previously, lower-case h has designated the "middleman" function. Upper-case H is of no relation, and is used out of convenience.) Notationally, F receives inputs named x, G receives inputs named y, and H receives inputs named $z = (x, y)$. So in the tth step, we incur an error of

$$\mathrm{err}_t := |\,\mathbf{E}[\Psi(H(\boldsymbol{z}^{(t-1)}))] - \mathbf{E}[\Psi(H(\boldsymbol{z}^{(t)}))]|.$$

Let us write

$$\begin{aligned}
\boldsymbol{F} &:= \mathrm{E}_t F(\boldsymbol{x}^{(t-1)}) & \boldsymbol{G} &:= \mathrm{E}_t G(\boldsymbol{y}^{(t-1)}) & \boldsymbol{H} &:= \mathrm{E}_t H(\boldsymbol{z}^{(t-1)}) \\
\boldsymbol{F}_{t-1} &:= \mathrm{L}_t F(\boldsymbol{x}^{(t-1)}) & \boldsymbol{G}_{t-1} &:= \mathrm{L}_t G(\boldsymbol{y}^{(t-1)}) & \boldsymbol{H}_{t-1} &:= \mathrm{L}_t H(\boldsymbol{z}^{(t-1)}) \\
\boldsymbol{F}_t &:= \mathrm{L}_t F(\boldsymbol{x}^{(t)}) & \boldsymbol{G}_t &:= \mathrm{L}_t G(\boldsymbol{y}^{(t)}) & \boldsymbol{H}_t &:= \mathrm{L}_t H(\boldsymbol{z}^{(t)})
\end{aligned}$$

Clearly, $\boldsymbol{H} = (\boldsymbol{F}, \boldsymbol{G})$, $\boldsymbol{H}_{t-1} = (\boldsymbol{F}_{t-1}, \boldsymbol{G}_{t-1})$, and $\boldsymbol{H}_t = (\boldsymbol{F}_t, \boldsymbol{G}_t)$. Thus,

$$\mathrm{err}_t = |\,\mathbf{E}[\Psi(\boldsymbol{H} + \boldsymbol{H}_{t-1})] - \mathbf{E}[\Psi(\boldsymbol{H} + \boldsymbol{H}_t)]|.$$

Now we apply Taylor's theorem to Ψ, centered at $\boldsymbol{H}$, out to the third partial derivatives:

$$\left|\Psi(x+y) - \sum_{|\beta|<3} \frac{\Psi^{(\beta)}(x)}{\beta!}(y)^\beta\right| \le \sum_{|\beta|=3} \frac{\Psi^{(\beta)}(y)}{\beta!}|y|^\beta, \text{ for all } x, y \in \mathbb{R}^q.$$

This gives us the following upper bound on err_t:

$$\mathrm{err}_t \le \left|\sum_{\beta<3}\left(\mathbf{E}\left[\frac{\Psi^{(\beta)}(\boldsymbol{H})}{\beta!}(\boldsymbol{H}_{t-1})^\beta\right] - \mathbf{E}\left[\frac{\Psi^{(\beta)}(\boldsymbol{H})}{\beta!}(\boldsymbol{H}_t)^\beta\right]\right)\right| + \sum_{|\beta|=3}\frac{C_3}{\beta!}\left(|\,\mathbf{E}[\boldsymbol{H}_{t-1}]|^\beta + |\,\mathbf{E}[\boldsymbol{H}_t]|^\beta\right).$$

Here we have used the upper bound on the third partial derivatives of Ψ. We will upper bound these two terms in the next two sections.

A.4 The quadratic part

We will now show that the degree-0, degree-1, and degree-2 terms cancel, in expectation. In other words, for all multi-indices $\beta \in \mathbb{N}^{2q}$ such that $|\beta| \le 2$,

$$\mathbf{E}\left[\frac{\Psi^{(\beta)}(\boldsymbol{H})}{\beta!}(\boldsymbol{H}_{t-1})^{\beta}\right] = \mathbf{E}\left[\frac{\Psi^{(\beta)}(\boldsymbol{H})}{\beta!}(\boldsymbol{H}_{t})^{\beta}\right].$$

First, note that $\Psi^{\beta}(\boldsymbol{H})/\beta!$ is independent of the value of $\boldsymbol{z}_t$ or $\boldsymbol{z}'_t$. Thus, it is sufficient to show that when everything is fixed except for $\boldsymbol{z}_t$ and $\boldsymbol{z}'_t$, then $(\boldsymbol{H}_{t-1})^{\beta}$ and $(\boldsymbol{H}_t)^{\beta}$ are distributed identically. Note also that fixing everything but $\boldsymbol{z}_t$ and $\boldsymbol{z}'_t$ turns $\boldsymbol{H}_{t-1}$ and $\boldsymbol{H}_t$ into functions which depend only on $\boldsymbol{z}'_t$ and $\boldsymbol{z}_t$, respectively.

When $|\beta| = 0$, this statement is trivial. For the other cases, split β into its two halves, $\beta_1, \beta_2 \in \mathbb{N}^q$, so that $\beta = (\beta_1, \beta_2)$. The first of these, β_1, covers the output indices of F, and the second of these, β_2, covers the output indices of G. When either $|\beta_1| = 0$ or $|\beta_2| = 0$, then $(\boldsymbol{H}_{t-1})^{\beta}$ depends either entirely on $\boldsymbol{F}_{t-1}$ or $\boldsymbol{G}_{t-1}$, and similarly $(\boldsymbol{H}_t)^{\beta}$ depends either entirely on $\boldsymbol{F}_t$ or $\boldsymbol{G}_t$. However, the marginial distributions $\boldsymbol{x}^{(t-1)}$ and $\boldsymbol{x}^{(t)}$ are identical, as are $\boldsymbol{y}^{(t-1)}$ and $\boldsymbol{y}^{(t)}$, so the expectations are equal. This covers the case of $|\beta| = 1$ and a portion of the case of $|\beta| = 2$.

What remains is when $|\beta_1| = |\beta_2| = 1$, in which we are trying to verify that

$$\mathbf{E}[P_{\Delta}(\boldsymbol{x}_t)Q_{\Delta}(\boldsymbol{y}_t)] = \mathbf{E}[P_{\Delta}(\boldsymbol{x}'_t)Q_{\Delta}(\boldsymbol{y}'_t)]$$

for some real-valued multilinearized functions $P_{\Delta} : \Delta_q^{d_1 K} \to \mathbb{R}$ and $Q_{\Delta} : \Delta_q^{d_2 K} \to \mathbb{R}$. Recall that these functions treat their inputs in Δ_q as probability distributions over Z_q, and output the expected value of some functions $P : Z_q^{d_1 K} \to \mathbb{R}$ and $Q : Z_q^{d_2 K} \to \mathbb{R}$ over these distributions. In other words,

$$P_{\Delta}(x) = \mathbf{E}[P(\boldsymbol{a}) \mid x], \qquad Q_{\Delta}(y) = \mathbf{E}[Q(\boldsymbol{b}) \mid y]$$

where and $\boldsymbol{a}_i$ is drawn from x_i (independently across i), and $\boldsymbol{b}_i$ is drawn from y_i. Each input variable to P_{Δ} (and Q_{Δ}) is either a randomly drawn element of Z_q or a $*$, which is then substituted by a randomly drawn element of Z_q. The result is that, from the perspectives of P and Q, the $*$'s don't matter; they are simply being evaluated on uniformly random inputs in $Z_q^{d_1 K}$ and $Z_q^{d_2 K}$. More formally,

$$\begin{aligned}\mathbf{E}[P_{\Delta}(\boldsymbol{x}_t)Q_{\Delta}(\boldsymbol{y}_t)] &= \mathbf{E}[\mathbf{E}[P(\boldsymbol{a}) \mid \boldsymbol{x}_t] \cdot \mathbf{E}[Q(\boldsymbol{b}) \mid \boldsymbol{y}_t]] \\ &= \mathbf{E}[\mathbf{E}[\mathbf{E}[P(\boldsymbol{a}) \cdot Q(\boldsymbol{b}) \mid \boldsymbol{x}_t] \mid \boldsymbol{y}_t]] = \mathbf{E}[P(\boldsymbol{w}) \cdot Q(\boldsymbol{v})],\end{aligned}$$

where $\boldsymbol{w}$ and $\boldsymbol{v}$ are distributed as independently uniform elements of Z_q^m. (This is where we use that π is a pseudo-independent distribution.) A similar deduction shows that $\mathbf{E}[P_{\Delta}(\boldsymbol{x}'_t)Q_{\Delta}(\boldsymbol{y}'_t)] = \mathbf{E}[P(\boldsymbol{w}) \cdot Q(\boldsymbol{v})]$ as well, so the two distributions are equivalent.

A.5 The cubic error term

The previous section shows that

$$\mathrm{err}_t \le \sum_{|\beta|=3} \frac{C_3}{\beta!}\left(|\,\mathbf{E}[\boldsymbol{H}_{t-1}]|^{\beta} + |\,\mathbf{E}[\boldsymbol{H}_t]|^{\beta}\right).$$

What remains is the cubic error terms. There are q^3 such terms, one for each of $\boldsymbol{H}_{t-1}$ and $\boldsymbol{H}_t$. Using $|\,\mathbf{E}[\boldsymbol{H}_{t-1}]| \le \mathbf{E}[|\boldsymbol{H}_{t-1}|]$, we get that

$$\mathrm{err}_t \le 2C_3 q^3 B, \qquad \text{assuming } \mathbf{E}[|\boldsymbol{H}_{t-1}|^{\beta}], \mathbf{E}[|\boldsymbol{H}_t|^{\beta}] \le B \quad \forall |\beta| = 3. \tag{17}$$

We'll focus on β's for which $|\beta_1| = 1$ and $|\beta_2| = 2$ (keeping in mind that all other cases can be handled similarly). When β is of this form, it selects one component, l_1, from the F side and two components, l_2 and l_3, from the G side (l_2 may equal l_3). Writing $(\boldsymbol{H}_t)_l$ for the lth component of $\boldsymbol{H}_t$, we have

$$\begin{aligned}\mathbf{E}[|\boldsymbol{H}_t|^\beta] &= \mathbf{E}[|(\boldsymbol{F}_t)_{l_1}(\boldsymbol{G}_t)_{l_2}(\boldsymbol{G}_t)_{l_3}|] \\ &\le \mathbf{E}[|(\boldsymbol{F}_t)_{l_1}|^3]^{1/3}\,\mathbf{E}[|(\boldsymbol{G}_t)_{l_2}|^3]^{1/3}\,\mathbf{E}[|(\boldsymbol{G}_t)_{l_3}|^3]^{1/3}, \end{aligned} \tag{18}$$

using Hölder's inequality. Let M_F equal $\max_l\{\mathbf{E}[|(\boldsymbol{F}_t)_l|^3]\}$ and M_G equal $\max_l\{\mathbf{E}[|(\boldsymbol{G}_t)_l|^3]\}$. Equation (18) is clearly less than $\max\{M_F, M_G\} \le M_F + M_G$. The result of this application of Hölder's inequality is to break up the dependence between the F and the G sides. Indeed, doing the same thing for $\mathbf{E}[|\boldsymbol{H}_{t-1}|^\beta]$ shows that it too is $\le M_F + M_G$, since independently of anything else, $\boldsymbol{x}^{(t-1)}$ and $\boldsymbol{x}^{(t)}$ are distributed identically, as are $\boldsymbol{y}^{(t-1)}$ and $\boldsymbol{y}^{(t)}$.

For an arbitrary l, let's now look at

$$\mathbf{E}[|(\boldsymbol{F}_t)_l|^3] = \mathbf{E}[|(\mathrm{L}_t F(\boldsymbol{x}^{(t)})_l)|^3] = \mathbf{E}[|(\mathrm{L}_t F(\boldsymbol{x}))_l|^3].$$

We claim that

$$\mathbf{E}[|(\mathrm{L}_t F(\boldsymbol{x}))_l|^3] \le \mathbf{E}[|\mathrm{L}_t F(\boldsymbol{w}_1, \dots, \boldsymbol{w}_K)_l|^3],$$

where $\boldsymbol{w}_i$ is independently distributed as a uniform element of $\mathbb{Z}_q^m$ for each i. In fact, this holds for a generic multilinearized function ϕ. To prove this, introduce the random variable $\boldsymbol{v} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_K)$ which is drawn from $\boldsymbol{x}$. Then,

$$\mathbf{E}[|\phi(\boldsymbol{x})|^3] = \mathbf{E}[|\,\mathbf{E}[\phi(\boldsymbol{v})|\boldsymbol{x}]|^3] \le \mathbf{E}[\mathbf{E}[|\phi(\boldsymbol{v})|^3|\boldsymbol{x}]] = \mathbf{E}[|\phi(\boldsymbol{w})|^3].$$

As a result, we now consider $\mathbf{E}[|(\mathrm{L}_t F(\boldsymbol{w}))_l|^3]$. Recall that $F = \mathrm{T}_{1-\eta} f$, so

$$\mathbf{E}[|(\mathrm{L}_t F(\boldsymbol{w}))_l|^3] = \mathbf{E}[|(\mathrm{L}_t \mathrm{T}_{1-\eta} f(\boldsymbol{w}))_l|^3] = \mathbf{E}[|(\mathrm{T}_{1-\eta}\mathrm{L}_t f(\boldsymbol{w}))_l|^3],$$

because $\mathrm{T}_{1-\eta}$ and L_t commute. The function $(\mathrm{T}_{1-\eta}\mathrm{L}_t f)_l$ takes values in $[0,1]$, so we have

$$\mathbf{E}[|\mathrm{T}_{1-\eta}\mathrm{L}_t f(\boldsymbol{w})_l|^3] = \mathbf{E}[|\mathrm{T}_{\sqrt{1-\eta}}\mathrm{T}_{\sqrt{1-\eta}}\mathrm{L}_t f(\boldsymbol{w})_l|^3] \le \mathbf{E}[|\mathrm{T}_{\sqrt{1-\eta}}\mathrm{T}_{\sqrt{1-\eta}}\mathrm{L}_t f(\boldsymbol{w})_l|^r]$$

for $r = 2+2c$, where $c = c(q,\eta) > 0$ is a small number chosen so that $\mathrm{T}_{\sqrt{1-\eta}}$ is $(2,r)$-hypercontractive on $\mathbb{Z}_q^m$. (By [LO00, Ole03, Wol07] one may take $c = \Theta(\eta/\log q)$.) Thus we may upper-bound the above by

$$\mathbf{E}[|\mathrm{T}_{\sqrt{1-\eta}}\mathrm{L}_t f(\boldsymbol{w})_l|^2]^{r/2} = \mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f_l]^{1+c} \le \mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f]^{1+c}.$$

So plugging all of these deductions (and the identical ones for g) back into (18), we get

$$\mathbf{E}[|\boldsymbol{H}_t|^\beta] \le \mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f]^{1+c} + \mathbf{Inf}^{(1-\eta)}_{B_1(t)}[g]^{1+c}.$$

For brevity, let's write $\mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f]$ as $I_t[f]$ and $\mathbf{Inf}^{(1-\eta)}_{B_1(t)}[g]$ as $I_t[g]$. Then this upper bound is

$$I_t[f]^{(1+c)} + I_t[g]^{(1+c)} \le (I_t[f] + I_t[g]) \cdot \max(I_t[f], I_t[g])^c. \tag{19}$$

A.6 Influences

Now, recall our assumption on influences, which is that

$$\min(I_t[f], I_t[g]) \le \epsilon.$$

Define $t \in [K]$ to be *half-influential* if $\max(I_t[f], I_t[g]) > \epsilon_0$, where $\epsilon_0 > \epsilon$ is a parameter to be set later. Assume for the moment that t is not half-influential. Then we may upper bound (19) by $(I_1 + I_2) \cdot \epsilon_0^c$, and hence (recall (17))

$$\mathrm{err}_t \le O(C_3 q^3) \cdot (I_t[f] + I_t[g]) \cdot \epsilon_0^c, .$$

We need the following fact:

Fact A.3. *Take* $c_\eta = \frac{2}{\eta} \ln\left(\frac{1}{\eta}\right)$. *Then* $\sum_{t=1}^K \mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f] \le c_\eta$.

Proof. For all $x > 0$ we can bound $x(1-\eta)^x$ by c_η. Thus,

$$\begin{aligned}\sum_{t=1}^K \mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f] &= \sum_{S \subseteq [d_1 K]} |\{t \mid S \cap B_1(t) \neq \emptyset\}| (1-\eta)^{|S|} \|f^S\|_2^2 \\ &\le \sum_{S \subseteq [d_1 K]} |S| (1-\eta)^{|S|} \|f^S\|_2^2 \le \sum_{S \subseteq [d_1 K]} c_\eta \|f^S\|_2^2 = c_\eta \|f\|_2^2 = c_\eta. \qquad \square\end{aligned}$$

An identical deduction holds for g. Using this, we may now sum this bound over all non-half-influential t and get that the bulk of the overall error (the non-half-influential part) is at most

$$2C_3 q^3 \cdot \epsilon_0^{c(q,\eta)} \cdot \sum_{t=1}^K (\mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f] + \mathbf{Inf}^{(1-\eta)}_{B_2(t)}[g]) \le 2C_3 c_\eta q^3 \cdot \epsilon_0^{c(q,\eta)}, \tag{20}$$

It remains to deal with the error from the half-influential blocks t.

A.7 Half-influences

Now we handle the half-influential t. The good thing about the half-influential t is that there are not too many of them. Specifically, if t is half-influential we have that a $(1-\eta)$-noisy influence exceeds ϵ_0; again, by Fact A.3, we deduce that there are at most c_η/ϵ_0 half-influential coordinates total.

Now assume t is a half-influential coordinate. We still know that one of the two noisy-influences is at most ϵ; say without loss of generality that $\mathbf{Inf}^{(1-\eta)}_{B_1(t)}[f] \le \epsilon$. We again want to bound err_t. To this end, define the random variables $\boldsymbol{z}_*^{(t-1)} = (\boldsymbol{x}_*^{(t-1)}, \boldsymbol{y}^{(t-1)})$ and $\boldsymbol{z}_*^{(t)} = (\boldsymbol{x}_*^{(t)}, \boldsymbol{y}^{(t)})$ where $\boldsymbol{z}_*^{(i)}$ has the same distribution as $\boldsymbol{z}^{(i)}$, only the t-th block of $\boldsymbol{x}_*^{(i)}$ is always set to $*^{d_1}$. Note in fact that $\boldsymbol{z}_*^{(t-1)}$ and $\boldsymbol{z}_*^{(t)}$ are distributed identically to each other, as the only difference between $\boldsymbol{z}^{(t-1)}$ and $\boldsymbol{z}^{(t)}$ is in the correlation of their t-th blocks, and we have removed that correlation by fixing the t-th blocks of $\boldsymbol{x}_*^{(t-1)}$ and $\boldsymbol{x}_*^{(t)}$. Thus,

$$\begin{aligned}\mathrm{err}_t &= \left|\mathbf{E}[\Psi(H(\boldsymbol{z}^{(t-1)}))] - \mathbf{E}[\Psi(H(\boldsymbol{z}^{(t)}))]\right| \\ &= \left|\mathbf{E}[\Psi(H(\boldsymbol{z}^{(t-1)}))] - \mathbf{E}[\Psi(H(\boldsymbol{z}_*^{(t-1)}))] + \mathbf{E}[\Psi(H(\boldsymbol{z}_*^{(t)}))] - \mathbf{E}[\Psi(H(\boldsymbol{z}^{(t)}))]\right| \\ &\le \left|\mathbf{E}[\Psi(H(\boldsymbol{z}^{(t-1)}))] - \mathbf{E}[\Psi(H(\boldsymbol{z}_*^{(t-1)}))]\right| + \left|\mathbf{E}[\Psi(H(\boldsymbol{z}_*^{(t)}))] - \mathbf{E}[\Psi(H(\boldsymbol{z}^{(t)}))]\right| \\ &\le \mathbf{E}\left[\left|\Psi(H(\boldsymbol{z}^{(t-1)})) - \Psi(H(\boldsymbol{z}_*^{(t-1)}))\right|\right] + \mathbf{E}\left[\left|\Psi(H(\boldsymbol{z}_*^{(t)})) - \Psi(H(\boldsymbol{z}^{(t)}))\right|\right].\end{aligned}$$

Let us focus here on the second error term. An analogous argument will give the same bound for the first term. Using the first-derivative bounds on Ψ, (i.e., its C_1-Lipschitzness in each coordinate) and also the fact that the second coordinates of $\boldsymbol{z}^{(t)}$ and $\boldsymbol{z}_*^{(t)}$ (i.e., $\boldsymbol{y}^{(t)}$) are identical, the first error is at most

$$\begin{aligned} & C_1 \cdot \mathbf{E}\left[\left\|F(\boldsymbol{x}^{(t)}) - F(\boldsymbol{x}_*^{(t)})\right\|_1\right] \\ \leq\ & C_1\sqrt{q} \cdot \mathbf{E}\left[\left\|F(\boldsymbol{x}^{(t)}) - F(\boldsymbol{x}_*^{(t)})\right\|_2\right] \\ \leq\ & C_1\sqrt{q} \cdot \mathbf{E}\left[\left\|F(\boldsymbol{x}^{(t)}) - F(\boldsymbol{x}_*^{(t)})\right\|_2^2\right]^{1/2}, \end{aligned}$$

where the two inequalities are by Cauchy-Schwarz. Consider the function $\phi = F - F_*$, where F_* is F with $*^{d_1}$ hard-wired into the input coordinates of its t-th block. Then ϕ is a multilinearized function, and a similar argument as before shows that $\mathbf{E}[\|\phi(\boldsymbol{x}^{(t)})\|_2^2] \leq \mathbf{E}[\|\phi(\boldsymbol{w})\|_2^2]$, where $\boldsymbol{w}$ is a uniformly random element of $\mathbb{Z}_q^{d_1 K}$. Let $\boldsymbol{w}_*$ be $\boldsymbol{w}$ with its t-th block fixed to $*^{d_1}$. This shows that

$$\mathbf{E}\left[\left\|F(\boldsymbol{x}^{(t)}) - F(\boldsymbol{x}_*^{(t)})\right\|_2^2\right]^{1/2} \leq \mathbf{E}\left[\|F(\boldsymbol{w}) - F(\boldsymbol{w}_*)\|_2^2\right]^{1/2} = \mathbf{E}\left[\|\mathrm{L}_t F(\boldsymbol{w})\|_2^2\right]^{1/2},$$

where the last step is by Proposition A.1. But here we are done, as

$$\begin{aligned} \mathbf{E}\left[\|\mathrm{L}_t F(\boldsymbol{w})\|_2^2\right]^{1/2} &= \mathbf{E}\left[\|\mathrm{L}_t \mathrm{T}_{1-\eta} f(\boldsymbol{w})\|_2^2\right]^{1/2} \\ &= \mathbf{E}\left[\|\mathrm{T}_{1-\eta} \mathrm{L}_t f(\boldsymbol{w})\|_2^2\right]^{1/2} \\ &\leq \mathbf{E}\left[\left\|\mathrm{T}_{\sqrt{1-\eta}} \mathrm{L}_t f(\boldsymbol{w})\right\|_2^2\right]^{1/2} \\ &= \mathbf{Inf}_{B_1(t)}^{(1-\eta)}[f]^{1/2}, \end{aligned}$$

which by assumption is at most $\epsilon^{1/2}$. Now, by the sum-of-influences lemma, there are at most c_η/ϵ_0 such t. Hence the overall contribution to the error from the half-influential t is at most

$$2 \cdot C_1\sqrt{q}\epsilon^{1/2} \cdot c_\eta/\epsilon_0 \leq 2C_1 c_\eta \sqrt{q}\epsilon^{1/4}, \qquad \text{taking } \epsilon_0 = \epsilon^{1/4}.$$

A.8 Conclusion

Combining the non-half-influential and the half-influential errors, we get a final overall error bound for $|\mathbf{E}[M(H(\boldsymbol{z}))] - \mathbf{E}[M(H(\boldsymbol{z}'))]|$ of

$$2C_1 c_\eta \sqrt{q}\epsilon^{1/4} + 2C_3 c_\eta q^3 \epsilon^{c(q,\eta)/4}.$$

Again, $c_\eta = \frac{2}{\eta}\log\left(\frac{1}{\eta}\right)$ and $c(q,\eta) = \Theta(\eta/\log(q))$. This completes the theorem.

B The probability the tests pass

Let T be a 2-party, q-ary, (c,d)-blocked, η-noise correlated test using the correlated distribution π with marginals μ and ν. We now apply Theorem A.2 to show that when performing the soundness analysis of T, we may decouple the correlation between π's marginals without paying too high a price. We restate Theorem 4.18:

Theorem 4.18 restated. *Let T be a 2-party, q-ary, (d_1, d_2)-blocked, η-noise correlated test. Let T' denote its uncorrelated version. Assume we are applying these test under the blocking maps π_u and π_v. Let $f : Z_q^{d_1 K} \to \triangle_q$ and $g : Z_q^{d_2 K} \to \triangle_q$ satisfy*

$$\min(\mathbf{Inf}^{(1-\eta)}_{\pi_u^{-1}(t)}[f], \mathbf{Inf}^{(1-\eta)}_{\pi_v^{-1}(t)}[g]) \le \kappa \quad \forall t \in [K].$$

Then
$$\mathrm{Val}_T(f,g) \le \mathrm{Opt}(T') + \epsilon(\eta,\kappa),$$
and $\mathrm{Val}_T(f,g) \le \mathrm{Opt}_{\mathrm{folded}}(T') + \epsilon(\eta,\kappa)$ *if f, g folded.*

Here $\epsilon(\eta,\kappa) \le (q/\eta)^{O(1)} \cdot \kappa^{\Omega(\eta/\log q)}$; in particular, for each fixed $\eta > 0$, $\epsilon(\eta,\kappa) \to 0$ as $\kappa \to 0$.

Proof. We will show that $|\mathrm{Val}_T(f,g) - \mathrm{Val}_{T'}(f,g)| \le \epsilon(\eta,\kappa)$. This shows the theorem by $\mathrm{Val}_T(f,g) \le \mathrm{Val}_{T'}(f,g) + \epsilon(\eta,\kappa) \le \mathrm{Opt}(T') + \epsilon(\eta,\kappa)$. When f and g are folded, $\mathrm{Val}_{T'}(f,g) \le \mathrm{Opt}_{\mathrm{folded}}(T')$, so we may instead write $\mathrm{Val}_T(f,g) \le \mathrm{Opt}_{\mathrm{folded}}(T') + \epsilon(\eta,\kappa)$.

Extend f and g to functions $f : \triangle_q^{d_1 K} \to \triangle_q$ and $g : \triangle_q^{d_2 K} \to \triangle_q$. The outputs of f and g, conditioned on $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \sim \pi$, are distributed as

$$F(\tilde{\boldsymbol{x}}) := (T_{1-\eta} f)(\tilde{\boldsymbol{x}}) \qquad \text{and} \qquad G(\tilde{\boldsymbol{y}}) := (T_{1-\eta} g)(\tilde{\boldsymbol{y}}),$$

respectively.

Let (ϕ, j) be a test drawn from $\mathcal{T}$. Without loss of generality, assume that $j = 1$. Write $\phi_b(a) := \phi(a,b)$. Say that $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ is drawn from π, and that $h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) = b$. The only values that f can output which will satisfy ϕ are those in $\phi_b^{-1}(1)$. Indeed, the probability that ϕ is satisfied is $\sum_{a \in \phi_b^{-1}(1)} F_a(\widetilde{\boldsymbol{x}})$, the probability that the output of $f(\dot{\boldsymbol{x}})$ falls in $\phi_b^{-1}(1)$. A similar argument holds for tests involving g. Write

$$\psi_b(s_0, \dots, s_{q-1}, t_0, \dots, t_{q-1}) := \sum_{(\phi,1) \in \mathcal{T}} \mathcal{T}(\phi, 1) \sum_{a \in \phi_b^{-1}(1)} s_a + \sum_{(\phi,2) \in \mathcal{T}} \mathcal{T}(\phi, 2) \sum_{a \in \phi_b^{-1}(1)} t_a.$$

Then the probability the test is passed, conditioned on $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \sim \pi$ and $h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) = b$, is precisely $\psi_b(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))$. Note that ψ_b is a linear function of its inputs, all of which have coefficients between 0 and 1, inclusive.

An optimal h for f and g will simply output the element b which maximizes this probability. If we write

$$M(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}})) := \mathrm{Max}_q(\psi_0(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}})), \dots, \psi_{q-1}(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))),$$

then an optimal h passes the test with probability $\mathbf{E}_{(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \sim \pi}[M(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] = \mathrm{Val}_T(f,g)$. Note that $\mathbf{E}_{(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \sim \mu \times \nu}[M(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] = \mathrm{Val}_{T'}(f,g)$. Thus, we are seeking to upper bound the expression

$$|\mathrm{Val}_T(f,g) - \mathrm{Val}_{T'}(f,g)| = \left| \mathop{\mathbf{E}}_{(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \sim \pi} [M(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] - \mathop{\mathbf{E}}_{(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \sim \mu \times \nu} [M(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] \right|. \tag{21}$$

To apply Theorem A.2 to this expression, we need to replace M with a smooth approximation. Theorem C.1 provides a function $\Psi_{q,\delta}$ such that $\|\mathrm{Max}_q - \Psi_{q,\delta}\| \le 4\delta \log q$, and

$$|\partial^{(\beta)} \Psi_{q,\delta}| \le O\left(q^{\alpha_1}\right) \quad \forall |\beta| = 1, \qquad |\partial^{(\beta)} \Psi_{q,\delta}| \le O\left(\frac{1}{\delta^2} q^{\alpha_3}\right) \quad \forall |\beta| = 3,$$

where α_1 and α_3 are absolute constants. Now, define

$$\Psi(F(\widetilde{x}), G(\widetilde{y})) := \Psi_{q,\eta}(\psi_0(F(\widetilde{x}), G(\widetilde{y})), \dots, \psi_{q-1}(F(\widetilde{x}), G(\widetilde{y}))).$$

It is clear that $\|M - \Psi\|_\infty \le 4\delta \log q$. Furthermore, because for all b, ψ_b is a linear function whose coefficients lie between 0 and 1,

$$|\partial^{(\beta)}\Psi| \le O\left(q^{\alpha_1+1}\right) \quad \forall|\beta| = 1, \qquad |\partial^{(\beta)}\Psi| \le O\left(\frac{1}{\delta^2} q^{\alpha_3+3}\right) \quad \forall|\beta| = 3,$$

By the triangle inequality, Equation (21) is at most

$$\left| \mathop{\mathbf{E}}_{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\sim\pi} [M(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] - \mathop{\mathbf{E}}_{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\sim\pi} [\Psi(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] \right| \tag{22}$$

$$+ \left| \mathop{\mathbf{E}}_{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\sim\pi} [\Psi(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] - \mathop{\mathbf{E}}_{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\sim\mu\times\nu} [\Psi(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] \right| \tag{23}$$

$$+ \left| \mathop{\mathbf{E}}_{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\sim\mu\times\nu} [\Psi(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] - \mathop{\mathbf{E}}_{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\sim\mu\times\nu} [M(F(\widetilde{\boldsymbol{x}}), G(\widetilde{\boldsymbol{y}}))] \right|. \tag{24}$$

We may upper bound lines (22) and (24) by $4\delta \log q$ each. It remains to apply Theorem A.2 to line (23), which gives a total bound of

$$8\delta \log q + 2 \cdot O\left(q^{\alpha_1+1}\right) \frac{2}{\eta} \ln\left(\frac{1}{\eta}\right) \sqrt{q}\kappa^{1/4} + 2 \cdot O\left(\frac{1}{\delta^2} q^{\alpha_3+3}\right) \frac{2}{\eta} \ln\left(\frac{1}{\eta}\right) q^3 \kappa^{\Theta(\eta/\log q)/4}. \tag{25}$$

Note that the second and third terms are bounded from above by

$$\frac{1}{\delta^2} \left(\frac{q}{\eta}\right)^{O(1)} \kappa^{\Theta(\eta/\log q)},$$

so Equation 25 is at most

$$8\delta \log q + \frac{1}{\delta^2} \left(\frac{q}{\eta}\right)^{O(1)} \kappa^{\Theta(\eta/\log q)}.$$

By selecting δ to make these terms equal, we may set

$$\delta := \left(\frac{1}{\log q} \left(\frac{q}{\eta}\right)^{O(1)} \kappa^{\Theta(\eta/\log q)}\right)^{1/3} = \left(\frac{q}{\eta}\right)^{O(1)} \kappa^{\Theta(\eta/\log q)}.$$

Because δ was selected to make both terms equal, the total bound on $|\mathrm{Val}_T(f,g) - \mathrm{Val}_{T'}(f,g)|$ is $\epsilon(\eta,\kappa) := (q/\eta)^{O(1)} \kappa^{\Theta(\eta/\log q)}$. □

C Max approximator

Define $\mathrm{Max}_n : \mathbb{R}^n \to \mathbb{R}$ as

$$\mathrm{Max}_n(x_1, \ldots, x_n) := \max_i |x_i|.$$

We can approximate Max_n with a smooth function:

Theorem C.1. *For $n \ge 2$ and $0 < \delta < 1/2$, there exists a function $\Psi_{n,\delta}$ such that $\|\Psi_{n,\delta} - \mathrm{Max}_n\|_\infty \le 4\delta \log n$. In addition, $\Psi_{n,\delta}$ has bounded derivatives:*

$$|\partial^{(\beta)}\Psi_{n,\delta}| \le O\left(n^{\alpha_1}\right) \quad \forall|\beta| = 1, \quad \textit{and} \qquad |\partial^{(\beta)}\Psi_{n,\delta}| \le O\left(\frac{1}{\delta^2} n^{\alpha_3}\right) \quad \forall|\beta| = 3,$$

where α_1 and α_3 are absolute constants.

Proof. More explicitly, we will first prove the following bounds on the derivatives of $\Psi_{n,\delta}$ for values of n which are powers of two:

$$|\partial^{(\beta)}\Psi_{n,\delta}| \le n^{\log_2 C_1} \quad \forall|\beta| = 1, \qquad |\partial^{(\beta)}\Psi_{n,\delta}| \le \frac{C_2}{\delta} n^{2\log_2 C_1} \log_2 n \quad \forall|\beta| = 2,$$

$$|\partial^{(\beta)}\Psi_{n,\delta}| \le \frac{3C_2^2 C_3}{\delta^2} n^{3\log_2 C_1} \log_2^2 n \quad \forall|\beta| = 3,$$

where C_1, C_2, and C_3 are absolute constants. Note that these bounds are all polynomial in n. To do this, we will construct an approximating function for the case of $n = 2$, and will apply this function recursively for the case when n is a power of two, using the fact that $\mathrm{Max}_{2n}(x_1, x_2, \ldots, x_{2n-1}, x_{2n}) = \mathrm{Max}_n(\mathrm{Max}_2(x_1, x_2), \ldots, \mathrm{Max}_2(x_{2n-1}, x_{2n}))$. The general case for Max_n comes from taking the approximator for $\mathrm{Max}_{n_{pow}}$, where n_{pow} is the smallest power of two larger than n, and hardwiring $n_{pow} - n$ of the inputs to 0. As $n_{pow} \le 2n$, the bounds on the derivatives of $\Psi_{n,\delta}$ are still polynomial in n, implying the theorem in full generality.

The base case. We can write

$$\mathrm{Max}_2(x_1, x_2) = \frac{|x_1 + x_2| + |x_1 - x_2|}{2}.$$

Define $\mathrm{abs}(x) = |x|$. By Lemma 3.21 of [MOO10], for all $0 < \delta < 1/2$ there is a C^∞ function abs_δ which satisfies

$$\|\mathrm{abs}_\delta - \mathrm{abs}\|_\infty \le 2\delta; \text{ and, } \|(\mathrm{abs}_\delta)^{(r)}\|_\infty \le O(\delta^{1-r}).$$

Let C_1, C_2, and C_3 be constants greater than or equal to 1 such that $\|(\mathrm{abs}_\delta)^{(r)}\|_\infty \le C_r \delta^{(1-r)}$ for all $r \in \{1, 2, 3\}$. Set $\Psi_{2,\delta} := \mathrm{abs}_\delta(x_1 + x_2)/2 + \mathrm{abs}_\delta(x_1 - x_2)/2$. It is an easy exercise to check that $\|\mathrm{Max}_2 - \Psi_{2,\delta}\|_\infty \le 2\delta$, and $\|\partial\Psi_{2,\delta}^{(\beta)}\|_\infty \le C_{|\beta|}\delta^{1-|\beta|}$, for all β with $|\beta| \in \{1, 2, 3\}$.

The inductive step. Let $n = 2^m$ for some integer $m \ge 2$. By induction, there exists a function $\Psi_{n/2,\delta}$ for which $\|\Psi_{n/2,\delta} - M_{n/2}\|_\infty \le 2\delta \log(n/2)$, and

$$|\partial^{(\beta)}\Psi_{n/2,\delta}| \le \frac{1}{C_1} n^{\log_2 C_1} \quad \forall|\beta| = 1, \qquad |\partial^{(\beta)}\Psi_{n/2,\delta}| \le \frac{C_2}{\delta C_1^2} n^{2\log_2 C_1} \log_2(n/2) \quad \forall|\beta| = 2,$$

$$|\partial^{(\beta)}\Psi_{n/2,\delta}| \le \frac{3C_2^2 C_3}{\delta^2 C_1^3} n^{3\log_2 C_1} \log_2^2(n/2) \quad \forall|\beta| = 3.$$

Construct

$$\Psi_{n,\delta}(x_1, \ldots, x_n) := \Psi_{n/2,\delta}(\Psi_{2,\delta}(x_1, x_2), \ldots, \Psi_{2,\delta}(x_{n-1}, x_n)).$$

Because $\mathrm{Max}_{n/2}$ is 1-Lipschitz,

$$\begin{aligned}\|\Psi_{n,\delta} - \mathrm{Max}_n\|_\infty &\le \|\Psi_{n/2,\delta} - \mathrm{Max}_{n/2}\|_\infty + \|\Psi_{2,\delta} - \mathrm{Max}_2\|_\infty \\ &\le 2\delta \log(n/2) + 2\delta = 2\delta \log n.\end{aligned}$$

For all k, let $\|\partial^{(r)}\Psi_{k,\delta}\|_\infty$ denote the maximum over all β with $|\beta| = 1$ of $\|\partial^{(\beta)}\Psi_{k,\delta}\|_\infty$. The remainder of this proof makes liberal use of the chain rule. We may bound the first derivatives of $\Psi_{n,\delta}$ by

$$\|\partial^{(1)}\Psi_{n,\delta}\|_\infty \le \|\partial^{(1)}\Psi_{n/2,\delta}\|_\infty \cdot \|\partial^{(1)}\Psi_{2,\delta}\|_\infty. \le \frac{1}{C_1} n^{\log_2 C_1} \cdot C_1 = n^{\log_2 C_1},$$

where the second inequality uses the inductive hypothesis and the base case.

We may next bound the second derivatives of $\Psi_{n,\delta}$ by

$$\begin{aligned}\|\partial^{(2)}\Psi_{n,\delta}\|_\infty &\le \|\partial^{(2)}\Psi_{n/2,\delta}\|_\infty \cdot \|\partial^{(1)}\Psi_{2,\delta}\|_\infty^2 + \|\partial^{(1)}\Psi_{n/2,\delta}\|_\infty \cdot \|\partial^{(2)}\Psi_{2,\delta}\|_\infty \\ &\le \frac{C_2}{\delta C_1^2} n^{2\log_2 C_1} \log_2(n/2) \cdot C_1^2 + \frac{1}{C_1} n^{\log_2 C_1} \cdot \frac{C_2}{\delta} \\ &\le \frac{C_2}{\delta} n^{2\log_2 C_1} (\log_2(n/2) + 1) \le \frac{C_2}{\delta} n^{2\log_2 C_1} \log_2 n.\end{aligned}$$

Finally, we may bound the third derivatives of $\Psi_{n,\delta}$ by

$$\begin{aligned}\|\partial^{(3)}\Psi_{n,\delta}\|_\infty &\le \|\partial^{(3)}\Psi_{n/2,\delta}\|_\infty \cdot \|\partial^{(1)}\Psi_{2,\delta}\|_\infty^3 + 3\|\partial^{(2)}\Psi_{n/2,\delta}\|_\infty \cdot \|\partial^{(2)}\Psi_{2,\delta}\|_\infty \cdot \|\partial^{(1)}\Psi_{2,\delta}\|_\infty \\ &\quad + \|\partial^{(1)}\Psi_{n/2,\delta}\|_\infty \cdot \|\partial^{(3)}\Psi_{2,\delta}\|_\infty \\ &\le \frac{3C_2^2C_3}{\delta^2 C_1^3} n^{3\log_2 C_1} \log_2^2(n/2) \cdot C_1^3 + 3\frac{C_2}{\delta C_1^2} n^{2\log_2 C_1} \log_2(n/2) \cdot \frac{C_2}{\delta} \cdot C_1 + \frac{1}{C_1} n^{\log_2 C_1} \cdot \frac{C_3}{\delta^2} \\ &\le \frac{3C_2^2C_3}{\delta^2} n^{3\log_2 C_1} \log_2^2(n/2) + \frac{3C_2^2C_3}{\delta^2} n^{3\log_2 C_1} \log_2 n \le \frac{3C_2^2C_3}{\delta^2} n^{3\log_2 C_1} \log_2^2 n. \qquad \square\end{aligned}$$

D Padding

We show here that a (c, s) point of NP-hardness for a CSP immediately yields other related points of NP-hardness.

Lemma D.1. *For a given CSP Γ, let c_0 be the infimum of $\mathrm{Opt}(\mathcal{I})$ over all instances $\mathcal{I}$ of Γ. Then if (c, s) is a point of NP-hardness for Γ, so is any convex combination of (c, s), (c_0, c_0), and $(1, 1)$.*

Proof. Consider any convex combination $\lambda_1(c, s) + \lambda_2(c_0, c_0) + \lambda_3(1, 1)$, where $\lambda_i >= 0$, for all i, and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. For $0 \le \alpha \le 1$, we will show a simple method of mapping α-satisfiable instances of Γ into $(\lambda_1\alpha + \lambda_2 c_0 + \lambda_3)$-satisfiable instances of Γ. This is enough to prove the lemma. Let $\mathcal{I}$ be the c_0-satisfiable instance of Γ, and let ϕ be a (satisfiable) predicate in Γ.

Let $\mathcal{J}$ be an instance of Γ with optimum α. Consider the Γ instance $\mathcal{J}'$ constructed as follows: copy over $\mathcal{J}$ into $\mathcal{J}'$, but multiply all of its weights by λ_1. Next, copy over $\mathcal{I}$ into $\mathcal{J}'$ (with a set of variables disjoint to $\mathcal{J}$), but multiply all of its weights by λ_2. Finally, introduce the constraint ϕ over k new variables, and give it weight λ_3. As the three parts of $\mathcal{J}'$ contain disjoint variables, the optimum assignment to $\mathcal{J}'$ just assigns the variables to each part optimally. This means that $\mathrm{Opt}(\mathcal{J}') = \lambda_1\alpha + \lambda_2 c_0 + \lambda_3$. $\square$

E Upper Bound

Let 2Lin(Z_q)-Test′ denote the uncorrelated version 2Lin(Z_q)-Test. We derived the upper bound on the optimum value of 2Lin(Z_q)-Test′ for $2 < q < 7$ by analyzing the following test:

2Lin(Z_q)-Upper-Test

- Given functions $f, g : Z_q^2 \to Z_q$ and $h : (Z_q \cup \{*\})^4 \to Z_q$:
- Pick x_1, x_2, y_1, and y_2 uniformly at random from Z_q.
- Give h either:

1. $(x_1, *, y_1, *)$
2. $(x_1, *, *, y_2)$
3. $(*, x_2, y_1, *)$
4. $(*, x_2, *, y_2)$

- Test h against either $f(x_1, x_2)$ or $g(y_1, y_2)$, each with equal probability.

We claim the following:

Theorem E.1. $\mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_q)\text{-}\textsc{Upper-Test}) \geq \mathrm{Opt}_{\mathrm{folded}}(2\mathsf{Lin}(Z_q)\text{-}\textsc{Test}')$.

Proof. Let (f, g, h) be a folded strategy for $2\mathsf{Lin}(Z_q)$-$\textsc{Test}'$. We will construct a folded strategy (f', g', h') for which $\mathrm{Val}_{2\mathsf{Lin}(Z_q)\text{-}\textsc{Upper-Test}}(f', g', h') \geq \mathrm{Val.}(f, g, h)$. Consider selecting $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$ independently from $(\mu_{q,1/2})^{\otimes n}$ and forming $\boldsymbol{x}$ (respectively, $\boldsymbol{y}$) from $\widetilde{\boldsymbol{x}}$ (respectively, $\widetilde{\boldsymbol{y}}$) by filling in the $*$'s with uniformly random elements of Z_q. Now, let $\widetilde{\boldsymbol{x}}'$ be the unique element of $(Z_q \cup *)^n$ for which $\widetilde{\boldsymbol{x}} \circ \widetilde{\boldsymbol{x}}' = \boldsymbol{x}$, and similarly for $\widetilde{\boldsymbol{y}}'$. Then $\widetilde{\boldsymbol{x}}'$ and $\widetilde{\boldsymbol{y}}'$ are also distributed as strings drawn from $(\mu_{q,1/2})^{\otimes n}$; furthermore, $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{x}}'$ each are independent of both of $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{y}}'$, and vice versa.

Next, pick a, b, c, and d independently and uniformly at random from Z_q. Then

$$\mathrm{Val}_{2\mathsf{Lin}(Z_q)\text{-}\textsc{Test}'}(f, g, h) = \mathrm{avg} \begin{cases} \mathbf{Pr}[f((\widetilde{\boldsymbol{x}} + a) \circ (\widetilde{\boldsymbol{x}}' + b)) = h(\widetilde{\boldsymbol{x}} + a, \widetilde{\boldsymbol{y}} + c)], \\ \mathbf{Pr}[f((\widetilde{\boldsymbol{x}} + a) \circ (\widetilde{\boldsymbol{x}}' + b)) = h(\widetilde{\boldsymbol{x}} + a, \widetilde{\boldsymbol{y}}' + d)], \\ \mathbf{Pr}[g((\widetilde{\boldsymbol{y}} + c) \circ (\widetilde{\boldsymbol{y}}' + d)) = h(\widetilde{\boldsymbol{x}}' + b, \widetilde{\boldsymbol{y}} + c)], \\ \mathbf{Pr}[g((\widetilde{\boldsymbol{y}} + c) \circ (\widetilde{\boldsymbol{y}}' + d)) = h(\widetilde{\boldsymbol{x}}' + b, \widetilde{\boldsymbol{y}}' + d)]. \end{cases}$$

By the probabilistic method, there must exist strings $\widetilde{x}$, $\widetilde{x}'$, $\widetilde{y}$, and $\widetilde{y}'$ for which, conditioned on these strings, the average of the four probabilities, α, is at least $\mathrm{Val.}(f, g, h)$. (The randomness is now only over the choice of a, b, c, and d.) Now, set

$$f'(a, b) := f((\widetilde{x} + a) \circ (\widetilde{x}' + b)), \quad g'(c, d) := g((\widetilde{y} + c) \circ (\widetilde{y}' + d)),$$

and define h' as follows:

$$h'(a, *, c, *) := h(\tilde{x} + a, \tilde{y} + c), \quad h'(a, *, *, d) := h(\tilde{x} + a, \tilde{y}' + d),$$

$$h'(*, b, c, *) := h(\tilde{x}' + b, \tilde{y} + c), \quad \text{and} \quad h'(*, b, *, d) := h(\tilde{x}' + b, \tilde{y}' + d).$$

Because f and g are folded, so are f' and g'. Furthermore, it is easy to see that $\mathrm{Val}_{2\mathsf{Lin}(Z_q)\text{-}\textsc{Upper-Test}}(f', g', h') = \alpha$. By the definition of α, this is at least $\mathrm{Val}_{2\mathsf{Lin}(Z_q)\text{-}\textsc{Test}'}(f, g, h)$, which concludes the proof. □

This dramatically simplifies our problem: to prove an upper bound for the optimum of $2\mathsf{Lin}(Z_q)$-$\textsc{Test}'$, we need only show one for the optimum of $2\mathsf{Lin}(Z_q)$-$\textsc{Upper-Test}$. As f and g are folded and only have two inputs apiece in $2\mathsf{Lin}(Z_q)$-$\textsc{Upper-Test}$, this problem becomes feasible, at least for small values of q. There are q^{2q} total choices for f and g. For any fixed f and g, we may find and evaluate the optimal h in time $O(q^3)$ as follows: there are $4q^2$ possible inputs to h. For each input, there are q possible inputs for the remaining coordinate of f, and q possible inputs for the remaining coordinate of g. The optimal h will simply output the value which occurs most frequently among the outputs of f and g, and it succeeds with probability the number of times this value occurs divided by $2q$. This gives an $O(q^3 q^{2q})$ algorithm for determining $\mathrm{Opt}(2\mathsf{Lin}(Z_q)\text{-}\textsc{Upper-Test})$. We coded a Java implementation, which ran quickly (enough) when $q < 7$, and found the following:

Theorem E.2. $\mathrm{Opt}(2\mathsf{Lin}(Z_q)\text{-}\textsc{Upper-Test}) = 3/8 + 5/8q$ *when* $q < 7$. *Thus, for these values of* q, $\mathrm{Opt}(2\mathsf{Lin}(Z_q)\text{-}\textsc{Test}')$ *is at most* $3/8 + 5/8q$ *as well.*

F A Fourier-based proof of $2\mathsf{Lin}(Z_2)$ hardness

In this section, we give a Fourier-analytic proof of Theorem 2.4 for $2\mathsf{Lin}(Z_2)$. We assume familiarity with Fourier analysis over Z_2. We are interested in showing that it is hard to $(c-2\epsilon, s+2\epsilon)$-decide the $2\mathsf{Lin}(Z_2)$ problem. Consider the $2\mathsf{Lin}(Z_2)$ instance produced by running the ϵ-noisy version of the $2\mathsf{Lin}(Z_2)$-$\textsc{Test}_d$ through Theorem 4.19. We have already covered the completeness case, and we will jump into the soundness case at the point where we have found the "good" edges — those edges $(u,v) \in E$ for which $\mathrm{Val}_{2\mathsf{Lin}(Z_2)\text{-}\textsc{Test}_d}(f_u, g_v, h_{uv}) \geq \frac{11}{16} + \epsilon$. Let (u,v) be such a good edge, and set $f := f_u$, $g := g_v$, and $h := h_{uv}$. We will now sketch how to complete the proof, showing how we can use Fourier analysis to decode the functions into a good assignment.

Proof sketch of Theorem 2.4 for $2\mathsf{Lin}(Z_2)$. Generate $\boldsymbol{x}$, $\widetilde{\boldsymbol{x}}$, $\dot{\boldsymbol{x}}$, $\boldsymbol{y}$, $\widetilde{\boldsymbol{y}}$, and $\dot{\boldsymbol{y}}$ as in $2\mathsf{Lin}(Z_2)$-$\textsc{Test}$. Define $\boldsymbol{x}' \in \{-1,1\}^{dK}$ to be $\boldsymbol{x}'[t] = \boldsymbol{x}[t]$ if $\widetilde{\boldsymbol{x}}[t] \neq *^d$ and $\boldsymbol{x}'[t] = -\boldsymbol{x}[t]$ otherwise. Let $\dot{\boldsymbol{x}}'$ be a $(1-\epsilon)$-correlated copy of $\boldsymbol{x}'$. Define $\boldsymbol{y}'$ and $\widetilde{\boldsymbol{y}}'$ similarly. Note that $(\boldsymbol{x}', (\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}))$ has the same distribution as $(\boldsymbol{x}, (\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}))$, and $(\boldsymbol{y}', (\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}))$ has the same distribution as $(\boldsymbol{y}, (\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}))$.

$$\begin{aligned}\mathbf{Pr}[\text{test fails}] &= \mathrm{avg}\{\mathbf{Pr}[f(\dot{\boldsymbol{x}}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \mathbf{Pr}[g(\dot{\boldsymbol{y}}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})]\} \\ &= \mathrm{avg}\left\{\begin{array}{l}\mathbf{Pr}[f(\dot{\boldsymbol{x}}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[f(\dot{\boldsymbol{x}}') \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[g(\dot{\boldsymbol{y}}) \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})], \\ \mathbf{Pr}[g(\dot{\boldsymbol{y}}') \neq h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})]\end{array}\right\} \geq \frac{1}{4}\,\mathbf{E}[\mathsf{mixed}(f(\dot{\boldsymbol{x}}), f(\dot{\boldsymbol{x}}'), g(\dot{\boldsymbol{y}}), g(\dot{\boldsymbol{y}}'))], \end{aligned} \tag{26}$$

where

$$\mathsf{mixed}(c_1,c_2,c_3,c_4) = \begin{cases} 0 & \text{if all four inputs are the same,} \\ 1 & \text{if exactly three inputs are the same and one is different,} \\ 2 & \text{if exactly two of the inputs are the same and two are different.}\end{cases}$$

This is because whatever $f(\dot{\boldsymbol{x}})$, $f(\dot{\boldsymbol{x}}')$, $g(\dot{\boldsymbol{y}})$, and $g(\dot{\boldsymbol{y}}')$ turn out to be, $h(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ must differ from at least $\mathsf{mixed}(f(\dot{\boldsymbol{x}}), f(\dot{\boldsymbol{x}}'), g(\dot{\boldsymbol{y}}), g(\dot{\boldsymbol{y}}'))$ of them.

Next, let $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4 \in \{-1,1\}$ be chosen uniformly at random conditioned on $\boldsymbol{\lambda}_1\boldsymbol{\lambda}_2\boldsymbol{\lambda}_3\boldsymbol{\lambda}_4 = 1$. There are eight such possibilities. It is the case that $(\dot{\boldsymbol{x}}, \dot{\boldsymbol{y}}', \dot{\boldsymbol{y}}, \dot{\boldsymbol{y}}')$ has the same distribution as $(\boldsymbol{\lambda}_1\dot{\boldsymbol{x}}, \boldsymbol{\lambda}_2\dot{\boldsymbol{y}}', \boldsymbol{\lambda}_3\dot{\boldsymbol{y}}, \boldsymbol{\lambda}_4\dot{\boldsymbol{y}}')$. To see this, it suffices to show that $(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}, \boldsymbol{y}')$ has the same distribution as $(\boldsymbol{\lambda}_1\boldsymbol{x}, \boldsymbol{\lambda}_2\boldsymbol{x}', \boldsymbol{\lambda}_3\boldsymbol{y}, \boldsymbol{\lambda}_4\boldsymbol{y}')$. This holds because for each block t, it is either the case that

$$\boldsymbol{x}[t] = \boldsymbol{x}'[t] \text{ and } \boldsymbol{y}[t] = -\boldsymbol{y}'[t]$$

or

$$\boldsymbol{x}[t] = -\boldsymbol{x}[t] \text{ and } \boldsymbol{y}[t] = \boldsymbol{y}'[t],$$

each with equal probability. Flipping an even number of signs maintains this.

Note that for fixed $c_1, c_2, c_3, c_4 \in \{-1,1\}$, if $\mathsf{mixed}(c_1,c_2,c_3,c_4) = 1$ (i.e., $c_1c_2c_3c_4 = -1$), then $\mathsf{mixed}(c_1\boldsymbol{\lambda}_1, c_2\boldsymbol{\lambda}_2, c_3\boldsymbol{\lambda}_3, c_4\boldsymbol{\lambda}_4) = 1$ always. On the other hand, if $\mathsf{mixed}(c_1,c_2,c_3,c_4) = 0$ or 2 (i.e., $c_1c_2c_3c_4 = 1$), then there are two settings of the four $\boldsymbol{\lambda}$'s for which $\mathsf{mixed}(c_1\boldsymbol{\lambda}_1, c_2\boldsymbol{\lambda}_2, c_3\boldsymbol{\lambda}_3, c_4\boldsymbol{\lambda}_4) = 0$ and six settings for which $\mathsf{mixed}(c_1\boldsymbol{\lambda}_1, c_2\boldsymbol{\lambda}_2, c_3\boldsymbol{\lambda}_3, c_4\boldsymbol{\lambda}_4) = 2$. In other words, in this case, $\mathbf{E}[\mathsf{mixed}(c_1\boldsymbol{\lambda}_1, c_2\boldsymbol{\lambda}_2, c_3\boldsymbol{\lambda}_3, c_4\boldsymbol{\lambda}_4)] = 3/2$, where the expectation is taken over the choice of $\boldsymbol{\lambda}$'s. Therefore,

$$
\begin{aligned}
&\frac{1}{4}\,\mathbf{E}[\mathsf{mixed}(f(\dot{\boldsymbol{x}}), f(\dot{\boldsymbol{x}}'), g(\dot{\boldsymbol{y}}), g(\dot{\boldsymbol{y}}'))] \\
=&\frac{1}{4}\,\mathbf{E}[\mathsf{mixed}(f(\boldsymbol{\lambda}_1\dot{\boldsymbol{x}}), f(\boldsymbol{\lambda}_2\dot{\boldsymbol{x}}'), g(\boldsymbol{\lambda}_3\dot{\boldsymbol{y}}), g(\boldsymbol{\lambda}_4\dot{\boldsymbol{y}}'))] \\
=&\frac{1}{4}\,\mathbf{E}\left[1\cdot\mathbf{1}[f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')=-1]+\frac{3}{2}\cdot\mathbf{1}[f(\dot{\boldsymbol{x}}')f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')=1]\right] \\
=&\frac{1}{4}\,\mathbf{E}\left[\frac{5}{4}+\frac{1}{4}f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')\right] \\
=&\frac{5}{16}+\frac{1}{16}\,\mathbf{E}[f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')].
\end{aligned}
$$

By the assumption of the lemma, the probability that f, g, and h fail this test is at most $5/16-\epsilon$. Thus,

$$\mathbf{E}[f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')]\le -16\epsilon. \tag{27}$$

We now complete the proof using Fourier analysis. The first step is to show the following claim via a standard computation:

Claim F.1. $\mathbf{E}[f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')]=\sum_{S,T\subseteq[dK]}(1-\epsilon)^{2|S|}\hat{f}(S)^2(1-\epsilon)^{2|T|}\hat{g}(T)^2\prod_{i=1}^{K}\mathsf{same}(|S[k]|,|T[k]|)$,

where

$$\mathsf{same}(c,d)=\begin{cases}1 & \text{if } c \text{ and } d \text{ are both even,}\\ -1 & \text{if } c \text{ and } d \text{ are both odd,}\\ 0 & \text{otherwise.}\end{cases}$$

Proof.

$$\mathbf{E}[f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')]=\sum_{S,T,U,V\subseteq[dK]}\hat{f}(S)\hat{f}(T)\hat{g}(U)\hat{g}(V)\,\mathbf{E}[\chi_S(\dot{\boldsymbol{x}})\chi_T(\dot{\boldsymbol{x}}')\chi_U(\dot{\boldsymbol{y}})\chi_V(\dot{\boldsymbol{y}}')]$$

For a set $S\subseteq[dK]$, write $S(t):=S\cap B(t)$. The expectation on the far right is

$$\mathbf{E}[\chi_S(\dot{\boldsymbol{x}})\chi_T(\dot{\boldsymbol{x}}')\chi_U(\dot{\boldsymbol{y}})\chi_V(\dot{\boldsymbol{y}}')]=\prod_{t=1}^{K}\mathbf{E}[\chi_{S(t)}(\dot{\boldsymbol{x}})\chi_{T(t)}(\dot{\boldsymbol{x}}')\chi_{U(t)}(\dot{\boldsymbol{y}})\chi_{V(t)}(\dot{\boldsymbol{y}}')]$$

If for any t it is the case that $S(t)\neq T(t)$, then the expectation will be 0. The same holds for U and V. Thus, we need only concern ourselves with the case when $S=T$ and $U=V$, meaning that

$$\mathbf{E}[f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')]=\sum_{S,T\subseteq[dK]}\hat{f}(S)^2\hat{g}(T)^2\prod_{t=1}^{K}\mathbf{E}[\chi_{S(t)}(\dot{\boldsymbol{x}})\chi_{S(t)}(\dot{\boldsymbol{x}}')\chi_{T(t)}(\dot{\boldsymbol{y}})\chi_{T(t)}(\dot{\boldsymbol{y}}')] \tag{28}$$

To understand the expectation for the tth block, let's first consider it in its noiseless form,

$$\mathbf{E}[\chi_{S(t)}(\boldsymbol{x})\chi_{S(t)}(\boldsymbol{x}')\chi_{T(t)}(\boldsymbol{y})\chi_{T(t)}(\boldsymbol{y}')]. \tag{29}$$

With probability $1/2$, $\widetilde{\boldsymbol{x}}[t]\neq *^d$ and $\widetilde{\boldsymbol{y}}[t]=*^d$, and with probability $1/2$, $\widetilde{\boldsymbol{x}}[t]=*^d$ and $\widetilde{\boldsymbol{y}}[t]\neq *^d$. Let's focus on the first case. Whenever $\widetilde{\boldsymbol{x}}[t]\neq *^d$, $\boldsymbol{x}[t]=\boldsymbol{x}'[t]$ and $\boldsymbol{y}[t]=-\boldsymbol{y}'[t]$. The conclusion is that $\chi_{S(t)}(\boldsymbol{x})\chi_{S(t)}(\boldsymbol{x}')=1$ and $\chi_{T(t)}(\boldsymbol{y})\chi_{T(t)}(\boldsymbol{y}')=(-1)^{|T(t)|}$, meaning that the entire expectation

equals $(-1)^{|T(t)|}$. Similarly, when $\widetilde{\boldsymbol{y}}[t] \neq *^d$, the expectation equals $(-1)^{|S(t)|}$. Thus, the quantity in (29) reduces to

$$\frac{1}{2}((-1)^{|S(t)|} + (-1)^{|T(t)|}) = \mathsf{same}(|S(t)|, |T(t)|).$$

From here, adding the noise back in is simple. Let $\boldsymbol{z}_1 \in \{-1, 1\}^{dK}$ be a randomly chosen string with $\mathbf{E}[(\boldsymbol{z}_1)_i] = 1 - \epsilon$. Then $\dot{\boldsymbol{x}}$ is distributed as $\boldsymbol{z}_1 \circ \boldsymbol{x}$, where $\circ$ denotes coordinate-wise multiplication. We can define $\boldsymbol{z}_2$, $\boldsymbol{z}_3$, and $\boldsymbol{z}_4$ identically and note that $\dot{\boldsymbol{x}}'$ is distributed as $\boldsymbol{z}_2 \circ \boldsymbol{x}'$, $\boldsymbol{y}_1$ is distributed as $\boldsymbol{z}_3 \circ \boldsymbol{y}$, and $\dot{\boldsymbol{y}}'$ is distributed as $\boldsymbol{z}_4 \circ \boldsymbol{y}'$. Then

$$\begin{aligned}
&\mathbf{E}[\chi_{S(t)}(\dot{\boldsymbol{x}})\chi_{S(t)}(\dot{\boldsymbol{x}}')\chi_{T(t)}(\dot{\boldsymbol{y}})\chi_{T(t)}(\dot{\boldsymbol{y}}')] \\
=&\mathbf{E}[\chi_{S(t)}(\boldsymbol{x})\chi_{S(t)}(\boldsymbol{x}')\chi_{T(t)}(\boldsymbol{y})\chi_{T(t)}(\boldsymbol{y}')\chi_{S(t)}(\boldsymbol{z}_1)\chi_{S(t)}(\boldsymbol{z}_2)\chi_{T(t)}(\boldsymbol{z}_3)\chi_{T(t)}(\boldsymbol{z}_4)] \\
=&\ \mathsf{same}(|S(t)|, |T(t)|) \cdot (1-\epsilon)^{2|S(t)|}(1-\epsilon)^{2|T(t)|}.
\end{aligned}$$

This means that we can write equation (28) as

$$\begin{aligned}
&\mathbf{E}[f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')] \\
=&\sum_{S,T\subseteq[dK]} \hat{f}(S)^2\hat{g}(T)^2 \prod_{k=1}^{K} \mathsf{same}(|S(t)|, |T(t)|) \cdot (1-\epsilon)^{2|S(t)|}(1-\epsilon)^{2|T(t)|} \\
=&\sum_{S,T\subseteq[dK]} (1-\epsilon)^{2|S|}\hat{f}(S)^2(1-\epsilon)^{2|T|}\hat{g}(T)^2 \prod_{i=k}^{K} \mathsf{same}(|S(t)|, |T(t)|), \qquad (30)
\end{aligned}$$

exactly as claimed. □

Note that the product on the right is nonzero exactly when $|S(t)| \equiv |T(t)| \pmod 2$ for all t. We now find it convenient to state the following definition: for two sets S, $T \subseteq [dK]$, we write $S \equiv_d T$ if $|S(t)| \equiv |T(t)| \pmod 2$ for all t. Plugging this into (30) yields:

$$\begin{aligned}
\mathbf{E}[f(\dot{\boldsymbol{x}})f(\dot{\boldsymbol{x}}')g(\dot{\boldsymbol{y}})g(\dot{\boldsymbol{y}}')] &= \sum_{S,T\subseteq[dK]} (1-\epsilon)^{2|S|}\hat{f}(S)^2(1-\epsilon)^{2|T|}\hat{g}(T)^2 \prod_{i=k}^{K} \mathsf{same}(|S(t)|, |T(t)|) \\
&\geq -\sum_{S,T\subseteq[dK]} \hat{f}(S)^2\hat{g}(T)^2\mathbf{1}[S \equiv_d T](1-\epsilon)^{2|S|+2|T|}
\end{aligned}$$

Combining this with (27),

$$16\epsilon \leq \sum_{S,T\subseteq[dK]} \hat{f}(S)^2\hat{g}(T)^2 \cdot \mathbf{1}[S \equiv_d T] \cdot (1-\epsilon)^{2|S|+2|T|}.$$

Finally, since f and g are folded we may equivalently write

$$16\epsilon \leq \sum_{|S|\equiv|T|\equiv 1 \pmod 2} \hat{f}(S)^2\hat{g}(T)^2 \cdot \mathbf{1}[S \equiv_d T] \cdot (1-\epsilon)^{2|S|+2|T|}.$$

Note that when both $|S| \equiv |T| \equiv 1 \pmod 2$ and $S \equiv_d T$, there must exist some t such that $|S(t)| \equiv |T(t)| \equiv 1 \pmod 2$; in particular, both $S(t)$ and $T(t)$ are nonempty. It is therefore easy to check that the "Håstad-style decoding" procedure from [Hås01] succeeds on f and g, completing the proof (sketch). □

G Constructing the 2Lin(Z_2)-Test

In this section, we show how the 2Lin(Z_2)-Test can be viewed as a composition of Håstad's two-function linearity tests [Hås01] with the gadget of [TSSW00]. Interestingly, we could find no similar interpretation for our 2Lin(Z_q)-Test, which suggests that the function-in-the-middle approach may be necessary for deriving it. The particular linearity test of Håstad we will use as a starting point is the 4Lin(Z_2) test (a natural extension of his 3Lin(Z_2) test), which we state below without noise:

4Lin(Z_2)-Test

- Given functions $f : \{0,1\}^K \to \{0,1\}$ and $g : \{0,1\}^{dK} \to \{0,1\}$:
- Choose $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \{0,1\}^K$ and $\boldsymbol{y}_1 \in \{0,1\}^{dK}$ independently and uniformly at random.
- Form $\boldsymbol{y}_2 \in \{0,1\}^K$ by setting $\boldsymbol{y}_2 := 1 + (\boldsymbol{x}_1)_\pi + (\boldsymbol{x}_2)_\pi + \boldsymbol{y}_1$.
- Test $f(\boldsymbol{x}_1) + f(\boldsymbol{x}_2) + g(\boldsymbol{y}_1) + g(\boldsymbol{y}_2) \equiv 1 \pmod 2$.

The 4Lin(Z_2)-to-2Lin(Z_2) gadget: Now, we will extrapolate from [TSSW00] the 4Lin(Z_2)-to-2Lin(Z_2) gadget. We're given the 4Lin(Z_2) equation

$$x_1 + x_2 + x_3 + x_4 \equiv b \pmod 2,$$

where $x_i, b \in \{0,1\}$. Pick random $y_1, y_2, y_3, y_4 \in \{0,1\}$ satisfying $y_1 + y_2 + y_3 + y_4 \equiv 1 - b \pmod 2$. Note that

- If $x_1 + x_2 + x_3 + x_4 \equiv b \pmod 3$, then $(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) + (x_4 + y_4) \equiv 1 \pmod 2$.
- If $x_1 + x_2 + x_3 + x_4 \not\equiv b \pmod 3$, then $(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) + (x_4 + y_4) \equiv 0 \pmod 2$.

In the first case, the string $(x_1 + y_1, x_2 + y_2, x_3 + y_3, x_4 + y_4)$ is distributed like a randomly chosen string of four bits which sums to 1. Such a string will always have three entries the same and one entry which is different. In the second case, the string is distributed like a randomly chosen string of four bits which sum to zero. Such a string will with probability 3/4 have two 0's and two 1s and will with probability 1/4 be either all 0's or all 1's. So, introduce a new variable m, and generate the set of 2Lin(Z_2) equations

$$(x_i + y_i) - m \equiv 0 \pmod 3, i \in [4].$$

Clearly, the m which maximizes the number of these 2Lin(Z_2) equations is the majority bit of $(x_1 + y_1, x_2 + y_2, x_3 + y_3, x_4 + y_4)$. In the first case, this m will satisfy 3/4 of the equations. In the second case, it will satisfy $3/4 * 1/2 + 1/4 * 1 = 5/8$ of the equations. Because 4Lin(Z_2) is NP-hard to $(1 - \epsilon, \frac{1}{2} + \epsilon)$-approximate, this gives NP-hardness of $(\frac{3}{4}, \frac{11}{16} + \epsilon)$-approximating 2Lin($Z_2$).

Composing the two: Let us begin by considering the case in the 4Lin(Z_2)-Test when f and g are matching dictators, though we don't necessarily know which ones. How can we determine the majority bit of $(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), g(\boldsymbol{y}_1), g(\boldsymbol{y}_2))$? Set $\boldsymbol{m} \in \{0,1\}^{dK}$ to be the bitwise majority of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{y}_1$, and $\boldsymbol{y}_2$. Then the majority bit we're looking for is simply the corresponding dictator bit of $\boldsymbol{m}$. Thus, as a preliminary composition step, consider the following test, where we have added a new function h:

Preliminary $2\mathsf{Lin}(Z_2)$-Test

- Given functions $f : \{0,1\}^K \to \{0,1\}$ and $g, h : \{0,1\}^{dK} \to \{0,1\}$:
- Choose $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \{0,1\}^K$ and $\boldsymbol{y}_1 \in \{0,1\}^{dK}$ independently and uniformly at random.
- Form $\boldsymbol{y}_2 \in \{0,1\}^K$ by setting $\boldsymbol{y}_2 := 1 + (\boldsymbol{x}_1)_\pi + (\boldsymbol{x}_2)_\pi + \boldsymbol{y}_1$.
- Set $\boldsymbol{m} \in \{0,1\}^{dK}$ to be the bitwise majority of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{y}_1$, and $\boldsymbol{y}_2$.
- Test either $f(\boldsymbol{x}_1) = h(\boldsymbol{m})$, $f(\boldsymbol{x}_2) = h(\boldsymbol{m})$, $g(\boldsymbol{y}_1) = h(\boldsymbol{m})$, or $g(\boldsymbol{y}_2) = h(\boldsymbol{m})$, each with equal probability.

Note that in relation to everything else, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are identically distributed. This means, for instance, that the $f(\boldsymbol{x}_2) = h(\boldsymbol{m})$ test can be replaced with a second $f(\boldsymbol{x}_1) = h(\boldsymbol{m})$ test. Similarly, in relation to everything else $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are identically distributed. As a result, we need only test $g(\boldsymbol{y}_1) = h(\boldsymbol{m})$. The result of these two simplifications is that the test in the last step is equivalent to testing $f(\boldsymbol{x}_1) = h(\boldsymbol{m})$ or $g(\boldsymbol{y}_1) = h(\boldsymbol{m})$, each with probability $1/2$.

Next, condition the test on some values for $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$, and consider the distribution on $\boldsymbol{m}$ that this induces. For any block $i \in [K]$, whenever $(\boldsymbol{x}_1)_i = (\boldsymbol{x}_2)_i$, which happens with $1/2$ probability, then $\boldsymbol{y}_2[i] = 1 + \boldsymbol{y}_1[i]$. In other words, $\boldsymbol{y}_1[i]$ and $\boldsymbol{y}_2[i]$ are exact opposites, so $\boldsymbol{m}[i]$ is uniquely defined to be $((\boldsymbol{x}_1)_i)^d$. On the other hand, when $(\boldsymbol{x}_1)_i = 1 + (\boldsymbol{x}_2)_i$, which happens with the remaining $1/2$ probability, then $\boldsymbol{y}_2[i] = \boldsymbol{y}_1[i]$. As $(\boldsymbol{x}_1)_i$ and $(\boldsymbol{x}_2)_i$ are exact opposities, $\boldsymbol{m}[i]$ is uniquely defined to be $\boldsymbol{y}_1[i]$. Thus, conditioned on $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$, $\boldsymbol{m}[i]$ is always either $((\boldsymbol{x}_1)_i)^d$ or $\boldsymbol{y}_1[i]$, each with probability $1/2$. Although for each block h is not "told" which event occurred, it is easy for it to "figure out" (with high probability) which of $\boldsymbol{x}_1$ or $\boldsymbol{y}_1$ a given block in $\boldsymbol{m}$ came from: if $\boldsymbol{m}[i]$ is all 0's or 1's, then it very likely came from $\boldsymbol{x}_1$. Otherwise, it must have come from $\boldsymbol{y}_1$.

At this point, as we no longer compare $f(\boldsymbol{x}_2)$ with $h(\boldsymbol{m})$ or $g(\boldsymbol{y}_2)$ with $h(\boldsymbol{m})$, and we have found how to generate $\boldsymbol{m}$ directly from $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$, we no longer need any reference to $\boldsymbol{x}_2$ or $\boldsymbol{y}_2$. Furthermore, as h is effectively given the information about which blocks of $\boldsymbol{m}$ came from where because constantly 0 or 1 blocks so conclusively point to $\boldsymbol{x}_1$ as the source, we might as well give it this information explicitly. With these changes and simplifications, it is clear that we have derived the $2\mathsf{Lin}(Z_2)$-Test as stated previously in the paper.

Going backwards: We have now shown a (somewhat inexact) methodology for composing a test with a gadget to get a function-in-the-middle test. It is natural to ask whether this process can be reversed: if one is given a function-in-the-middle test which can be naturally viewed as the composition of a test with a gadget, is it possible to recover the test and the gadget? In fact, although this is not an exact science, it turns out that it is indeed possible: consider the Fourier analytic proof of $2\mathsf{Lin}(Z_2)$ hardness given in Appendix F. The proof begins by taking the $\boldsymbol{x}$ and $\boldsymbol{y}$ defined in the $2\mathsf{Lin}(Z_2)$-Test and generating from them the strings $\boldsymbol{x}'$ and $\boldsymbol{y}'$, which seemingly exist only to help the proof. It can be checked that these strings are distributed identically to the strings $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{y}_1$, and $\boldsymbol{y}_2$ from the above $4\mathsf{Lin}(Z_2)$-Test (accounting for the $\{-1,1\} \leftrightarrow \{0,1\}$ switch). This suggests that underlying the $2\mathsf{Lin}(Z_2)$-Test is another test T which checks some constraint involving $f(\boldsymbol{x})$, $f(\boldsymbol{x}')$, $g(\boldsymbol{y})$, and $g(\boldsymbol{y}')$.

Next, Equation 26 indicates that the test T probably uses a four bit predicate whose satisfying assignments tend to have larger majorities than the unsatisfying assignments do. The $4\mathsf{Lin}(Z_2)$ predicate is a natural choice which satisfies this condition.

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for Unique Games and related problems. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 563–572, 2010. 1, 2.1, 8

[AEH01] Gunnar Andersson, Lars Engebretsen, and Johan Håstad. A new way of using semidefinite programming with applications to linear equations mod p. *Journal of Algorithms*, 39(2):162–204, 2001. 2.1

[BHK+11] Boaz Barak, Aram Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractive inequalities, sums of squares proofs, and their applications. Manuscript, 2011. 8

[BOL90] Michael Ben-Or and Nathan Linial. Collective coin flipping. In Silvio Micali, editor, *Randomness and Computation*. Academic Press, New York, 1990. 5.2

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, 2011. 1, 2.1, 8

[CC02] Miroslav Chlebík and Janka Chlebíková. Approximation hardness of the Steiner tree problem on graphs. In *Proceedings of the 8th Annual Scandinavian Workshop on Algorithm Theory*, pages 95–99, 2002. 8

[CC04] Miroslav Chlebík and Janka Chlebíková. On approximation hardness of the minimum 2SAT-DELETION problem. In *Proceedings of the 29th Annual International Symposium on Mathematical Foundations of Computer Science*, pages 263–273, 2004. 2.1

[DH09] Irit Dinur and Prahladh Harsha. Composition of low-error 2-query PCPs using decodable PCPs. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 472–481, 2009. 2

[DMR09] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. *SIAM Journal on Computing*, 39(3):843–873, 2009. 2

[DS05] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1):439–485, 2005. 2.1

[Fei99] Uriel Feige. *Randomized rounding of semidefinite programs — variations on the Max-Cut example*, volume 1761 of *Lecture Notes in Computer Science*, pages 189–196. Springer, 1999. 1

[Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 543–543, 2002. 1

[FK94] Uriel Feige and Joe Kilian. Two prover protocols: low error at affordable rates. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, pages 172–183, 1994. 2

[FR04] Uriel Feige and Daniel Reichman. On systems of linear equations with two variables per equation. In *Proceedings of the 7th Annual International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pages 117–127, 2004. 1.1, 2.1, 2.1, 2.1

[GS11] Venkatesan Guruswami and Ali Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with PSD objectives. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, 2011. 1, 8

[GW95] Michel Goemans and David Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995. 1

[Hås97] Johan Håstad. Some optimal inapproximability results. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pages 1–10, 1997. 1, 1.1, 1.2

[Hås99] Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182(1):105–142, 1999. 1

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001. 1, 2.1, 2.1, 3, 3.2, 4.3, 5, F, G

[IP01] Russell Impagliazzo and Ramamohan Paturi. On the complexity of k-SAT. *Journal of Computer and System Sciences*, 62(2):367–375, 2001. 1

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symposium on Theory of Computing*, pages 767–775, 2002. 1, 2

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007. 1, 2

[KO09] Subhash Khot and Ryan O'Donnell. SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory of Computing*, 5(1):83–117, 2009. 1

[Kol10] Alexandra Kolla. Spectral algorithms for Unique Games. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity*, pages 122–130, 2010. 1

[LO00] Rafal Latała and Krzysztof Oleszkiewicz. Between Sobolev and Poincaré. In Vitali Milman and Gideon Schechtman, editors, *Geometric aspects of functional analysis*, pages 147–168. Springer, 2000. A.5

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1), 2010. 2, C

[Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010. 2, 4.2, 5.1, 5.1, A.3

[MR10] Dana Moshkovitz and Ran Roz. Two-query PCP with subconstant error. *Journal of the ACM*, 57(5):29, 2010. 1, 2, 2.4, 2.1

[New00] Alantha Newman. Approximating the maximum acyclic subgraph. Master's thesis, Massachusetts Institute of Technology, 2000. 8

[O'D02] Ryan O'Donnell. Hardness amplification within NP. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 751–760, 2002. 5.1

[Ole03] Krzysztof Oleszkiewicz. On a nonsymmetric version of the Khinchine–Kahane inequality. In Evariste Giné, Christian Houdré, and David Nualart, editors, *Stochastic inequalities and applications*, volume 56, pages 157–168. Birkhäuser, 2003. A.5

[OW08] Ryan O'Donnell and Yi Wu. An optimal SDP algorithm for Max-Cut, and equally optimal Long Code tests. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 335–344, 2008. 1, 2.1

[OWZ11] Ryan O'Donnell, Yi Wu, and Yuan Zhou. Hardness of Max-2Lin and Max-3Lin over integers, reals, and large cyclic groups. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity*, pages 23–33, 2011. 2

[PV06] Christos Papadimitriou and Santosh Vempala. On the approximability of the traveling salesman problem. *Combinatorica*, 26(1):101–120, 2006. 8

[Rag09] Prasad Raghavendra. *Approximating NP-hard problems: efficient algorithms and their limits*. PhD thesis, University of Washington, 2009. 1, 2

[Raz95] Ran Raz. A parallel repetition theorem. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing*, pages 447–456, 1995. 2

[Ste10] David Steurer. Subexponential algorithms for d-to-1 two-prover games and for certifying almost perfect expansion. Available at the author's website, 2010. 1, 2.1

[Ste11] David Steurer. Personal communication, 2011. 1

[TSSW00] Luca Trevisan, Gregory Sorkin, Madhu Sudan, and David Williamson. Gadgets, approximation, and linear programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000. 1, 1.2, 1.2, 2.1, 3.2, 5, G, G

[Tul09] Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pages 303–312, 2009. 8

[Wol07] Paweł Wolff. Hypercontractivity of simple random variables. *Studia Mathematica*, 180(3):219–236, 2007. A.5

[WZ10] Yi Wu and Yuan Zhou. Personal communication, 2010. 2

[Yan04] Ke Yang. On the (im)possibility of non-interactive correlation distillation. In *Proceedings of the 6th Annual Latin American Informatics Symposium*, pages 222–231, 2004. 3, 3.1

[Yan07] Ke Yang. On the (im)possibility of non-interactive correlation distillation. *Theoretical Computer Science*, 382(2):157–166, 2007. 3, 5, 5.1, 5.1