THEORY OF COMPUTING, Volume 9 (23), 2013, pp. 703–757
www.theoryofcomputing.org

SPECIAL ISSUE: APPROX-RANDOM 2012

Circumventing d-to-1 for Approximation Resistance of Satisfiable Predicates Strictly Containing Parity of Width at Least Four*

Cenny Wenner†

Received October 31, 2012; Revised June 8, 2013; Published September 13, 2013

Abstract: Håstad established that any predicate $P \subseteq \{0,1\}^m$ containing Parity of width at least three is approximation resistant for almost-satisfiable instances. In comparison to for example the approximation hardness of 3SAT, this general result however left open the hardness of perfectly-satisfiable instances. This limitation was addressed by O'Donnell and Wu, and subsequently generalized by Huang, to show the threshold result that predicates *strictly* containing Parity of width at least three are approximation resistant also for perfectly-satisfiable instances, assuming the d-to-1 Conjecture.

We extend modern hardness-of-approximation techniques by Mossel et al., eliminating the dependency on projection degrees for a special case of decoupling/invariance and—when

*A preliminary version of this paper appeared in the International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'12) [28], and a technical preprint appeared in the Electronic Colloquium on Computational Complexity (ECCC'12), 2012 [27].

†Supported by ERC Advanced Investigator Grant 226203.

ACM Classification: G.1.2., G.1.6, G.3

AMS Classification: 68Q17, 68Q25, 68Q87, 68W25, 90C59

Key words and phrases: Boolean functions, hardness of approximation, inapproximability, constraint satisfaction, Fourier analysis, approximation resistance, dictatorship test, label cover, smooth label cover, correlation bounds, perfect completeness, invariance, d-to-1 games

 DOI: 10.4086/toc.2013.v009a023

reducing from SMOOTH LABEL COVER—the dependency on projection degrees for noise introduction. Tools in hand, we prove the same approximation-resistance result for predicates of width at least four, subject only to $\mathsf{P} \neq \mathsf{NP}$.

1 Introduction

We study the approximation limits of NP-hard Constraint Satisfaction Problems (CSPs). A canonical example being MAX-3SAT which in the CSP framework can be denoted as MAX-CSP$^+$(3OR).[1] In MAX-3SAT, we are given Boolean variables $x_1, \ldots, x_n$ and clauses of the form "$a \vee b \vee c$," where each literal a, b, and c is either a variable x_i or its negation; a solution to an instance is an assignment to the variables, the (optimal) value of a solution is the number of clauses it satisfies, and the value of an instance is the maximum value over all solutions. In the CSP framework, we substitute the value "true" for 1 and "false" for 0. In greater generality, the Boolean problem MAX-CSP$^+(P)$ is defined by specifying the *width-m* Boolean predicate $P \subseteq \{0,1\}^m$ applied to the set of m literals instead of 3OR.

It is known that 3SAT is NP-hard to solve exactly and we turn our attention to efficient approximations. We say that a solution is a c-approximation if its value is at least c times the optimal value of an instance. In particular, for MAX-3SAT, choosing a random assignment yields a $7/8$-approximation in expectation and unfortunately this is essentially the best efficient approximation of the problem as MAX-3SAT is NP-hard to approximate better than $7/8 + \varepsilon$ for every $\varepsilon > 0$ [11]. In fact, even if the instance is perfectly satisfiable, i. e., positive instances can have all clauses satisfied, it is NP-hard to satisfy more than a fraction $7/8 + \varepsilon$.

When a random assignment to an almost-satisfiable instance essentially achieves the best polynomial-time approximation factor assuming $\mathsf{P} \neq \mathsf{NP}$, we say that a predicate is *approximation resistant*. For simplicity, our treatise work under the assumption $\mathsf{P} \neq \mathsf{NP}$. A benefit of showing that a predicate is approximation resistant is that it establishes the optimal polynomial-time approximation factor of the predicate up to lower-order terms. In particular, this quantity is called the *random assignment threshold* and equals $2^{-m}|P|$ where m is the width of the predicate P and $|P|$ is the number of assignments $\mathbf{x} \in \{0,1\}^m$ which satisfies the predicate. The celebrated work by Håstad [11] demonstrated that a number of well-studied predicates are approximation resistant and the techniques thereof have been the starting point of a long line of strong inapproximability results. Most proofs showing the approximation resistance of a predicate P in fact establishes something stronger: that the predicate is *hereditarily approximation resistant* or even *useless*, either property implying that every predicate $Q \supseteq P$ is also approximation resistant. In fact, recent developments show that there are hereditarily approximation-resistant predicates accepting very few assignments and in particular these predicates are known to be contained in "most" predicates, formally implying that the fraction of approximation-resistant predicates of a particular width m approaches one as m goes to infinity [12].

Of particular interest to us is the predicate *Odd Parity* defined by $(a_1, \ldots, a_m) \in P$ if the number of $a_i = 1$ is odd, and the predicate *Even Parity* is defined analogously. Håstad showed that (either) Parity is *hereditarily approximation resistant*, meaning that not only is Parity approximation resistant, but so is

[1] The definition of MAX-CSP is at times ambiguous and we have chosen to add a superscripted plus to signify that constraints may involve negations of variables or more general operations for larger domains. As an example, if NEQ denotes the binary Boolean not-equal predicate, then MAX-CSP(NEQ) equals MAX-CUT while MAX-CSP$^+$(NEQ) equals MAX-3LIN-2.

any predicate $Q \subseteq \{0,1\}^m$ containing Parity, whereby containing, we mean in the set sense. However, in comparison to for instance MAX-3SAT, this result only holds with respect to *almost satisfiable* instances. Formally, letting Q be an arbitrary predicate containing Parity, for any $\eta, \varepsilon > 0$, given a MAX-CSP$^+(Q)$ instance with value at least $1-\eta$, it is NP-hard to find a solution with value at least $2^{-m}|Q| + \varepsilon$.

For Parity, the use of almost-satisfiable instances is necessary: perfectly-satisfiable instances can via Gaussian elimination be solved in polynomial time, whereas almost-satisfiable instances are hard to approximate within $1/2+\varepsilon$. It is not immediately clear whether other approximation-resistant predicates containing Parity should be easy or hard for satisfiable instances, and indeed 3SAT is as hard to approximate for almost-satisfiable as it is for perfectly-satisfiable instances.

Assuming Khot's d-to-1 Conjecture [17], this question was settled by O'Donnell and Wu [22] for $m = 3$ and later generalized to $m \geq 3$ by Huang [14]. They showed the remarkable threshold result that any predicate *strictly* containing Parity is approximation resistant also for perfectly-satisfiable instances. More specifically, the width-three case can on account of symmetry be reduced to showing the hereditary approximation resistance of the "Not Two" predicate (NTW), defined as accepting all triples of bits if they are all zeroes or have odd parity, i. e., does not contain two ones.

The result of O'Donnell and Wu follows from the construction of a Probabilistically Checkable Proof (PCP) reducing from an outer verifier to MAX-CSP$^+(NTW)$. The outer verifier may be taken as a black-box (parametrized) CSP called LABEL COVER. In LABEL COVER, one is given a bipartite graph $G = (U \cup V, E)$, a "small" label set K, a "large" label set L, and for each edge $e \in E$ an associated projection $\pi_e : L \to K$. Solutions assign each vertex $u \in U$ a label $\lambda(u)$ from K and each vertex $v \in V$ a label $\lambda(v)$ from L, and the value of a solution is the fraction of edges $\{u,v\} \in E$ for which $\lambda(u) = \pi_{\{u,v\}}(\lambda(v))$. One can show that it is NP-hard for every $\varepsilon_{\mathrm{LC}} > 0$ to distinguish whether a LABEL COVER instance has value 1 (the *completeness*) or value at most $\varepsilon_{\mathrm{LC}}$ (the *soundness*) for sufficiently large label sets K and L depending on $\varepsilon_{\mathrm{LC}}$.

Reductions from LABEL COVER are today standard in hardness of approximation. For Boolean constraints, such proofs typically involve semantically replacing $\lambda(u)$ and $\lambda(v)$ with $2^{2^{|K|}}$ and $2^{2^{|L|}}$ Boolean variables, respectively. These variables are respectively viewed as functions $f^u : \{-1,1\}^K \to \{-1,1\}$ and $g^v : \{-1,1\}^L \to \{-1,1\}$. The intention, for positive instances, is to set these functions to *dictators*. That is, setting $f^u(\mathbf{x}) = x_{\lambda(u)}$ and $g^v(\mathbf{y}) = y_{\lambda(v)}$. For negative instances, there are however no guarantees that the functions are set according to this coding scheme. Reducing to a MAX-CSP$^+(P)$ instance, and viewing P as the indicator of its set, points of such functions are passed as arguments to P. The value of an edge $\{u,v\}$ in the LABEL COVER instance is thereby reduced to, for some integer T, to the value of the expectation

$$\mathbf{E}_{(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(T)},\mathbf{y}^{(T+1)},\ldots,\mathbf{y}^{(m)})\sim\mathcal{T}}\left[P\left(f^u(\mathbf{x}^{(1)}),\ldots,f^u(\mathbf{x}^{(T)}),g^v(\mathbf{y}^{(T+1)}),\ldots,g^v(\mathbf{y}^{(m)})\right)\right], \tag{1.1}$$

where the arguments are chosen according to a *test distribution* $\mathcal{T}$. Equation (1.1) is the starting point for Fourier analysis of PCPs. For approximation resistance, this involves first taking the Fourier expansion of P and proceeding to bounds terms of the forms $\mathbf{E}[\prod f^u]$, $\mathbf{E}[\prod g^v]$, and/or $\mathbf{E}[\prod f^u \prod g^v]$. For work most similar to this treatise, T is typically one, rendering the first kind of term(s) trivial to bound while terms of the third kind become $\mathbf{E}[f^u \prod g^v]$. Finally, a central parameter to this work is the *(maximum) degree of projections*, $d = d(\varepsilon_{\mathrm{LC}}) = \max_{e\in E}\max_{i\in K}|\pi_e^{-1}(i)|$. That is, the greatest number of labels from the large label set which share projections. For present NP-hard constructions of LABEL COVER, d grows polynomially with $\varepsilon_{\mathrm{LC}}^{-1}$ and one would like to avoid bounds depending on both $\varepsilon_{\mathrm{LC}}$ and d.

The construction by O'Donnell and Wu is similar to that of Håstad for MAX-3-LIN-2, i. e., Even Parity on three bits. Working with almost-satisfiable instances, Håstad could define his test distribution such that each argument to a function was somewhat "noised." O'Donnell and Wu, working with perfectly-satisfiable instances, could not afford this. Instead they made use of the subtle "unpredictability" of a predicate which strictly contains Parity. Defining a test distribution close to that for MAX-3-LIN-2, but with somewhat bounded correlation between the arguments to the functions, they used theorems by Mossel [19] to argue that the analysis behave roughly as though the arguments were somewhat noised. Following this, the effect of only being "close" to the uniform distribution over Parity had to be bounded. For this, they extended modern techniques for analyzing PCP's. They introduced a "matrix-notation technique" to bound terms of the form $\mathbf{E}[\prod g^v]$ while for terms of the form $\mathbf{E}[f^u \prod g^v]$, they used a coordinate-wise distribution-substitution method, also known as "Lindeberg's method," to bound the terms by influences. Their method has subsequently found other applications [26, 25].

We note that all of the steps in the preceding paragraph involve degenerative dependencies on d, the degrees of projections. This prompted the use of the d-to-1 Conjecture which states that LABEL COVER remains NP-hard for arbitrarily low soundness error $\varepsilon_{\mathrm{LC}}$ even when the degrees of projections is fixed to a constant d. The d-to-1 Conjecture is a sibling to the more well-known Unique Games Conjecture (UGC) which hypothesizes that LABEL COVER is hard even for unique projections, i. e., projection degree $d = 1$, albeit for distinguishing instances of value $1 - \varepsilon_{\mathrm{LC}}$ from instances of value $\varepsilon_{\mathrm{LC}}$ as satisfiable instances can be recognized in polynomial time.

The two conjectures, and in particular the UGC, imply strong or even tight results for a remarkably large class of NP-hard problems. For instance, that Parity is hereditarily approximation resistant is equivalently to the statement "*any width-$m \geq 3$ predicate supporting an $(m-1)$-wise-independent distribution is approximation resistant*." Assuming the UGC, Austrin and Mossel [2] showed that in fact it suffices for approximation resistance that a predicate supports a pairwise-independent distribution.

Despite considerable efforts to either prove or refute these conjectures, we appear to be nowhere near settling the conjectures nor theorems serving equivalent purposes. There has however been recent progress towards circumventing the conjectures for particular problems [25, 10] which is also the strategy of this treatise. In fact, recently, the UGC-based result of Austrin and Mossel was essentially circumvented by Chan [5], establishing NP-hard approximation resistance of every predicate containing a subgroup supporting a pairwise-independent distribution. This result carried numerous implications such as an almost tight hardness of general width-constrained constraint satisfaction, or, in other words, "*how few satisfying assignments can an approximation-resistant predicate have?*," improving the bound from $2^{\mathcal{O}(m^{1/2})}$ of the 2^m assignments to simply $2m$. It is worth noting that Chan's result and implications again hold with regard to distinguishing almost-satisfiable instances. In follow-up by Huang [15], the techniques of this work as well as of Chan's were extended to show, with regard to perfectly satisfiable instances, the approximation resistance of a predicate with $2^{\mathcal{O}(m^{1/3})}$ accepting assignments.

Related work on the width-three case We note that our techniques only permit us to show the approximation resistance of predicates strictly containing Parity of width at least four while we fail to address the arguably most interesting case of width three. As mentioned above, due to symmetry, the width-three case reduces to showing the approximation resistance for satisfiable instances of the predicate NTW $= \{000, 001, 010, 100, 111\}$.

In a parallel result, Håstad [13], handles this case with similar but slightly different methods from ours. In fact, Håstad shows that with slight modifications to the outer verifier and choice of parameters, the original protocol by O'Donnell and Wu in fact yields approximation resistance of NTW even without assuming the d-to-1 Conjecture. In both our approach as well as in Håstad's, reducing from SMOOTH LABEL COVER is integral to limit the effect of projection degrees in correlations. For the definition of terms used in this section, we refer the reader to Section 1.2. Instead of sacrificing completeness for noise in the protocol, one defines in addition to the main distribution over arguments, a noising distribution on the predicate which is able to break perfect correlations between arguments. In our methods, to get correlations independent of projection degrees, we employ this distribution independently for different labels sharing projections. Consequently, for our method to work, the noising distribution needs to have uniform marginals conditioned on the small-side argument or else the projected marginals will not be uniform and give away the tested LABEL COVER edge. For the NTW predicate $\{000, 001, 010, 100, 111\}$, conditioned on the first argument being 0, the choices of the remaining two arguments are 00, 01, and 10. Unfortunately, placing any weight on the 000 outcome cannot produce a uniform marginal with projections, while putting weight on the other outcomes yields perfect correlation between the three arguments, and therefore our approach falls short.

O'Donnell and Wu, as well as Håstad, retain uniform marginals while putting weight on the outcome 000 by instead for each small-side label i using the noising distribution at most once among the set of all large-side labels with projection i and using the outcome 111 to counteract the bias. Naturally, a consequence of this limited use of the noising distribution is that the correlations between arguments will depend on the projection degrees. Håstad counteracts this dependency by using, in the analysis, sufficiently large smoothness parameters so that, e. g., "large" Fourier terms which should be killed by noise depend on sufficiently many labels as a function of the projection degrees that the dependent correlations suffice. Normally, such a dependency in smoothness parameters is problematic because it instead becomes difficult to bound "small" terms. Håstad handles this by dividing terms into "small," "medium," and "large" components where the difference in sizes between small and large terms is sufficiently extreme that both parts can be suitably handled and what remains is to bound "medium"-sized terms. The sizes in question depend on the probability that one samples from the noising distribution and Håstad handles middle terms with a trick from the protocol of MAX-3SAT: randomizing this noising parameter and in effect the size of small and large terms. Suitably chosen, terms are unlikely to fall in the medium-size component and hence the contribution of this component is insignificant in expectation. The downside with this approach is that it introduces a massive, "tower of exponentials," blowup in the final smoothness parameters and resulting instance size, although this is still merely a constant.

1.1 Our contributions and techniques

Our main contribution is to circumvent the d-to-1 Conjecture to show that any predicate strictly containing Parity of width at least four is approximation resistant for satisfiable instances unless $\mathsf{P} = \mathsf{NP}$. The overarching steps of our proof follow those of O'Donnell and Wu, and our main technical contribution is to extend the methods of Mossel et al. [20, 19] to limit the effects of projection degrees. Subject to smoothness, explained below, we show that our PCP behaves roughly the same subject to what we call projected noise as it does subject to independent noise; more on this below. Additionally, we employ a multivariate invariance principle extended to projection games which avoids dependencies on the degree of projections d. We note that a similar elimination of the dependency on d, using different methods, was recently shown by O'Donnell and Wright [21] for a particular two-variable case employed to show that it is NP-hard to distinguish $1/2$-satisfiable UNIQUE GAMES instances from $3/8+\varepsilon$ satisfiable.

The SMOOTH LABEL COVER problem serves an integral role in our proofs and is a variant of LABEL COVER which roughly states that if one looks at a vertex $v \in V$ and two labels $j \neq j' \in L$, over the random choice e of edges incident v, the two labels are unlikely to share projection; that is, the event "$\pi_e(j) = \pi_e(j')$" has arbitrarily low positive measure over the choice of $e \in E$. SMOOTH LABEL COVER was first defined by Khot to show approximation hardness of COLORING [16]. Subsequently, Feldman et al. [9] used it for the hardness of learning monomials, and Guruswami et al. [10] to establish exciting optimal inapproximability results for two geometric results where previously only optimal UG-hardness results were known. More intimately related to our work, Khot and Saket [18] used smoothness to show $20/27+\varepsilon$ approximation hardness of MAX-CSP of width three on satisfiable instances.

Subject to smoothness, we relate what we call *projected noise* to *non-projected* or *independent noise*. By projected noise, we mean noise where coordinates which share projections are jointly resampled with some small noise probability while the latter does the same independently for each coordinate. Projected noise is introduced by conventional techniques from correlation bounds, while independent noise is typically needed to decode from influences without a dependency on projection degrees. The issue with the former is that projected noise does not significantly affect functions which depend on a large number of coordinates with the same projection. However, under SMOOTH LABEL COVER, any function which depends on many coordinates must essentially depend in expectation on many coordinates with different projections. With the limited unpredictability of the distribution we define, we can via correlation bounds introduce projected noise independent of d and subsequently turn it into independent noise because of smoothness.

With a test distribution which behaves roughly as though arguments were independently noised, we wish to bound expectations of the form $\mathbf{E}[\prod g^v]$ and $\mathbf{E}[f^u \prod g^v]$. For the former, we employ smoothness, partial independence of the test distribution, and hypercontractivity to argue that the expectation is roughly the same as for a distribution where all coordinates $j \in L$ are drawn independently, as in UNIQUE GAMES. Since our test distribution is arbitrarily close to being independent over the arguments $\{\mathbf{y}^{(t)}\}_t$ in this setting, the expectation $\mathbf{E}[\prod g^v]$ is close to 0. Finally, we extend the coordinate-wise distribution-substitution method of O'Donnell and Wu, to show a multivariate invariance theorem similar to Mossel's [19] but where bounds do not depend on the degree of projections d. This permits us to effortlessly bound terms of the form $\mathbf{E}[f^u \prod g^v]$. In fact, the soundness analysis of a term $\mathbf{E}[f^u \prod g]$ involving functions on both the small and large label sets—often considered the hardest part of soundness analysis—ecomes the easiest step subject to this theorem.

It may be pedagogical to discuss what we require to employ our steps. For noise introduction, it suffices, with smoothness, that each string $\mathbf{y}^{(r)}$ has in the marginal distribution over a label $j \in L$ bounded correlation to arguments $\{\mathbf{y}^{(t)}\}_{t \neq r}$ conditioned on $\{\mathbf{x}^{(t)}\}_t$. For bounding products of the form $\mathbf{E}[\prod g^v]$, we require noise, smoothness, and a roughly $m/2$-wise independent balanced distribution for $\{\mathbf{y}^{(t)}\}_t$. For bounding products of the form $\mathbf{E}[f^u \prod g^v]$ in terms of influences, we require the weak conditions of uniform marginals and that any single string $\mathbf{y}^{(r)}$ is independent of $\{\mathbf{x}^{(t)}\}_t$.

1.2 Preliminaries

We assume that the reader is familiar with basic probability theory and computational complexity theory.

1.2.1 Basic notation

When clear from the context sub- and superscripts may be omitted. For any real p, we denote by $\bar{p} = 1 - p$, while for a set A from a possibly implicit universe $\mathcal{U}$, $\bar{\mathrm{A}}$ refers to the complementary set $\mathcal{U} \setminus \mathrm{A}$. We use Iverson notation $[S]$ where S is a true/false statement to denote 1 whenever S is true and 0 otherwise. For a natural number n, the integral interval $\{1, \ldots, n\}$ is denoted $[n]$. In this treatise, we deal extensively with correlated spaces $\mathcal{P} = (\prod_{t=1}^{m} \Omega_t, \mu)$ over finite domains. When the sample space is clear from the context, we may also specify measures instead of probability spaces, and vice versa. Given an index set $A \subseteq [m]$, we call Ω_A the product space $\prod_{t \in A} \Omega_t$. For a single index $r \in [m]$, Ω_{-r} denotes the product of all sample spaces besides Ω_r, i. e., $\prod_{t \in A \setminus r} \Omega_t$, where $A \setminus r$ denotes the set $A \setminus \{r\}$. Similarly, $\Omega_{-r,-r'} \triangleq \Omega_{A \setminus \{r,r'\}}$ (where $\triangleq$ denotes equality by definition). On a related note, for a set S and element x, we may denote the difference $S \setminus \{x\}$ simply by $S \setminus x$ or $S - x$. Vectors may for clarity be denoted either by bold font, as in $\mathbf{x}$, or with an overset arrow, as in $\vec{\mu}$. Given a tuple $\mathbf{x} = (x_i)_{i \in A}$ and a bijection $\sigma : A \leftrightarrow A$, $\mathbf{x} \circ \sigma$ denotes the tuple $(x_{\sigma(i)})_{i \in A}$. For functions $\pi : A \to B$, where A and B are arbitrary domains, we may also see π as a relation, i. e., the set of tuples $\{(a, b) \in A \times B \mid \pi(a) = b\}$. The ℓ_p norm of f is denoted by $\|f\|_{\mu,p}$ and is defined as

$$\mathbf{E}_{\mathbf{x} \sim \mu}[|f(\mathbf{x})|^p]^{1/p}$$

for real $p \geq 1$ and $\max_{\mathbf{x}} f(\mathbf{x})$ for $p = \infty$. When clear from the context, we shall omit the distribution μ.

1.2.2 Operators on probability spaces

Tensoring Given a probability space $\mathcal{P} = (\Omega, \mu)$, the n^{th} tensor power of $\mathcal{P}$ is $\mathcal{P}^{\otimes n} = (\Omega^n, \mu' = \mu^{\otimes n})$ where $\mu'(\omega_1, \ldots, \omega_n) = \mu(\omega_1) \cdots \mu(\omega_n)$.

Noise operators So called noised functions are standard when analyzing PCPs and we extend the notion somewhat to encompass also probability spaces.

Definition 1.1. Let $\mathcal{P} = (\Omega_1, \mu)$ be a probability space, n an natural number, and $f : \Omega^n \to \mathbb{R}$ a function on $\mathcal{P}^{\otimes n}$. The *noise operator*, also called the *Bonami-Beckner operator*, $\mathrm{T}_{\mathcal{P},\bar{\gamma}}(f) : \Omega^n \to \Omega^n$ with parameter $\bar{\gamma} \in [0, 1]$ applied to f takes an argument $\mathbf{x} = (x_i)_{i \in [n]}$, and yields the expectation of f where for every i, x_i is independently resampled from $\mathcal{P}$ with probability γ and otherwise preserved.

We shall typically omit the distribution $\mathcal{P}$ when it is clear from the context. The noise operator is more commonly defined by a parameter specifying the noise, whereas we specify the *correlation*, a more natural quantity in our eyes. The relation between the two definitions is immediate, substituting $\bar{\gamma}$ for γ.

It is convenient for our proofs to extend the definition of noise operators to probability spaces. In particular, let $\mathcal{P} = (\prod^m \Omega_t, \mu)$ be a correlated probability space, $A \subseteq [m]$ an index set, and $\bar{\gamma}$ a parameter. Then, $\mathrm{T}^A_{\bar{\gamma}} \mathcal{P}$ is defined as the probability space which first draws from $\mathcal{P}$ and with probability γ resamples Ω_A from its marginal of μ. When A is a singleton $\{x\}$, we merely denote the noise operator by $\mathrm{T}^x_{\bar{\gamma}}$ rather than $\mathrm{T}_{\bar{\gamma}}^{\{x\}}$. Abusing notation, we also use the shorthand $\mathrm{T}_{\bar{\gamma}_1,\dots,\bar{\gamma}_k}^{i_1,\dots,i_k}$ for $\mathrm{T}_{\bar{\gamma}_1}^{i_1}\cdots\mathrm{T}_{\bar{\gamma}_k}^{i_k}\,\mathcal{P}$. We note that the order of application of noise operators acting on individual sample spaces does not matter. Furthermore, noise operators do not affect the marginal distribution of any sample space.

The projection operator In order to conveniently analyze projection-game-based PCPs, we introduce a *projection operator* on correlated spaces. Intuitively, the operator yields a correlated space which first samples a subset of spaces Ω_A and then a number of times independently samples the remaining spaces $\Omega_{\bar{A}}$ conditioned on Ω_A. In other words, the projection operator is a formal notation for a typical PCP construction: one samples arguments for a few functions, here indexed by A, followed by "projecting," or sampling independently d times, the arguments for the remaining functions conditioned on the former.

Definition 1.2. The *degree-d projection*, $d \geq 1$, from an index set $A \subseteq [m]$ on a correlated space $\mathcal{P} = (\prod^m \Omega_t, \mu)$ is defined as $\mathcal{P}^{d\text{-proj-}A} \triangleq (\prod^m \Omega'_t, \mu')$, where $\Omega'_t = \Omega_t$ if $t \in A$ and otherwise Ω_t^d; $\omega'_t = (\omega'_{t,i})_{i\in[d]} \in \Omega'_t$ for $t \notin A$; and

$$\mu'(\omega'_1,\dots,\omega'_m) = \mathbf{P}_\mu\big(\Omega_A = \vec{\omega}'_A\big)\prod_{i=1}^{d}\mathbf{P}_\mu\big(\forall_{t\notin A}\Omega_t = \omega'_{t,i}\,|\,\Omega_A = \vec{\omega}'_A\big)\,.$$

For notational simplicity, we use the shorthand "$\mathcal{P}^{d\text{-proj-}1}$" to denote "$\mathcal{P}^{d\text{-proj-}\{1\}}$."

1.2.3 Orthogonal decompositions

Following previous work, our proofs make ample use of the *Fourier* and *Efron-Stein decompositions*. For a finite probability space (Ω,μ), let $\{\chi_\omega : \Omega \to R\}_{\omega\in\Omega}$ be an orthonormal basis of $L^2(\Omega,\mu)$ such that $\chi_{\omega'} = 1$ for some $\omega' \in \Omega$. Defining for $\vec{\omega} \in \Omega^n$ the character $\chi_{\vec{\omega}}(\mathbf{x}) = \prod_{i\in S}\chi_{\omega_i}(x_i)$, the collection $\{\chi_{\vec{\omega}}\}_{\vec{\omega}\in\Omega^n}$ is an orthonormal basis of $L^2(\Omega^n,\mu^n)$. In particular, if $\Omega = \mathbb{F}_q$ for some q and the distribution μ is uniform, one can take $\chi_\omega(x) = \mathrm{e}^{-\mathrm{i}2\pi x\omega/q}$ and for $\{\chi_\sigma\}_{\sigma\in\mathbb{F}_q^n}$the Fourier characters (Ω^n,μ^n). For the Boolean case which we are interested in, i. e., $\Omega = \{-1,1\}$, we shall simply index the Fourier characters by subsets $S \subseteq [n]$ and defining $\chi_S(\mathbf{x}) = \prod_{i\notin S}\chi_0(x_i)\prod_{i\in S}\chi_1(x_1) \triangleq \prod_{i\in S} x_i$. A Fourier basis of $L^2(\Omega^n,\mu^n)$ decomposes a function $f : \Omega^n \to \mathbb{R}$ into coefficients $\{\hat{f}_{\vec{\omega}}\}_{\vec{\omega}}$ of orthonormal characters $\{\chi_{\vec{\omega}}\}_{\vec{\omega}}$ which is suitable for PCP analysis. In particular, for the Boolean domain, $f(\mathbf{x}) = \sum_{S\subseteq[n]}\hat{f}_S\chi_S(\mathbf{x})$ where $\hat{f}_S = \mathbf{E}_\mathbf{x}[f(\mathbf{x})\chi_S(\mathbf{x})]$ and consequently, $\big|\hat{f}_S\big| \leq \max f$.

Slightly simpler for this work in the non-Boolean domain is however the Efron-Stein decomposition. To define this decomposition, we introduce "restriction means" of a function to a subset of arguments relative a measure. We note that these functions have in other works been referred to as *projections on coordinates*, in the linear-algebra sense rather than in the sense of LABEL COVER constraints.

Definition 1.3. Let $f : \Omega^{(1)} \times \cdots \times \Omega^{(n)} \to \mathbb{R}$, μ a measure on $\prod \Omega^{(t)}$, and $S \subseteq [n]$. Then the *restriction mean* of f to the argument set S is defined as $f_{\subseteq S}(\mathbf{x}) = \mathbf{E}[f(\mathbf{X})\,|\,\mathbf{X}_S = \mathbf{x}_S]$.

With this definition in hand, we define the Efron-Stein decomposition as follows.

Definition 1.4. Let $f : \Omega^{(1)} \times \cdots \times \Omega^{(n)} \to \mathbb{R}$ and μ a measure on $\prod \Omega^{(t)}$. Then the *Efron-Stein decomposition* of f with respect to μ is $\{f_S\}_{S\subseteq[n]}$ where

$$f_S(\mathbf{x}) = \sum_{T\subseteq S} (-1)^{|S\setminus T|} f_{\subseteq T}(\mathbf{x}) \,. \tag{1.2}$$

Whenever the sample spaces $\{\Omega^{(t)}\}_t$ are independent, the decomposition satisfies the following properties. We note that these properties were in previous work [7, 19] taken as the definition of an Efron-Stein decomposition, for which (1.2) was the unique construction.

Lemma 1.5 (Efron and Stein [7] and Mossel [19])**.** *Assuming $\{\Omega^{(t)}\}_t$ are independent, the Efron-Stein decomposition $\{f_S\}_S$ of $f : \prod \Omega^{(t)} \to \mathbb{R}$ satisfies:*

- *$f_S(\mathbf{x})$ depends only on $\mathbf{x}_S$,*
- *and for any $S, T \subseteq [m]$, and $\mathbf{x}_T \in \prod_{t\in T} \Omega^{(t)}$ such that $S \setminus T \neq \emptyset$,*

$$\mathbf{E}[f_S(\mathbf{X}) \,|\, \mathbf{X}_T = \mathbf{x}_T] = 0 \,.$$

We note that the relation between Efron-Stein and Fourier decompositions is straightforward, setting $f_S(\mathbf{x}) = \sum\{\hat{f}_S \chi_S(\mathbf{x}) : \bar{\omega} \in \Omega^n, S = \{i \mid \omega_i \neq \omega'\}\}$. In the Boolean case, and in effect for our approximation-hardness application, albeit not for our general techniques, the transforms are interchangeable where in particular $f_S(\mathbf{x}) = \hat{f}_S \chi_S(\mathbf{x})$.

The effect of the noise operator $\mathrm{T}_{\bar{\gamma}}$ additionally has a simply characterization in terms of the decompositions. In particular, for a function f with Efron-Stein decomposition $\{f_S\}_S$, the Efron-Stein decomposition $\{f'_S\}_S$ of $f' = \mathrm{T}_{\bar{\gamma}} f$ satisfies $f'_S = \bar{\gamma}^{|S|} f_S$.

A noised function roughly behaves as a degree-bounded function as follows.

Definition 1.6. The *k-low-degree expansion* $f^{\leq k}$ of a function $f : \Omega^n \to \mathbb{R}$ with Efron-Stein decomposition $\{f_S\}_{S\subseteq[n]}$ is defined as

$$f^{\leq k}(\mathbf{x}) = \sum_{S\subseteq[n]\,:\,|S|\leq k} f_S(\mathbf{x}) \,.$$

Similarly,

$$f^{> k}(\mathbf{x}) = \sum_{S\subseteq[n]\,:\,|S|> k} f_S(\mathbf{x}) \,.$$

We also note the following well-known corollary of the Hypercontractivity Theorem [4, 3].

Lemma 1.7. *Let $q \geq 2$ be a parameter and $f : \{-1,1\}^n \to \mathbb{R}$ a function of degree at most k. Then,*

$$\|f\|_q \leq (q-1)^{k/2} \|f\|_2 \,.$$

The term *shattered* was coined by Håstad [13] and denotes an expansion $\sum_{S\in\mathbb{S}} g_S$ where every non-zero g_S satisfies $|\pi(S)| = |S|$. One of our goals is to show that for smooth projections, low-degree functions are essentially shattered.

Definition 1.8. Let L, K, and Ω be arbitrary finite sets and consider a projection $\pi : L \to K$ and a function $g : \Omega^L \to \mathbb{R}$. We say that g is *shattered* when non-zero Efron-Stein terms $g_T, T \subseteq L$, satisfy $|\pi(T)| = |T|$, i. e., the elements of T have unique projections.

Definition 1.9. With respect to finite sets L, K, and Ω; a projection $\pi : L \to K$; and a function $g : \Omega^L \to \mathbb{R}$, we define the *shattered part of g with respect to π*, $g^{\rightarrowtail\pi}$, as

$$g^{\rightarrowtail\pi}(\mathbf{y}) = \sum_{T \subseteq L : |\pi(T)| = |T|} g_T(\mathbf{y}) .$$

In the following lemma, we show that the sum of all Efron-Stein terms f_S in a powerset $S \subseteq B$ behaves nicely. More specifically, the following lemma implies that $\left|\sum_{S:A\subseteq S\subseteq B} f_S\right| \leq 2^{|A|}\|f\|_\infty$. In our applications, we apply this lemma with constant-sized A.

Lemma 1.10. *Let $A \subseteq B \subseteq [n]$ and $\{f_S\}_{S\subseteq[n]}$ be the Efron-Stein decomposition of a function $f : \Omega^n \to \mathbb{R}$. Then,*

$$\sum_{S:A\subseteq S\subseteq B} f_S = \sum_{B\setminus A\subseteq S\subseteq B} (-1)^{|B\setminus S|} f_{\subseteq S} .$$

Proof. By construction, the LHS equals

$$\begin{aligned}
\sum_{S:A\subseteq S\subseteq B} \sum_{T\subseteq S} (-1)^{|S-T|} f_{\subseteq T} &= \sum_{T\subseteq B} (-1)^{|T|} f_{\subseteq T} \sum_{S:A\cup T\subseteq S\subseteq B} (-1)^{|S|} \\
&= \sum_{T\subseteq B} (-1)^{|T|} f_{\subseteq T} (-1)^{|A\cup T|} \sum_{U\subseteq B\setminus(A\cup T)} (-1)^{|U|} . \qquad (1.3)
\end{aligned}$$

Whenever $B \setminus (A \cup T) \neq \emptyset$, the last expression evaluates to 0. Consequently, non-zero terms satisfy $A \cup T = B$ and, as desired, $(1.3) = \sum_{B\setminus A\subseteq T\subseteq B} (-1)^{|B\setminus T|} f_{\subseteq T}$. □

1.2.4 Influences

A useful concept of functions is the *influence* of a coordinate. Intuitively, for a function $f : \{-1,1\}^n \to \mathbb{R}$, the influence of coordinate i is how much $f(\mathbf{x})$ changes on average with x_i. When analyzing positive instances in long-code-based PCPs, the functions in question are *dictators* of the encoded assignments; formally, $f^u(\mathbf{x}) = x_{\lambda(u)}$ where $\lambda(u)$ is the assignment to the vertex u in the reduced-from LABEL COVER instance. In the other direction, whenever a protocol accepts with a non-negligible probability over a random assignment, one would like to argue that the functions must essentially have significant influences and additionally so, for multiple functions, of coordinates consistent with projections.

Definition 1.11. Let $f : \Omega^n \to \mathbb{R}$ be a function and $i \in [n]$ a coordinate. The *influence* of coordinate i is $\mathrm{Inf}_i(f) = \mathbf{E}_{\mathbf{x}_{-i}}[\mathbf{Var}_{x_i}[f(\mathbf{x})]]$, where the implicit distributions are uniform over Ω^n.

The influence of a coordinate has a nice representation in terms of the Efron-Stein decomposition with respect to the uniform distribution.

Lemma 1.12. *Let $f : \Omega^n \to \mathbb{R}$ be a function and $\{f_S\}_S$ its Efron-Stein decomposition with respect to the uniform distribution. Then,*

$$\mathrm{Inf}_i(f) = \sum_{S\ni i} \mathbf{E}\left[f_S^2\right] \leq \mathbf{Var}[f] .$$

In a similar way, *noisy influences* are defined as $\mathrm{Inf}_i^{(\bar{\gamma})}(f) \triangleq \mathrm{Inf}_i(\mathrm{T}_{\bar{\gamma}} f)$ where $\gamma \in [0,1]$ is a noise parameter. Consequently,

$$\mathrm{Inf}_i^{(\bar{\gamma})}(f) = \sum_{S \ni i} \rho^{2|S|}\, \mathbf{E}\left[f_S^2\right] .$$

We note that the total influence of a function with codomain $[-1,1]$ can be of the order n, achieved by for instance Parity on n bits, while the total noisy influence for $\gamma > 0$ is always bounded from above by a constant depending only on γ.

Lemma 1.13. *Let $f : \Omega^n \to [-1,1]$ be a function and γ a parameter in $(0,1]$. Then,*

$$\mathrm{TotInf}^{(\bar{\gamma})}(f) \triangleq \sum_i \mathrm{Inf}_i^{(\bar{\gamma})}(f) \le \gamma^{-1} . \tag{1.4}$$

Proof. Expressed in Efron-Stein terms, the total noisy influence equals

$$(1.4) = \sum_i \sum_{S \ni i} \bar{\gamma}^{2|S|}\, \mathbf{E}\left[f_S^2\right] = \sum_S |S| \bar{\gamma}^{2|S|}\, \mathbf{E}\left[f_S^2\right] . \tag{1.5}$$

Using Parseval's identity, $\sum_S \mathbf{E}\left[f_S^2\right] = \mathbf{E}[f]^2 \le 1$ for functions of codomain $[-1,1]$ and

$$(1.5) \le \max_S |S| \bar{\gamma}^{2|S|} \le \max_k \sum_{i=1}^{k} \bar{\gamma}^{2i} \le (1-\bar{\gamma}^2)^{-1} = (1-(1-\gamma)^2)^{-1} \le \gamma^{-1}(2-\gamma)^{-1} \le \gamma^{-1} . \qquad \square$$

1.2.5 Correlations

Intimately connected with noise operators is the concept of *correlation* between sample spaces. We note that correlations are always bounded by one and noise operators applied to individual sample spaces can only decrease correlation.

Definition 1.14. The *correlation* $\rho(\Omega_1, \Omega_2;\ \mu)$ between Ω_1 and Ω_2 with respect to the probability space $\mathcal{P} = (\Omega_1 \times \Omega_2, \mu)$ is

$$\rho_{\mathcal{P}}(\Omega_1, \Omega_2) \triangleq \rho(\Omega_1, \Omega_2;\ \mathcal{P}) \triangleq \max_{\phi, \psi} \mathbf{E}_\mu[\phi \psi] ,$$

where the maximum is over functions $\phi : \Omega_1 \to \mathbb{R}$, $\psi : \Omega_2 \to \mathbb{R}$ s.t. $\mathbf{E}[\phi] = 0$ and $\mathbf{Var}[\phi] = \mathbf{Var}[\psi] = 1$.

1.2.6 LABEL COVER variants

LABEL COVER is a constraint-satisfaction characterization of probabilistically checkable proofs which, for sufficiently large domains, has completeness 1 and soundness error arbitrarily close to 0. In particular, we in the following choose to reduce from the bipartite unweighted multigraph projection-game variant.

Definition 1.15. Instances (U, V, K, L, E, Π) of the gap problem LABEL COVER$_k$ consists of a bipartite multigraph $G = (U, V, E)$, label sets K and L, $|K| = k$, for the vertices U and V, respectively, and for every edge $e \in E$, a *projection* $\pi_e : L \to K$. A solution to LABEL COVER consists of a *labeling* $\lambda : U \to K, V \to L$ and the value of a solution is given by the fraction of edges $\{u, v\} \in E$ such that $\pi^{\{u,v\}}(\lambda(v)) = \lambda(u)$. Finally, the value of an instance is the maximum value of any solution.

We additionally refer to K as the small label set and to L as the large, similarly to U and V as the small- and large-side vertices, respectively.

A useful construction for limiting the effect of projection degrees is smoothness as introduced by Khot [16]. In Definition 1.16, we have adapted the characterizing definition of smoothness somewhat to the property typically exploited in PCP analysis. Intuitively, the smoothness property states that whenever we look at a constant-sized set of labels for a right-side vertex v, the labels have unique projections over the choice of neighbors of v.

Definition 1.16. A LABEL COVER instance is (J,ξ)-*smooth* if for any vertex $v \in V$ and any set of labels $S \subseteq L, |S| \leq J$, over a uniformly at random neighbor $u \in U$ of v, $\mathbf{P}_{u\sim v}\big(|\pi^{\{u,v\}}(S)| < |S|\big) \leq \xi$.

Subject to a LABEL COVER instance with the smoothness property, we also say that the a vertex $u \in U$ or projection is chosen (J,ξ)-smoothly.

The following definition and subsequent hardness proof of SMOOTH LABEL COVER closely follows standard constructions. We have adapted the definition insignificantly, choosing bipartite projection games over multi-layered games, characterizing Definition 1.16 as the essential property of smoothness, and for notational simplicity derive regular projection degrees.

Theorem 1.17. *For any parameters $\varepsilon_{LC} > 0, \xi > 0, J \in \mathbb{N}$, there exists $k = k(\varepsilon_{LC}, J, \xi)$ such that Gap-$(1,\varepsilon_{LC})$ LABEL COVER$_k$ with the following properties is* NP*-hard:*

- *the constraint graph is left- and right regular,*
- *projections are $d(\varepsilon_{LC}) = 5^{R_1(\varepsilon_{LC})}$-regular,*
- *and the instances are (J,ξ)-smooth.*

Proof. As in Khot [16], a consequence of Papadimitriou and Yannakakis [23], and more specifically Feige [8], there exists a constant $c < 1$ such that it is NP-hard to distinguish MAX-3SAT instances with exactly three literals per clause and five occurrences per variable of value 1 from instances of value at most c. As a projection game, or LABEL COVER instance, a random clause is sent to the second player and a random variable from the clause to the first player, accepting if the answers are consistent and satisfy the clause. The approximation hardness gap implies that distinguishing value-1 instances of this game from instances of value at most c' for some constant $c' < 1$ is NP-hard. One can extend the game by adding to each of the three weight-one clause assignments, such as "001," a duplicate label, e. g., "001*," with identical acceptance criteria. This ensures that for every assignment to the queried variable, there are exactly five satisfying clause assignments while the game remains hard for the same gap. Conventionally applying the Parallel Repetition Theorem of Raz [24], for some $R_1 = R_1(\varepsilon_{\mathrm{LC}})$ rounds of repetition yields a LABEL COVER instance with the desired completeness, soundness, and graph- and projection-degree regularity.

To introduce smoothness, let $R_2 \triangleq \xi^{-1}J^2$ and recall Khot [16]: define new label sets $K' \triangleq [R_2] \times K \times L^{R_2-1}$ and $L' \triangleq L^{R_2}$. The new projections $\pi' : L' \to K'$ are defined as follows: choose $t \in [R_2]$ uniformly at random and project $\vec{j} = (j_1,\ldots,j_{R_2}) \in L'$ to $(t, \pi(j_t), j_1,\ldots,j_{t-1},j_{t+1},\ldots,j_{R_2}) \in K'$. We note by the construction that the label-set cardinality is $k = k(\varepsilon_{\mathrm{LC}},J,\xi) = R_2 2^{R_1} 10^{R_1(R_2-1)}$.

Proving the smoothness property, let T be an arbitrary subset of L' of cardinality at most J. In order for $|\pi'(T)| < |T|$, at least one of the $\binom{T}{2} \leq J^2$ pairs of labels in $T \subseteq L'$ must project to the same label in

K'. Any two distinct labels, $\vec{j}$ and $\vec{j}'$, differ in at least one of their R_2 coordinates and unless each of these coordinates is chosen in the random projection, $\vec{j}$ and $\vec{j}'$ have distinct projections. As only one coordinate is projected, this happens with probability at most $1/R_2$ and a union bound implies that $|\pi'(T)| < |T|$ with probability at most J^2/R_2 which is no greater than ξ by the choice of R_2. $\square$

Permutations of labels When reducing from a LABEL COVER instance with regular projection degrees, we shall identify the "small" label set K with the integer range $[k]$ for some constant k and with the "large" label set L, the set $K \times [d]$. Several of the quantities we derive, such as correlations between sample spaces, are properties of the underlying correlated spaces alone and do not depend on the particular projection, possibly conditioned on projection degrees. For this reason, we prefer to abstract away the projection and permute the sample spaces suitably when applying said bounds to analyze a LABEL COVER reduction. Given a projection $\pi : L \to K$ with projection degrees exactly d, we denote by $\bar{\pi}$ an arbitrary bijection $L \leftrightarrow L$ such that if, for labels $i \in K$ and $(i', r') \in L$, $\pi(i', r') = i$, then there exists $r \in [d]$ such that $\bar{\pi}(i', r') = (i, r)$. In other words, the coordinates are permuted such that $(i, r) \in L$ projects to $i \in K$. We note that the bounds we derive do not depend on the particular choice of $\bar{\pi}$ and we consequently do not specify these bijections further than being consistent with a projection π. This renaming of coordinates is straightforward and it is only for smoothness where particular care must be taken.

To demonstrate the defined notation, consider the classical PCP of Parity on three bits courtesy of Håstad [11]: for a particular LABEL COVER edge e and associated projection π_e, choose strings $\mathbf{x} \in \{-1,1\}^K$ and $\mathbf{y} \in \{-1,1\}^L$ uniformly at random, and set for each $j \in L$, $z_j = x_{\pi_e(j)} \cdot y_j$; the test makes three queries to two tables indexed by the generated strings and accepts if the queries have even parity. Casting the test distribution in our notation, we consider the base distribution $\mathcal{P} = (\Omega_1 \times \Omega_2 \times \Omega_3, \mu)$ where $\Omega_1 = \Omega_2 = \Omega_3 = \{-1,1\}$ and $\mu(x,y,z) = 1/4 \cdot [xy = z]; x,y,z \in \{-1,1\}$. Assuming the projection π_e is d-regular for some constant d, the test distribution then corresponds to

$$\left(\mathcal{P}^{d\text{-proj-}1^{\otimes K}}\right)^{\bar{\pi}_e}$$

where $\bar{\pi}_e$ is an arbitrary bijection consistent with π_e and the superscript denotes permuting the coordinates of the sample spaces $\Omega_2^{[d] \times L}$ and Ω_3^L. In Håstad's construction, the coordinates in $\mathbf{z} = \{z_j\}_{j \in L}$ are additionally resampled independently with some small probability $\gamma > 0$; we express this with the noise operator, choosing as base distribution $\mathrm{T}_{\bar{\gamma}}^{(3)}\,\mathcal{P}$ and as test distribution

$$\mathcal{T} = \left(\left(\mathrm{T}_{\bar{\gamma}}^{(3)}\,\mathcal{P}\right)^{d\text{-proj-}1^{\otimes K}}\right)^{\bar{\pi}_e}.$$

In this setting, lemmas which hold for, e. g.,

$$\left(\mathrm{T}_{\bar{\gamma}}^{(3)}\,\mathcal{P}\right)^{d\text{-proj-}1^{\otimes K}}$$

also hold for $\mathcal{T}$ by identifying coordinates.

1.3 Main theorem

The main theorem of the paper is the following.

Theorem 1.18. *Any predicate $P \subseteq \{0,1\}^m, m \geq 4$, strictly containing odd or even Parity on m bits is approximation resistant for satisfiable instances.*

Our proof defines a distribution on the predicate P and shows that every non-constant term in the Fourier expansion of P must be small in the negative case. Indeed, our proof establishes so-called uselessness of P introduced in Austrin and Håstad [1], and in extension hereditary approximation resistance.

For clarity and being the simplest case of interest, we begin by proving the theorem for width-four predicates. In particular, it suffices to prove the theorem for the arity-four predicate "0,1, or 3" as other predicates either follows from symmetry or the hereditary approximation resistance of the symmetrical predicates.

1.4 Organization of the paper

We define the reduction, its claimed soundness, and completeness in Section 2.2. The completeness argument is straightforward and can be found in Section 2.2.1. As usual with PCP-based hardness results, the soundness analysis is more involved and deferred to a number of separate sections. In particular, Section 2.3 establishes bounds on the correlation between sample spaces of our test distribution; Section 2.4 shows that, due to the correlation bounds, the analyzed expressions behave roughly as though they were noised; Section 2.5 bounds unmixed terms, i. e., terms of the form $\mathbf{E}[\prod g]$; while Section 2.6 bounds mixed expression, i. e., expressions of the form $\mathbf{E}[f \prod g]$, in terms of common influences consistent with projections. The established lemmas are stated in Section 2.2.2 were it is also shown how they prove the desired soundness.

In the last part of this paper, we generalize the arguments to predicates of greater width. Sections 3.2, 3.3, 3.4, 3.5, respectively, generalize the PCP protocol, correlation bounds, noise introduction arguments, and bounds on unmixed terms. In Section 3.6, we establish an invariance-style theorem building on preceding work which we later apply to bound mixed terms of greater width in Section 3.7. Finally, the results of these sections are tied together in Section 3.2.1 where it is shown how they imply the soundness bound of the generalized PCP.

2 Predicates of width four

Let *"0, 1, or 3"* be the arity-4 Boolean predicate accepting strings $\mathbf{x} \subseteq \{0,1\}^4$ containing exactly 0, 1, or 3 ones. In the following, we prove a special case of the main theorem.

Theorem 2.1. *The arity-4 predicate* "0, 1, or 3" *with negation is approximation resistant for satisfiable instances. Put differently, for every $\varepsilon_{CSP} > 0$, it is NP-hard to distinguish whether a* MAX-CSP$^+$("0, 1, or 3") *instance has value 1 or value at most $|P|/2^{-4} + \varepsilon_{CSP} = 9/16 + \varepsilon_{CSP}$.*

2.1 Proof outline

We define a standard Long Code-based protocol reducing from SMOOTH LABEL COVER with the desired completeness and soundness. The protocol works by sampling a string $\mathbf{x}$ uniformly at random from the small table and subsequently defining three strings on the large table by sampling $(y_j^{(2)}, y_j^{(3)}, y_j^{(4)})$ conditioned on $\mathbf{x}_{\pi(j)}$. This second step draws with high probability from the standard three-wise independent distribution on Odd Parity, and otherwise from a distribution which has positive weight on the additional all-zeroes assignment.

For the former distribution, if we had noise, we would by standard arguments have that the acceptance probability is essentially limited to that of a random assignment unless the SMOOTH LABEL COVER instance has a non-trivial labeling. The second distribution is used precisely to have correlations bounded away from one, permitting us to introduce noise with the help of smoothness, even when the projection degrees depend on the desired soundness error.

Having introduced noise, we have to bound terms of the form $\mathbf{E}[\prod g]$ and $\mathbf{E}[f \prod g]$. For the former, we argue that smoothness and partial independence makes the product behave roughly as though we had unique projections; for this case, the distribution is close to three-wise independent and the products are insignificantly correlated. For terms involving functions on both tables, i. e., $\mathbf{E}[f \prod g]$, we show via a multivariate invariance argument that the product is close to $\mathbf{E}[f]\,\mathbf{E}[\prod g] = 0$ unless a non-trivial labeling exists.

2.2 The protocol

The hardness of MAX-CSP$^+(P)$ follows by a reduction from SMOOTH LABEL COVER as it appears in Theorem 1.17 with soundness $\varepsilon_{\mathrm{LC}} = \varepsilon_{\mathrm{LC}}(\varepsilon_{\mathrm{CSP}})$, and label sets $K = [k(\varepsilon_{\mathrm{LC}}, J, \xi)]$ and $L = K \times [d(\varepsilon_{\mathrm{LC}})]$.

To define the reduction R from an instance $\mathcal{I}$, take as variables for the CSP$^+(P)$ instance $\mathrm{R}(\mathcal{I})$ for every vertex $u \in U$, $2^{|K|}$ Boolean variables and for every vertex $v \in V$, $2^{|L|}$ variables. As is standard, we see these variables as functions $f^u : \{0,1\}^K \to \{0,1\}$ and $g^v : \{0,1\}^L \to \{0,1\}$. Let $\mathcal{D}$ be the uniform distribution on "1 or 3" and let $\mathcal{E}$ be the distribution which chooses uniformly at random from $\{0000, 0111\}$ with probability 0.5 and otherwise uniformly at random from $\{1000, 1110, 1101, 1011\}$. Define a polynomial number of constraints corresponding to the following probabilistic verifier.

1. Pick a random vertex $u \in U$ and a random neighbor $v \in V$. Sample $\pi = \pi^{\{u,v\}}$ as defined by the SMOOTH LABEL COVER instance and let $\bar{\pi}$ be an arbitrary bijection $L \leftrightarrow L$ such that for every $i, i' \in K$ and $r \in [d]$, $\pi(i,r) = i'$ iff $\exists_{r' \in [d]} \bar{\pi}(i,r) = (i', r')$.

2. Sample random folding constants $a, b \sim \{0,1\}$. Define $f_a(\mathbf{x}) = a \oplus f^u(a \oplus \mathbf{x})$ and $g_b(\mathbf{y}) = b \oplus g^v(b \oplus \mathbf{y} \circ \bar{\pi})$.

3. For each $i \in K$, independently choose x_i uniformly at random from $\{0,1\}$. For each $j \in L$, independently sample $(x_{\pi(j)}, y_j^{(2)}, y_j^{(3)}, y_j^{(4)})$ conditioned on $x_{\pi(j)}$ from $\mathcal{D}$ with probability $\bar{\delta}$ and otherwise $\mathcal{E}$.

4. Accept iff $\left(f_a(\mathbf{x}), g_b(\mathbf{y}^{(2)}), g_b(\mathbf{y}^{(3)}), g_b(\mathbf{y}^{(4)})\right) \in P$.

We note that queries $a \oplus f(a \oplus \cdot)$ are permitted in MAX-CSP$^+$ where the operation $a \oplus \cdot$ acts as a possible negation of a variable. This construct is called *folding* and ensures that $\mathbf{E}_\mathbf{x}[f] = \mathbf{E}_\mathbf{y}[g] = 0$.

The goal is to show the following two properties of the protocol from which Theorem 2.1 follows. Completeness is argued in Section 2.2.1 and soundness in Section 2.2.2.

Proposition 2.2. *The protocol has completeness 1. Said equivalently, if* $\mathrm{Val}(I) = 1$*, then* $\mathrm{Val}(\mathrm{R}_P(I)) = 1$.

Proposition 2.3. *For arbitrary fixed* $\varepsilon_{CSP} > 0$*, the protocol has soundness* $|P|/16 + \varepsilon_{CSP} = 9/16 + \varepsilon_{CSP}$*. More specifically, if* $\mathrm{Val}(I) \leq \varepsilon_{LC} = \varepsilon_{LC}(\varepsilon_{CSP})$*, then* $\mathrm{Val}(\mathrm{R}_P(I)) \leq 9/16 + \varepsilon_{CSP}$.

Constants To be precise, m, the width of the predicate equals 4. Let $\delta \leq 2^{-16}\varepsilon_{\mathrm{CSP}}^2$, define

$$\rho_0 \triangleq \sqrt{1/2 + 1/2(1-\delta^2/8)^2}\,,$$

and choose $\gamma > 0$ sufficiently close to 0 such that $\sup_k \rho_0^k(1-\bar{\gamma}^k) \leq 2^{-8}\varepsilon_{\mathrm{CSP}}$.[2] Again, define

$$\rho_1 \triangleq \sqrt{1-\gamma^3}$$

and choose $\eta > 0$ sufficiently close to 0 such that $\sup_k \rho_1^k(1-\bar{\eta}^k) \leq 2^{-8}\varepsilon_{\mathrm{CSP}}$. Set the smoothness parameters $J \geq 1 + 2\log(2^{-8}\varepsilon_{\mathrm{CSP}}/10)/\log(\bar{\gamma})$, and $\xi \leq 2^{-18}3^{-3J/2-2}\varepsilon_{\mathrm{CSP}}^2$. Finally, $\varepsilon_{\mathrm{LC}}(\varepsilon_{\mathrm{CSP}}) = 2^{-12}\eta\gamma^3{\varepsilon_{\mathrm{CSP}}}^2$.

Proof of Theorem 2.1

Proof. A random assignment satisfies the predicate P with probability $|P|/16 = 9/16$. To establish the theorem, we wish to show for every $\varepsilon_{\mathrm{CSP}} > 0$ that Gap-$(1, 9/16 + \varepsilon_{\mathrm{CSP}})$ CSP$^+(P)$ is NP-hard. This follows by the following Propositions 2.2 and 2.3, together with the fact that the defined reduction yields at most a polynomial blow-up of instance size. □

2.2.1 Completeness

Proving perfect completeness of the protocol is standard and essentially follows by inspection.

Proof of Proposition 2.2. By the value of the Label Cover instance, there is an assignment $\lambda : U \to K, V \to L$ satisfying for every constraint $\pi^{\{u,v\}}, \{u,v\} \in E, \lambda(u) = \pi^{\{u,v\}}(\lambda(v))$. Define a solution to $\mathrm{R}_P(I)$ by the corresponding dictators, i. e., $\{f^u(\mathbf{x}) = x_{\lambda(u)}\}_{u \in U}$ and $\{g^v(\mathbf{y}) = y_{\lambda(v)}\}_{v \in V}$.

Following the protocol, let u, v, and $\bar{\pi}$ be as chosen, implying $\{u,v\} \in E$. Let $(i,r) = \bar{\pi}(\lambda(v))$; as the Label Cover instance satisfied all constraints, we have $i = \lambda(u)$ and consequently, f_a and g_b satisfies $f_a(\mathbf{x}) = x_i$ and $g_b(\mathbf{y}) = y_{(i,r)}$. The protocol hence accepts iff

$$\left(x_i, y^{(2)}_{(i,r)}, y^{(3)}_{(i,r)}, y^{(4)}_{(i,r)}\right) \in P\,.$$

As the protocol draws this tuple either from $\mathcal{D}$ or $\mathcal{E}$ and their respective support is in P, we conclude that the protocol always accepts. □

[2] For instance, one can take $\bar{\gamma} = \rho^{2^{-8}\varepsilon_{\mathrm{CSP}}}$.

2.2.2 Soundness

As is usual, we establish the soundness via the contradiction of its contrapositive: supposing that the acceptance probability of $\mathrm{R}_P(I)$ is greater than $|P|/16+\varepsilon_{\mathrm{CSP}}$, we show that there is a labeling of the SMOOTH LABEL COVER instance I achieving value greater than $\varepsilon_{\mathrm{LC}}=\varepsilon_{\mathrm{LC}}(\varepsilon_{\mathrm{CSP}})$. The dependency in particular is $\varepsilon_{\mathrm{LC}}(\varepsilon_{\mathrm{CSP}})=2^{-12}\eta\gamma^3{\varepsilon_{\mathrm{CSP}}}^2$ where the noise constants η and γ appear below.

Notation Define the following distributions which appear in our proofs,

$$\mathcal{T}_0=\bar{\delta}\mathcal{D}+\delta\mathcal{E},\quad \mathcal{T}_0'=\mathcal{T}_0^{d\text{-proj-1}\otimes K},\quad \mathcal{T}_1'=\left(\mathrm{T}_{\bar{\gamma}}^{2,3,4}\,\mathcal{T}_0^{d\text{-proj-1}}\right)^{\otimes K},$$

$$\mathcal{T}_2'=\left(\mathrm{T}_{\bar{\gamma}}^{2,3,4}\left(\mathrm{T}_{\bar{\eta}}^{1}\,\mathcal{T}_0\right)^{d\text{-proj-1}}\right)^{\otimes K},\quad \mathcal{T}_3=\mathrm{T}_{\bar{\eta}}^{1}\,\mathrm{T}_{\bar{\gamma}}^{2,3,4}\,\mathcal{T}_0,\quad \mathcal{T}_3'=\mathcal{T}_3^{d\text{-proj-1}\otimes K},\quad \mathcal{T}_3''=\mathcal{T}_3^{\otimes L}.$$

The test distribution of the protocol corresponds to $\mathcal{T}_0'$. Intuitively, $\mathcal{T}_1'$ is the distribution where projected noise is applied to $\mathbf{y}^{(2)},\mathbf{y}^{(3)}$, and $\mathbf{y}^{(4)}$, i. e., all coordinates which share projection are changed by noise simultaneously. $\mathcal{T}_2'$ is the same distribution but with noise applied also to $\mathbf{x}$. We note that projected and non-projected (independent) noise are the same for $\mathbf{x}$ as it is defined on the smaller table. $\mathcal{T}_3'$ is the distribution all strings—$\mathbf{x},\mathbf{y}^{(2)},\mathbf{y}^{(3)}$, and $\mathbf{y}^{(4)}$—all have independent noise. Finally, in $\mathcal{T}_3''$, we have independent noise as in $\mathcal{T}_3'$ but for each $j\in L$, the tuple $(y_j^{(2)},y_j^{(3)},y_j^{(4)})$ is drawn independently when, as in our analysis for this distribution, we are not concerned with $\mathbf{x}$. For notational simplicity, define $f=\mathbf{E}_a[f_a]$ and $g=\mathbf{E}_b[g_b]$. Let the queried points of the functions be

$$q_1\triangleq f(\mathbf{x}),\quad q_2\triangleq g(\mathbf{y}^{(2)}),\quad \ldots,\quad q_4\triangleq g(\mathbf{y}^{(4)})\,.$$

As is usual for PCP analysis, we substitute 1 for -1 and 0 for 1, and freely switch between the two conventions whenever convenient. Considering the Fourier transform $\{\hat{P}_\Gamma\}_{\Gamma\subseteq[4]}$ of the predicate, the acceptance probability of the protocol equals $\mathbf{E}_{E,\mathcal{T}_0'}\left[\sum_{\Gamma\subseteq[4]}\hat{P}_\Gamma\chi_\Gamma(\mathbf{q})\right]$ where the distribution is over a random edge $e\in E$ from the SMOOTH LABEL COVER instance and the arguments from $\mathcal{T}_0'$. For an arbitrary $\Gamma\neq\emptyset$ and distribution $\mathcal{R}$, let us denote by $\psi_\Gamma(\mathcal{R})=\mathbf{E}_{E,\mathcal{R}}[\chi_\Gamma(\mathbf{q})]$. Conceptually, we refer to these terms as $\mathbf{E}[\prod g]$ or $\mathbf{E}[f\prod g]$ for zero or more functions g. We also note that the acceptance probability in the new notation equals $\sum_\Gamma\hat{P}_\Gamma\psi_\Gamma(\mathcal{T}_0')$.

Properties of the protocol By inspection, we see that all distributions above have the basic property that all four arguments, $\mathbf{x},\mathbf{y}^{(2)},\ldots,\mathbf{y}^{(4)}$, have uniform marginals. Additionally, each $\mathbf{y}^{(t)}$ argument is on its own independent of $\mathbf{x}$.

Lemma 2.4. *Let $t\in\{2,3,4\}$ and consider either distribution $\mathcal{T}_r'$, $r=0,1,2,3$. The marginals on $\mathbf{x}$ and $\mathbf{y}^{(t)}$ are uniform and furthermore $\mathbf{y}^{(t)}$ is independent of $\mathbf{x}$.*

Proof. For simplicity of this proof, we use the 0,1-notation. By inspection, $\mathbf{x}$ has uniform marginals for both $\mathcal{D}$ and $\mathcal{E}$ and consequently $\mathcal{T}_0'$. By symmetry, consider the probability of the outcomes 00, 01, 10, and 11 for (x_1,y_2) in $\mathcal{D}$ and $\mathcal{E}$, respectively. Again by inspection, these outcomes each have probability 1/4. As $\mathbf{x}$ has uniform marginals and sampling is done independently for every coordinate j conditioned on $x_{\pi(j)}$, the lemma follows.

For either of the other distributions, it suffices to notice that the noise operators do not change marginals nor introduce dependencies. □

The aim is to show the following four propositions from which the soundness follows. We note that the first proposition establishes basic properties while the remaining three mimic the approach of O'Donnell and Wu [22]. The proofs of these latter three lemmas are deferred to Section 2.4, Section 2.5, and Section 2.6, respectively.

The first proposition states that, due to the preceding independence, terms involving at most one $\mathbf{y}^{(t)}$ argument are zero.

Proposition 2.5. $\psi_\Gamma(\mathcal{T}'_0) = 0$ *for* $\emptyset \neq \Gamma \subseteq [4], |\Gamma \cap \{2,3,4\}| \leq 1$.

Proof. As shown in Lemma 2.4, the test distribution has uniform marginals. Hence $\psi_{\{t\}}(\mathcal{T}'_0) = \mathbf{E}_{E,\mathcal{T}'_0}[q_t]$ which equals $\mathbf{E}_E[\mathbf{E}[f]]$ or $\mathbf{E}_E[\mathbf{E}[g]]$, both of which are 0 due to folding. Suppose $\Gamma = \{1,t\}$. Then $\psi_\Gamma(\mathcal{T}'_0) = \mathbf{E}_{E,\mathcal{T}'_0}[fg] = \mathbf{E}[f]\,\mathbf{E}[g] = 0$ since $\mathbf{y}^{(t)}$ is uniform and independent of $\mathbf{x}$ by Lemma 2.4, subsequently folding yields expectation 0. □

The second lemma, which involves smoothness and significant technical work, argues that terms are in expectation roughly the same with the original test distribution as the test distribution with noise, independent of d, on all arguments. We note that the constants

$$\rho_0 = \sqrt{1/2 + 1/2(1-\delta^2/8)^2} \quad \text{and} \quad \rho_1 = \sqrt{1-\gamma^3}$$

appearing in the proposition are correlation bounds appearing in the proofs and are bounded away from 1 depending only on δ and γ.

Proposition 2.6.

$$\left|\psi_\Gamma(\mathcal{T}'_0) - \psi_\Gamma(\mathcal{T}'_3)\right| \leq \sup_{k\geq 0} \rho_0^k (1-\bar{\gamma}^k) + \sup_{k \geq 0} \rho_1^k (1-\bar{\eta}^k) + 6\sqrt{\xi} + 6\bar{\gamma}^J \leq \varepsilon_{CSP}/256$$

for any $\Gamma \subseteq [4]$.

Third, over the noised distribution, terms of the form $\mathbf{E}[\prod g]$ are shown to have an expectation which approaches 0 as parameters are tweaked.

Proposition 2.7. $|\psi_\Gamma(\mathcal{T}'_3)| \leq 4\bar{\gamma}^{J/2-1} + 6 \cdot 3^{3J/4}\sqrt{\xi} + \sqrt{\delta} \leq \varepsilon_{CSP}/256$ *for* $1 \notin \Gamma \subseteq [4], |\Gamma| \geq 2$.

Finally, we bound terms of the form $\mathbf{E}[f\prod g]$. This is often considered the hardest part of PCP analysis. However, the argument is almost immediate after we extend Mossel's multivariate invariance principle [19] to projection games.

Proposition 2.8. $|\psi_\Gamma(\mathcal{T}'_3)| \leq 8\sqrt{\gamma^{-1}\,\mathbf{E}_E\left[\sum_{(i,j)\in\pi} \mathrm{Inf}_i^{(\bar{\eta})}(f)\mathrm{Inf}_j^{(\bar{\gamma})}(g)\right]}$ *for* $1 \in \Gamma \subseteq [4], |\Gamma| \geq 3$.

Proof of Proposition 2.3 Propositions in hand, we proceed to show how they imply the desired soundness.

Proof. As is usual, we establish the soundness through the contradiction of its contrapositive: supposing that the acceptance probability of $\mathrm{R}_P(I)$ is greater than $|P|/16+\varepsilon_{\mathrm{CSP}}$, we show that there is a labeling of the SMOOTH LABEL COVER instance I achieving value greater than $\varepsilon_{\mathrm{LC}} = \varepsilon_{\mathrm{LC}}(\varepsilon_{\mathrm{CSP}}) = 2^{-12}\eta\gamma^3{\varepsilon_{\mathrm{CSP}}}^2$.

The labeling in question is the $(\bar{\eta},\gamma)$-*Noisy Influence Assignment* which independently sets vertex u resp. v to label i resp. j with probability proportional to $\mathrm{Inf}_i^{(\bar{\eta})}(f^u)$ resp. $\mathrm{Inf}_j^{(\bar{\gamma})}(g^v)$. By Lemma 1.13, this defines probability measures with normalization constants bounded by η and γ, respectively. By this labeling, the SMOOTH LABEL COVER instance has value at least

$$\mathbf{P}_E(\lambda(u)=\pi(\lambda(v))\geq \eta\gamma\mathbf{E}_E\left[\sum_{(i,j)\in\pi}\mathrm{Inf}_i^{(\bar{\eta})}(f)\mathrm{Inf}_j^{(\bar{\gamma})}(g)\right]. \tag{2.1}$$

Suppose that the assignment $\{f^u\}_{u\in U},\{g^v\}_{v\in V}$ achieves value greater than $|P|/16+\varepsilon_{\mathrm{CSP}}$ for some $\varepsilon_{\mathrm{CSP}}>0$. Taking the Fourier expansion of the predicate P, the acceptance probability equals $\sum_\Gamma \hat{P}_\Gamma\psi_\Gamma(\mathcal{T}'_0)$. We note that the term with $\Gamma=\emptyset$ equals $|P|/16$. For the remaining terms, let A,B, and C be the respective choices of Γ appearing in Proposition 2.5, 2.7, and 2.8, respectively. By the first of these lemmas, for any $\Gamma\in A$, $\psi_\Gamma(\mathcal{T}'_0)=0$. We also note that $\left|\hat{P}_\Gamma\right|\leq 1$ for any Γ. Consequently,

$$\sum_{\Gamma\in B+C}\left|\psi_\Gamma(\mathcal{T}'_0)\right|>\varepsilon_{\mathrm{CSP}}\,.$$

By Proposition 2.6 and the choice of parameters,

$$\sum_{\Gamma\in B+C}\left|\psi_\Gamma(\mathcal{T}'_3)-\psi_\Gamma(\mathcal{T}'_0)\right|\leq\varepsilon_{\mathrm{CSP}}/4\,.$$

By Proposition 2.7 and the choice of parameters again,

$$\sum_{\Gamma\in B}\left|\psi_\Gamma(\mathcal{T}'_3)\right|\leq\varepsilon_{\mathrm{CSP}}/4\,.$$

Consequently, using Proposition 2.8 and $|C|\leq 4$, for some $\Gamma\in C$,

$$\varepsilon_{\mathrm{CSP}}/8\leq\left|\psi_\Gamma(\mathcal{T}'_3)\right|\leq 8\sqrt{\gamma^{-1}\,\mathbf{E}_E\left[\sum_{(i,j)\in\pi}\mathrm{Inf}_i^{(\bar{\eta})}(f)\mathrm{Inf}_j^{(\bar{\gamma})}(g)\right]}\,.$$

That is,

$$\mathbf{E}_E\left[\sum_{(i,j)\in\pi}\mathrm{Inf}_i^{(\bar{\eta})}(f)\mathrm{Inf}_j^{(\bar{\gamma})}(g)\right]>2^{-12}\gamma^2{\varepsilon_{\mathrm{CSP}}}^2.$$

Relating to (2.1), we see that the $(\bar{\eta},\gamma)$-noisy influence assignment achieves a value greater than $\varepsilon_{\mathrm{LC}}(\varepsilon_{\mathrm{CSP}})=2^{-12}\eta\gamma^3{\varepsilon_{\mathrm{CSP}}}^2$, as desired. □

2.3 Correlation bounds for the test distribution

In this subsection, we establish bounds on the correlation between strings in our test distribution. We view the distribution $\mathcal{T}_0'$ as sampling $|K|$ copies on the correlated space $\Omega_1 \times \Omega_2^d \times \Omega_3^d \times \Omega_4^d$ where each Ω_t equals $\{-1,1\}$ but is indexed for clarity. These samples form $\mathbf{x}, \mathbf{y}^{(2)}, \mathbf{y}^{(3)}$, and $\mathbf{y}^{(4)}$, respectively.

The correlation bounds we aim to establish for the test distribution are the following. The first lemma shows that for our test distribution $\mathcal{T}_0$, the correlation between arguments to g functions are bounded away from 1 independent of d. This in turn will enable us to introduce projected noise for g functions.

Lemma 2.9.

$$\rho\left(\Omega_1 \times \Omega_2^d \times \Omega_3^d, \Omega_4^d;\ \mathcal{T}_0^{d\text{-proj-1}}\right) \leq \sqrt{\frac{1}{2} + \frac{1}{2}\left(1 - \frac{\delta^2}{8}\right)^2}\,.$$

Proof. Proved in Section 2.3.2. □

The second lemma essentially says that after we have introduced projected noise for all g functions, the argument to f has correlation bounded away from 1 independent of d, enabling us to introduce noise for f.

Lemma 2.10.

$$\rho\left(\Omega_1, \Omega_2^d \times \Omega_3^d \times \Omega_4^d;\ \mathrm{T}_{\bar{\gamma}}^{2,3,4}\, \mathcal{T}_0^{d\text{-proj-1}}\right) \leq \sqrt{1-\gamma^3}\,.$$

Proof. Proved in Section 2.3.2. □

The third and final lemma is used to show that a product of g-functions is always small if we do not have projections. This will be the final step when we bound terms of the form $\mathbf{E}[\prod g]$ after we have argued that the product behaves roughly as though there were unique projections.

Lemma 2.11.

$$\rho\left(\Omega_2, \Omega_3 \times \Omega_4;\ \mathcal{T}_0\right) \leq \sqrt{\delta}\,.$$

Proof. Proved in Section 2.3.2. □

2.3.1 Preliminaries

We begin by introducing the concept of expected conditional expectation and present lemmas which simplify our proofs.

Definition 2.12. Let $\mathcal{P} = (\Omega = \Omega_A \times \Omega_B \times \Omega_C, \mu)$ be a correlated probability space and $\{\mathrm{E}_i\}_i$ a partition of Ω. The *expected conditional correlation* between Ω_A and Ω_B conditioned on $\{\mathrm{E}_i\}_i$ with respect to the measure μ is

$$\rho_\mu(\Omega_A, \Omega_B \mid \{\mathrm{E}_i\}_i;\ \mathcal{P}) = \mathbf{E}_{\mathrm{E}_i \sim \mu}\left[\rho_\mu(\Omega_A, \Omega_B;\ \mathcal{P} \mid \mathrm{E}_i)^2\right]^{1/2},$$

where "$\mathcal{P} \mid \mathrm{E}_i$" is the probability space conditioned on the event E_i.

In the special case when the partition $\{\mathrm{E}_i\}$ is the set of indicators of a correlated space Ω_C, we simplify notation as follows.

Definition 2.13. The *expected conditional correlation* between Ω_A and Ω_B conditioned on Ω_C with respect to the measure μ is

$$\rho_\mu(\Omega_A, \Omega_B \mid \Omega_C;\ \mathcal{P}) = \mathbf{E}_{\omega_C \sim \mu}\left[\rho_\mu(\Omega_A, \Omega_B;\ \mathcal{P} \mid \omega_C)^2\right]^{1/2},$$

where "$\mathcal{P} \mid \omega_C$" is the probability space conditioned on the event "$\Omega_C = \omega_C$."

We recall and slightly reformulate useful lemmas from Mossel [19]. The first lemma gives an explicit expression for computing correlations and is particularly useful for comparing correlations between different probability spaces.

Lemma 2.14 (Lemma 2.8, Mossel [19])**.**

$$\rho_\mu(\Omega_A, \Omega_B) = \sup_{\phi \in \mathrm{L}^2_\mu(\Omega_A),\, \mathbf{E}_\mu[\phi]=0, \mathbf{Var}_\mu[\phi]=1} \mathbf{E}_{\omega_B \sim_\mu \Omega_B}\left[\mathbf{E}_\mu[\phi \mid \omega_B]^2\right]^{1/2}.$$

The second lemma says that if two functions depend on a number of independent correlated spaces, then they can maximize their correlation by using only one set of correlated spaces.

Lemma 2.15 (Proposition 2.13, Mossel [19])**.** *Correlated probability spaces* $\left\{\left(\Omega_A^{(i)} \times \Omega_B^{(i)}, \mu_i\right)\right\}_i$ *satisfy*

$$\rho\left(\prod \Omega_A^{(i)}, \prod \Omega_B^{(i)};\ \prod \mu_i\right) \leq \max \rho\left(\Omega_A^{(i)}, \Omega_B^{(i)};\ \mu_i\right).$$

Finally, the following lemma provides a simple upper bound on the correlation of any probability space; however, it would be deteriorating with projection degrees if used immediately in our applications.

Lemma 2.16 (Lemma 2.9, Mossel [19])**.** *Let* $(\Omega_A \times \Omega_B, P)$ *be a correlated space where the minimum strictly positive probability of any outcome is* α*. Consider the bipartite graph* $G = (\Omega_A, \Omega_B, E)$ *where* $\{x, y\} \in E$ *iff* $(x, y) \in \Omega_A \times \Omega_B$ *has strictly positive probability. If* G *is connected, then*

$$\rho(\Omega_A, \Omega_B;\ P) \leq 1 - \alpha^2/2.$$

We establish in the following three lemmas useful for dealing with correlations. The first shows formally the intuitive property that the expected correlation between two sample spaces conditioned on a set of events partition the product sample space is greater than the correlation without knowledge of any events. This will prove useful in bounding correlations which involve sample spaces growing with projection degrees.

Lemma 2.17. *Let* $(\prod^m \Omega_t, \mu)$ *be a correlated probability space,* A, B *non-empty subsets of* $[m]$*, and* $\{\mathrm{E}_i\}_i$ *a partition of the sample space such that either* Ω_A *or* Ω_B *is independent of* $\{\mathrm{E}_i\}_i$*, i. e., the marginal distribution of the sample space* Ω_A *or* Ω_B *remains unchanged conditioned on any event* $E \in \{\mathrm{E}_i\}_i$*. Then,*

$$\rho(\Omega_A, \Omega_B;\ \mathcal{P}) \leq \rho(\Omega_A, \Omega_B \mid \{E_i\}_i;\ \mathcal{P}).$$

Proof. Without loss of generality, let Ω_B be independent of $\{\mathrm{E}_i\}_i$. Note that $\mathbf{E}_E[\cdot]$ in this proof denotes the expectation over events in $\{\mathrm{E}_i\}_i$; not the expectation over a random edge as in the analysis of our protocol.

By Lemma 2.14 and subsequently Jensen's inequality, the LHS equals

$$\sup_{\phi} \mathbf{E}_{\omega_A}\left[\mathbf{E}_{\Omega_B}[\phi \mid \omega_A]^2\right]^{1/2} \leq \sup_{\phi} \mathbf{E}_{E,\omega_A}\left[\mathbf{E}_{\Omega_B}[\phi \mid E, \omega_A]^2\right]^{1/2}, \tag{2.2}$$

where the supremum is over $\phi \in \mathrm{L}^2(\Omega_B)$ with expectation 0 and variance 1 with respect to μ. Using that

$$\sup_{x \in X} \mathbf{E}\left[x^2\right]^{1/2} = \left(\sup_{x \in X} \mathbf{E}\left[x^2\right]\right)^{1/2}$$

for $x \geq 0$ and subsequently independence between Ω_B and the events,

$$(2.2) = \left(\sup_{\phi} \mathbf{E}_{E,\omega_A}\left[\mathbf{E}_{\Omega_B}[\phi \mid E, \omega_A]^2\right]\right)^{1/2} \leq \mathbf{E}_E\left[\sup_{\phi} \mathbf{E}_{\omega_A \mid E}\left[\mathbf{E}_{\Omega_B}[\phi \mid \omega_A]^2\right]\right]^{1/2}. \tag{2.3}$$

Again from Lemma 2.14, we recognize this supremum as $\rho\left(\Omega_A, \Omega_B;\ \mathcal{P} \mid E\right)^2$. Consequently,

$$(2.3) = \mathbf{E}_{E \in \{E_i\}_i}\left[\rho\left(\Omega_A, \Omega_B \mid E;\ \mathcal{P}\right)^2\right]^{1/2},$$

which we recognize as the definition of $\rho\left(\Omega_A, \Omega_B \mid \{E_i\}_i;\ \mathcal{P}\right)$. □

As a corollary, there is a simple bound on the correlation between a mixture of distributions from the correlations of the respective distributions.

Corollary 2.18. *Let $(\Omega_A, \Omega_B, \bar{\delta}\mu + \delta\nu)$ be a correlated space such that the marginal distribution of at least one of Ω_A and Ω_B is identical on μ and ν. Then,*

$$\rho\left(\Omega_A, \Omega_B;\ \bar{\delta}\mu + \delta\nu\right) \leq \sqrt{\bar{\delta}\rho_\mu(\Omega_A, \Omega_B)^2 + \delta\rho_\nu(\Omega_A, \Omega_B)^2}.$$

Proof. Suppose that the marginal of Ω_A is identical on μ and ν. In consequence, the marginal coincides for these distributions with $\bar{\delta}\mu + \delta\nu$. We extend the sample space by an indicator of the drawn-from distribution and note that it does not change the correlation. Since Ω_A is independent of this outcome, we can apply Lemma 2.17 to the indicator for the desired corollary. □

Finally, there is a simple bound between sample spaces independent of projection degrees. Namely, if Ω_A are the sample spaces which do not grow with the projection degree d, and conditioning on Ω_A does not change the marginal distribution of one of the considered sample spaces, then the spaces cannot achieve higher correlation than having Ω_A revealed and using a single of the d independent samples.

Corollary 2.19. *Let A, B, C be disjoint subsets of $[m]$ and $\mathcal{P} = (\prod^m \Omega_t, \mu)$ a correlated probability space such that Ω_C is independent of Ω_A. Then,*

$$\rho\left(\Omega_A \times \Omega_B^d, \Omega_C^d;\ \mathcal{P}^{d\text{-proj-}A}\right) \leq \rho\left(\Omega_B, \Omega_C \mid \Omega_A;\ \mathcal{P}\right).$$

Proof. We use Lemma 2.17 with the events $\{E_i\}_i$ the outcomes of Ω_A, which are independent of Ω_C by the corollary hypothesis. The LHS is thusly bounded

$$\text{LHS} \leq \rho\left(\Omega_B^d, \Omega_C^d \mid \Omega_A;\ \mathcal{P}^{d\text{-proj-}A}\right) = \mathbf{E}_{\omega_A}\left[\rho\left(\Omega_B^d, \Omega_C^d;\ (\mathcal{P} \mid \omega_A)^{\otimes d}\right)^2\right]^{1/2}.$$

From Lemma 2.15, we know that these correlations are bounded by any the greatest correlation of any one sample, i. e.,

$$\mathbf{E}_{\omega_A}\left[\rho\left(\Omega_B, \Omega_C;\ \mathcal{P} \mid \omega_A\right)^2\right]^{1/2},$$

which we identify as the definition of $\rho\left(\Omega_B, \Omega_C \mid \Omega_A;\ \mathcal{P}\right)$. □

2.3.2 Proof of Lemmas 2.9 to 2.11

Preliminaries in hand, we turn to the proofs of the main lemmas of this subsection. The proofs are ordered by simplicity.

Proof of Lemma 2.11. With probability $\bar{\delta}$, Ω_4 is drawn from $\mathcal{D}$ in which case it is independent of $\Omega_2^d \times \Omega_3^d$, yielding a correlation of 0. In the other event, the correlation is bounded by 1. Hence, using Theorem 2.18,

$$\rho\left(\Omega_2 \times \Omega_3, \Omega_4;\ \mathcal{T}_3^{d\text{-proj-}1}\right) \leq \sqrt{\bar{\delta} \cdot 0^2 + \delta \cdot 1^2} \leq \sqrt{\delta}\,.$$ □

Proof of Lemma 2.10. We recall that the considered distribution is $\mathrm{T}_{\bar{\gamma}}^{2,3,4}\,\mathcal{T}_0^{d\text{-proj-}1}$. With probability γ^3, the outcome of $\Omega_2^d \times \Omega_3^d \times \Omega_4^d$ is independent of Ω_1 and the correlation is 0. Denote this event by A and let the correlations of the two possibilities be, respectively, $\rho_{\mathrm{A}} = 0$ and $\rho_{\bar{\mathrm{A}}} \leq 1$. Then,

$$\rho\left(\Omega_1, \Omega_2^d \times \Omega_3^d \times \Omega_4^d;\ \mathrm{T}_{\bar{\gamma}}^{2,3,4}\,\mathcal{T}_0^{d\text{-proj-}1}\right) \leq \sqrt{\mathbf{P}(\mathrm{A})\,\rho_{\mathrm{A}}^2 + \mathbf{P}(\bar{\mathrm{A}})\,\rho_{\bar{\mathrm{A}}}^2} \leq \sqrt{\gamma^3 \cdot 0^2 + (1-\gamma^3) \cdot 1^2} = \sqrt{1-\gamma^3}\,.$$
□

Proof of Lemma 2.9. Lemma 2.4 implies that Ω_4^d is independent of Ω_1. Applying Theorem 2.19 with $A = \{1\}, B = \{2,3\}, C = \{4\}$, we get

$$\rho\left(\Omega_1 \times \Omega_2^d \times \Omega_3^d, \Omega_4^d;\ \mathcal{T}_0^{d\text{-proj-}1}\right) \leq \rho\left(\Omega_2 \times \Omega_3, \Omega_4 \mid \Omega_1;\ \mathcal{T}_0\right)$$

which by definition equals $\mathbf{E}_{\omega_1}\left[\rho\left(\Omega_2 \times \Omega_3, \Omega_4;\ \mathcal{T}_0 \mid \omega_1\right)^2\right]^{1/2}$.

Switching to $\{0,1\}$ notation again, to establish that the considered correlation is bounded away from 1, it suffices that the conditioned correlation is bounded away from 1 for at least one of the cases $\omega_1 = 0$ and $\omega_1 = 1$. This is precisely what we do, we bound the latter by 1 and find a smaller bound for the case $\omega_1 = 0$.

To this end, we employ Lemma 2.16. The bipartite graph in question has left vertices $\Omega_2 \times \Omega_3$, right vertices Ω_4, and an edge $\{(\omega_2, \omega_3), \omega_4\}$ whenever $(\omega_2, \omega_3, \omega_4)$ has strictly positive probability, $\mathbf{P}(\{(\omega_1 = 0, \omega_2, \omega_3, \omega_4)\}) > 0$. Lemma 2.16 states that if this graph is connected, then the correlation is bounded away from 1. To be specific, at most $1-\alpha^2/2$ where α is the smallest strictly positive probability of any outcome.

Conditioned on $\omega_1 = 0$ and $\omega_4 = 0$, all choices of (ω_2, ω_3) except $(1,1)$ have positive probability: those of odd parity from $\mathcal{D}$ and the $(0,0)$ outcome from $\mathcal{E}$. In effect, all left vertices in the graph except for $(\omega_2, \omega_3) = (1,1)$ are connected to $\omega_4 = 0$. On the other hand, conditioned on $\omega_1 = 0$ and $\omega_4 = 1$, the distribution $\mathcal{D}$ has positive probability for both $(\omega_2, \omega_3) = (0,0)$ and $(1,1)$, allowing us to conclude that the graph indeed is connected.

By inspection, the minimum strictly positive probability of any outcome α for the distribution $\mathcal{T}_0 = \bar{\delta}\mathcal{D} + \delta\mathcal{E}$ conditioned on $\omega_1 = 0$ is given by $\mathcal{E}$ as δ is close to 0. $\mathcal{E}$ only has two outcomes conditioned on $\omega_1 = 0$, namely 0000 and 0111, each with conditioned probability 0.5. Consequently, $\alpha = \delta/2$, implying $\rho(\Omega_2 \times \Omega_3, \Omega_4;\ \mathcal{T}_0 \mid \omega_1 = 0) \leq 1 - \delta^2/8$ and, as desired,

$$\rho\left(\Omega_1 \times \Omega_2^d \times \Omega_3^d, \Omega_4^d;\ \mathcal{T}_0^{d\text{-proj-1}}\right) \leq \sqrt{\frac{1}{2} \cdot 1^2 + \frac{1}{2}\left(1 - \frac{\delta^2}{8}\right)^2}. \qquad \square$$

2.4 Noise introduction

We first show a theorem which allow us to go from projected noise to independent non-projected noise.

2.4.1 Equivalence of projected and non-projected noise

In this subsection, we show that when analyzing an expectation $\mathbf{E}[\prod g]$ or $\mathbf{E}[f \prod g]$ and projections are chosen as in SMOOTH LABEL COVER, the expected difference of this term with projected and non-projected noise is bounded by a constant depending only on and strictly decreasing with smoothness parameters ξ^{-1} and J.

By non-projected noise, we mean the classical noise in hardness of approximation: every coordinate $y_j \in \Omega$ of a string $\mathbf{y}$ is independently resampled from the marginal of the distribution with a fixed probability. By projected noise, we mean the noise as used by Mossel [19]: every group of coordinates $\{y_j\}_{\pi(j)=i} \in \Omega^{|\pi^{-1}(i)|}$ with the same projection are jointly resampled from the distribution's marginal with a fixed probability.

Lemma 2.20. *Consider label sets $K, L = K \times [d]$ and a function $g = g^v : \Omega_C^L \to \mathbb{R}$ defined on a probability space*

$$\mathcal{P} = (\Omega_A \times \Omega_B \times \Omega_C, \mu)^{d\text{-proj-}A^{\otimes K}}$$

where Ω_C is independent of Ω_A. Let $\pi = \pi^{\{u,v\}} : L \to K$ be chosen (J, ξ)-smooth and define $\mathcal{P}^{\bar{\pi}}$ by permuting the arguments of g in accordance with an arbitrary permutation $\bar{\pi}$ consistent with π. Let $\{g_S\}_{S \subseteq L}$ be the Efron-Stein decomposition of g and denote by U_S^π the event '$|\pi(S)| \geq \min\{J, |S|\}$'. Then,

$$\mathbf{E}_\pi\left[\left\|\sum_{S : \overline{\mathrm{U}_S^\pi}} g_S\right\|_2\right] \leq \sqrt{\xi \mathbf{Var}[g]},$$

where the norm is over $\mathcal{P}^{\bar{\pi}}$.

Proof. We let the implicit distribution of the proof draw $\mathcal{P}$, π and set $\mathcal{P}^{\bar{\pi}}$ accordingly. We also note that by the hypothesis that Ω_C is independent of Ω_A, the Efron-Stein decomposition of g with respect to $\mathcal{P}$ satisfies the usual properties of Lemma 1.5. In fact, as neither g nor its marginal distribution depends on π, the Efron-Stein decomposition with regard to the distribution $\mathcal{P}^{\bar{\pi}}$ is the same regardless of π. For this reason, we shall not specify further which distribution the decomposition is with respect to.

We begin by rewriting the left term,

$$\text{LHS} = \mathbf{E}\left[\left(\sum_S \left[\overline{\mathrm{U}_S^{\pi}}\right] g_S\right)^2\right]^{1/2} = \mathbf{E}_{\pi}\left[\sum_{S,T}\left[\overline{\mathrm{U}_S^{\pi}} \wedge \overline{\mathrm{U}_T^{\pi}}\right] \mathbf{E}_{\mathcal{P}^{\bar{\pi}}}[g_S g_T]\right]^{1/2}. \tag{2.4}$$

By the properties of Efron-Stein decompositions, $\mathbf{E}_{\mathcal{P}^{\bar{\pi}}}[g_S g_T]$ is 0 unless $S = T$. Furthermore, $\mathbf{E}_{\mathcal{P}^{\bar{\pi}}}\left[g_S^2\right]$ only depends on the marginal distribution of Ω_C^M and is independent of π. Hence,

$$(2.4) = \left(\sum_S \mathbf{P}_{\pi}\left(\left[\overline{\mathrm{U}_{\mathrm{S}}^{\pi}}\right]\right) \mathbf{E}_{\mathcal{P}}\left[g_S^2\right]\right)^{1/2}. \tag{2.5}$$

The permutation π was chosen to be (J,ξ)-smooth and consequently the probability in the expression is bounded by ξ, 0 in particular for $S = \emptyset$, and the remaining sum is the variance of g by Parseval's:

$$(2.5) \le \left(\xi \sum_{S \neq \emptyset} \mathbf{E}_{\mathcal{P}}\left[g_S^2\right]\right)^{1/2} = (\xi\, \mathbf{Var}[g])^{1/2}. \qquad \square$$

Theorem 2.21. *Let K and $L = K \times [d]$ be label sets, $\gamma \in (0,1]$ a parameter, $\mathcal{D}$ a distribution on $\Omega_A \times \Omega_B \times \Omega_C$ such that Ω_C is independent of Ω_A, and finally $h : \Omega_A^K \times \Omega_B^L \to [-1,1]$ and $g : \Omega_C^L \to [-1,1]$ functions. Define*

$$\mathcal{P} = \left(\mathrm{T}_{\bar{\gamma}}^{(C)}\, \mathcal{D}\right)^{d\text{-proj-}A^{\otimes K}}, \quad \mathcal{R} = \left(\mathrm{T}_{\bar{\gamma}}^{(C)}\, \mathcal{D}^{d\text{-proj-}A}\right)^{\otimes K}.$$

Additionally, let $\pi : L \to K$ be chosen (J,ξ)-smooth and define $\mathcal{P}^{\bar{\pi}}$ and $\mathcal{R}^{\bar{\pi}}$ by permuting coordinates accordingly. Then,

$$\left|\mathbf{E}_{\pi,\mathcal{P}^{\bar{\pi}}}[hg] - \mathbf{E}_{\pi,\mathcal{R}^{\bar{\pi}}}[hg]\right| \le 2\sqrt{\xi} + 2\bar{\gamma}^J. \tag{2.6}$$

Proof. We note that Ω_C^L is independent of Ω_A^K in both $\mathcal{P}$ and $\mathcal{R}$ despite noise. Furthermore, neither kind of noise affects the marginals of Ω_C^L; in fact the marginals of $\mathcal{P}, \mathcal{R}$, and $\mathcal{D}^{d\text{-proj-}A^{\otimes K}}$ coincide and consequently they share Efron-Stein decompositions. Rewriting the LHS,

$$\begin{aligned}\text{LHS} &= \left|\mathbf{E}_{\pi,\mathcal{P}^{\bar{\pi}}}\left[h \sum_{S:\mathrm{U}_S^{\pi}} g_S\right] - \mathbf{E}_{\pi,\mathcal{R}^{\bar{\pi}}}\left[h \sum_{S:\mathrm{U}_S^{\pi}} g_S\right] + \mathbf{E}_{\pi,\mathcal{P}^{\bar{\pi}}}\left[h \sum_{S:\overline{\mathrm{U}_S^{\pi}}} g_S\right] - \mathbf{E}_{\pi,\mathcal{R}^{\bar{\pi}}}\left[h \sum_{S:\overline{\mathrm{U}_S^{\pi}}} g_S\right]\right| \\ &\le \left|\mathbf{E}_{\pi,\mathcal{P}^{\bar{\pi}}}\left[h \sum_{S:\mathrm{U}_S^{\pi}} g_S\right] - \mathbf{E}_{\pi,\mathcal{R}^{\bar{\pi}}}\left[h \sum_{S:\mathrm{U}_S^{\pi}} g_S\right]\right| + \left|\mathbf{E}_{\pi,\mathcal{P}^{\bar{\pi}}}\left[h \sum_{S:\overline{\mathrm{U}_S^{\pi}}} g_S\right]\right| + \left|\mathbf{E}_{\pi,\mathcal{R}^{\bar{\pi}}}\left[h \sum_{S:\overline{\mathrm{U}_S^{\pi}}} g_S\right]\right|. \end{aligned} \tag{2.7}$$

Applying Cauchy-Schwarz to either of the two latter terms and subsequently Lemma 2.20,

$$(2.7) \leq \left| \mathbf{E}_{\pi,\mathcal{P}^{\bar{\pi}}}\left[h \sum_{S:\mathrm{U}^{\pi}_{S}} g_S \right] - \mathbf{E}_{\pi,\mathcal{R}^{\bar{\pi}}}\left[h \sum_{S:\mathrm{U}^{\pi}_{S}} g_S \right] \right| + 2\sqrt{\xi \mathbf{Var}[g]} \|h\|_2 .$$

Since the domain of g and h is $[-1,1]$ we note that their variance and l_2^2-norms are both bounded by 1. This yields the first term on the RHS of (2.6).

We proceed to bound the difference. The event U^{π}_{S} corresponds to $|\pi(S)| = |S|$ and/or $|\pi(S)| \geq J$. Whenever the former holds, the expectation over $\mathcal{P}^{\bar{\pi}}$ and $\mathcal{R}^{\bar{\pi}}$ coincide, i. e., projected and non-projected noise are the same.

It remains to bound the case when $|S| \neq |\pi(S)| \geq J$. However due to noise, the contribution of such terms is small. Let us show this formally through standard analysis, denoting

$$\mathcal{D}' = \mathcal{D}^{d\text{-proj-}A^{\otimes K}}$$

and noting that the larger of the two terms is over $\mathcal{R}'$ which can be seen in the following derivation by substituting $|\pi(S)|$ for $|S|$.

$$\begin{aligned}
&\left| \mathbf{E}_{\pi,\mathcal{P}^{\bar{\pi}}}\left[h \sum_{S:|S|\neq|\pi(S)|\geq J} g_S \right] - \mathbf{E}_{\pi,\mathcal{R}^{\bar{\pi}}}\left[h \sum_{S:|S|\neq|\pi(S)|\geq J} g_S \right] \right| \leq 2 \left| \mathbf{E}_{\pi,\mathcal{R}^{\bar{\pi}}}\left[h \sum_{S:|S|\neq|\pi(S)|\geq J} g_S \right] \right| \\
&= 2 \left| \mathbf{E}_{\pi,\mathcal{D}'^{\bar{\pi}}}\left[h \sum_{S:|S|\neq|\pi(S)|\geq J} \bar{\gamma}^{|\pi(S)|} g_S \right] \right| \leq 2\, \mathbf{E}_{\pi,\mathcal{D}'^{\bar{\pi}}}\left[h^2\right]^{1/2} \mathbf{E}_{\pi,\mathcal{D}'^{\bar{\pi}}}\left[\left(\sum_{S:|S|\neq|\pi(S)|\geq J} \bar{\gamma}^{|\pi(S)|} g_S \right)^2 \right]^{1/2} \\
&\leq 2 \sum_{S:|S|\neq|\pi(S)|\geq J} \mathbf{E}_{\pi,\mathcal{D}'^{\bar{\pi}}}\left[\bar{\gamma}^{2|\pi(S)|} g_S^2 \right]^{1/2} \leq 2\bar{\gamma}^{J}\, \mathbf{E}_{\pi,\mathcal{D}'^{\bar{\pi}}}\left[\sum_{S} g_S^2 \right]^{1/2} \leq 2\bar{\gamma}^{J}\, \mathbf{E}_{\pi,\mathcal{D}'^{\bar{\pi}}}\left[g^2\right]^{1/2} \leq 2\bar{\gamma}^{J} . \qquad \square
\end{aligned}$$

2.4.2 Proof of Proposition 2.6: Introducing noise in soundness analysis

The introduction of noise, independent of d, follows three steps. The first step is to argue that the correlation of the argument of a function is bounded away from one independent of d. As a second step, this permits us to employ a theorem by Mossel [19] to introduce a certain kind of noise, which we call *projected noise*, without changing the expectation too much. Finally we use our techniques of the preceding subsection to show that the introduced noise behaves roughly as independent noise.

To clarify these points, the noise introduced by the mentioned theorem, which we shall call projected noise, jointly resamples all d coordinates $j \in L$ which project to the same coordinate $i \in K$. What we would like is for every coordinate $j \in L$ to be resampled independently. This is an important distinction in soundness analysis where for instance the former may permit arbitrarily large parities of coordinates in $\pi^{-1}(i)$ to achieve a significant value. Standard decoding procedures, including Håstad's classical, and the more modern low-degree or noisy influences, fail in this setting as the number of potential coordinates grows with the soundness of the LABEL COVER instance and, in extension, the degree of projections.

We circumvent these problems with the machinery from the preceding subsection; smoothness essentially guarantees that functions which depend on many coordinates, depend on many coordinates with different projections. In effect, we show that projected noise behave roughly the same as independent noise in this setting.

Formalizing slightly, the first step of Proposition 2.6 is to introduce projected noise for every $\mathbf{y}^{(t)}$ string, i. e., to go from the distribution $\mathcal{T}'_0$ to $\mathcal{T}'_1$. Having done so, the correlation of $\mathbf{x}$ to $(\mathbf{y}^{(t)})_t$ is bounded away from one independent of d, permitting noise introducing for $\mathbf{x}$, defined as the distribution $\mathcal{T}'_2$. We note that projected and independent noise are the same for $\mathbf{x}$. Finally, we use smoothness and the projected noise of $\mathbf{y}^{(t)}$ to show that it behaves roughly as independent noise, yielding the distribution $\mathcal{T}'_3$. The three steps may be expressed through the following lemmas.

Lemma 2.22. *Let $\Gamma \subseteq [4]$, $|\Gamma| \geq 2$ and define $\rho_0 = \sqrt{1/2 + 1/2(1-\delta^2/8)^2}$. Then,*

$$\left|\mathbf{E}\left[\psi_\Gamma(\mathcal{T}'_0) - \psi_\Gamma(\mathcal{T}'_1)\right]\right| \leq 3 \sup_k \rho_0^k (1 - \bar{\gamma}^k)\,.$$

Proof. Proved in Section 2.4.2. □

Lemma 2.23. *Let $\Gamma \subseteq [4]$, $|\Gamma| \geq 2$ and define $\rho_1 = \sqrt{1-\gamma^3}$. Then,*

$$\left|\mathbf{E}\left[\psi_\Gamma(\mathcal{T}'_1) - \psi_\Gamma(\mathcal{T}'_2)\right]\right| \leq \sup_k \rho_1^k (1 - \bar{\eta}^k)\,.$$

Proof. Proved in Section 2.4.2. □

Lemma 2.24. *Let $\Gamma \subseteq [4]$, $|\Gamma| \geq 2$. Then,*

$$\left|\mathbf{E}\left[\psi_\Gamma(\mathcal{T}'_2) - \psi_\Gamma(\mathcal{T}'_3)\right]\right| \leq 6\sqrt{\xi} + 6\bar{\gamma}^J\,.$$

Proof. Proved in Section 2.4.2. □

Proof of Proposition 2.6. Follows directly from Lemmas 2.22, 2.23, and 2.24 by summing the respective differences. □

We begin by recapping a corollary of a lemma by Mossel [19] and next define the concept of *"lifted" functions* with notation borrowed from Dinur et al. [6].

Corollary 2.25 (Corollary of Lemma 6.1, Mossel [19])**.** *Let $\mathcal{P} = (\Omega_1 \times \Omega_2, \mu)$ be a correlated probability space satisfying $\rho(\Omega_1, \Omega_2) \leq \rho < 1$. Consider functions $\{f_t : \Omega_t^n \to [-1,1]\}_{t=1,2}$ and an arbitrary constant $\gamma \in (0,1]$. Then,*

$$\left|\mathbf{E}_{\mathcal{P}^{\otimes n}}[f_1 f_2] - \mathbf{E}_{\mathcal{P}^{\otimes n}}\left[(\mathrm{T}_{\bar{\gamma}} f_1) f_2\right]\right| \leq \sup_k \rho^k (1 - \bar{\gamma}^k)\,.$$

In the setting of Theorem 2.25, the involved functions are defined on a product space $(\Omega'_1 \times \Omega'_2, \mu)^{\otimes n}$. For proper treatment, we define equivalent functions where the sample spaces are products of all same sample spaces with the same projection.

Definition 2.26. The lifted function $\overline{g} : \{\{-1,1\}^d\}^K$ of $g : \{-1,1\}^{K\times[d]}$ is defined as

$$\overline{g}(\overline{\mathbf{y}}) = g(\mathbf{y})\,,$$

where its lifted argument $\overline{\mathbf{y}}$ satisfies $\overline{y}_{i,r} = y_{(i,r)}$, $i \in K, r \in [d]$.

We define $\Omega'_1 = \Omega_1, \Omega'_t = \Omega_t^d$ for $t = 2,3,4$, let $\overline{\mathbf{y}^{(t)}}$ be the lifted version of $\mathbf{y}^{(t)}$, and $\overline{\mathcal{T}}$ the lifted analogue of a distribution $\mathcal{T}$. Additionally, as we wish to claim simultaneously the lemmas for all $\Gamma \subseteq [4], |\Gamma| \geq 2$, let h_A for a subset $A \subseteq [4]$ denote

$$f^{[1\in A]} \prod_{t \in A\setminus 1} \overline{g^{(t)}}\,.$$

Proof of Lemma 2.22: Introducing projected noise for g functions

Proof. Let

$$\mathcal{D}_1 = \overline{\mathcal{T}_0^{d\text{-proj-1}}}, \quad \mathcal{D}_r = \mathrm{T}_{\bar{\gamma}}^{(r)}\, \mathcal{D}_{r-1}, \qquad r = 2,3,4\,.$$

We note that $\mathcal{D}_r^K$ is indeed the lifted analogue of $\mathcal{T}'_1$. Consequently, the lemma is proved by bounding the respective differences of expectations

$$\left|\psi_\Gamma(\mathcal{D}_r^K) - \psi_\Gamma(\mathcal{D}_{r-1}^K)\right|$$

for $r = 2,3,4$.

Working out the notation, for $r = 2,3,4$,

$$\begin{aligned}\left|\psi_\Gamma(\mathcal{D}_{r-1}^K) - \psi_\Gamma(\mathcal{D}_r^K)\right| &= \left|\mathbf{E}_{\mathcal{D}_{r-1}^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right] - \mathbf{E}_{\mathcal{D}_r^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right]\right| \\ = \left|\mathbf{E}_{\mathcal{D}_{r-1}^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right] - \mathbf{E}_{(\mathrm{T}_{\bar{\gamma}}^{(r)}\mathcal{D}_{r-1})^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right]\right| &= \left|\mathbf{E}_{\mathcal{D}_{r-1}^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right] - \mathbf{E}_{\mathcal{D}_{r-1}^K}\left[(\mathrm{T}_{\bar{\gamma}}\overline{g^{(r)}}) h_{\Gamma\setminus r}\right]\right|. \end{aligned} \tag{2.8}$$

This is the setting of Theorem 2.25.

By Lemma 2.9 and symmetry, the correlation $\rho(\Omega_1 \times \prod_{t\neq 1,r}\Omega_t^d, \Omega_r^d; \mathcal{T}_0^{d\text{-proj-1}})$, which equals $\rho(\Omega_1 \times \prod_{t\neq 1,r}\Omega'_t, \Omega'_r; \mathcal{D}_1)$, is bounded by

$$\rho_0 \triangleq \sqrt{1/2 + 1/2(1-\delta^2/8)^2}\,.$$

As noise can only decrease correlation, the same bound holds for $\mathcal{D}_{r-1}$. Similarly, this is a bound on any subset of sample spaces in the case $\Gamma \neq [4]$.

For $r = 2,3,4$, if $r \notin \Gamma$, the difference (2.8) is 0. Otherwise, we bound using Theorem 2.25 with $\rho \leq \rho_0$. That is, (2.8) $\leq \sup_k \rho_0^k(1-\bar{\gamma}^k)$. In conclusion,

$$\left|\psi_\Gamma(\mathcal{T}'_0) - \psi_\Gamma(\mathcal{T}'_1)\right| = \left|\psi_\Gamma(\mathcal{D}_1^K) - \psi_\Gamma(\mathcal{D}_4^k)\right| \leq \sum_{t=2}^{4}\left|\psi_\Gamma(\mathcal{D}_{t-1}^K) - \psi_\Gamma(\mathcal{D}_t^K)\right| \leq 3\sup_k \rho_0^k(1-\bar{\gamma}^k)\,. \qquad \square$$

Proof of Lemma 2.23: Introducing noise for the f function

Proof. By Lemma 2.10,

$$\rho(\Omega_1, \Omega_2^d \times \Omega_3^d \times \Omega_4^d; \mathcal{T}_1^{d\text{-proj-1}}) \leq \rho_1 \triangleq \sqrt{1-\gamma^3}\,.$$

The same bound holds for $\rho(\Omega_1, \prod_{t\in\Gamma\setminus 1}\Omega'_t; \overline{\mathcal{T}_1^{d\text{-proj-1}}})$. Hence, using Theorem 2.25 again,

$$\begin{aligned}\left|\psi_\Gamma(\mathcal{T}'_1)-\psi_\Gamma(\mathcal{T}'_2)\right| &= \left|\mathbf{E}_{\left(\mathrm{T}^{2,3,4}_{\bar\gamma}\mathcal{T}_0^{d\text{-proj-1}}\right)^K}\left[fh_{\Gamma\setminus 1}\right]-\mathbf{E}_{\left(\mathrm{T}^{2,3,4}_{\bar\gamma}\left(\mathrm{T}^1_{\bar\eta}\mathcal{T}_0\right)^{d\text{-proj-1}}\right)^K}\left[fh_{\Gamma\setminus r}\right]\right| \\ &= \left|\mathbf{E}_{\left(\mathrm{T}^{2,3,4}_{\bar\gamma}\mathcal{T}_0^{d\text{-proj-1}}\right)^K}\left[fh_{\Gamma\setminus r}\right]-\mathbf{E}_{\left(\mathrm{T}^{2,3,4}_{\bar\gamma}\mathcal{T}_0^{d\text{-proj-1}}\right)^K}\left[(\mathrm{T}_{\bar\eta}f)h_{\Gamma\setminus r}\right]\right| \le \sup_k \rho_1^k(1-\bar\eta^k)\,. \square\end{aligned}$$

Proof of Lemma 2.24: From projected noise to independent noise

Proof. The lemma follows immediately from Theorem 2.21. However, somewhat confusingly at this point, we have defined the functions g so that the sample spaces are correlated by the projection $\pi(i,r)=i$. To utilize the theorem, we unravel the definition from the protocol:

$$g(\mathbf{y}) \triangleq \mathbf{E}_{b\sim\{0,1\}}[b\oplus g^v(b\oplus\mathbf{y}\circ\bar\pi)]\,,$$

where, for $\pi\triangleq\pi^{\{u,v\}}$, $\bar\pi : K\times[d]\leftrightarrow K\times[d]$ is an arbitrary bijection such that if $(i',r')=\bar\pi(i,r)$, then $\pi(i)=i'$. Define

$$g'^v(\mathbf{y}) = \mathbf{E}_{b\sim\{0,1\}}[b\oplus g^v(b\oplus\mathbf{y})]$$

and let $\mathcal{T}'^{\pi}_2$ and $\mathcal{T}'^{\pi}_3$ permute coordinates of g' in accordance with an arbitrary bijection $\bar\pi$ consistent with π. Also, for $F\subseteq[4]$, let

$$h^v_F = f^{[1\in F]}\prod_{t\in F\setminus 1} g'^v(\mathbf{y}^{(t)})\,.$$

We recall the definition of the distributions

$$\mathcal{T}'_2 \triangleq \left(\mathrm{T}^{2,3,4}_{\bar\gamma}\left(\mathrm{T}^1_{\bar\eta}\,\mathcal{T}_0\right)^{d\text{-proj-1}}\right)^K \quad\text{and}\quad \mathcal{T}'_3 \triangleq \left(\left(\mathrm{T}^{2,3,4}_{\bar\gamma}\mathrm{T}^1_{\bar\eta}\,\mathcal{T}_0\right)^{d\text{-proj-1}}\right)^K\,.$$

The target difference equals

$$\left|\psi_\Gamma(\mathcal{T}'_2)-\psi_\Gamma(\mathcal{T}'_3)\right| = \left|\mathbf{E}_{u,v,\mathcal{T}'^{\pi^{\{u,v\}}}_2}\left[fh^v_{\Gamma\setminus 1}\right]-\mathbf{E}_{u,v,\mathcal{T}'^{\pi^{\{u,v\}}}_3}\left[fh^v_{\Gamma\setminus 1}\right]\right|. \tag{2.9}$$

We apply Theorem 2.21 up to three times, once for each of the coordinates appearing in Γ. Define $\mathcal{R}_t$ for $t\in[m]$ as follows and note that $\mathcal{R}_1$ corresponds to $\mathcal{T}'_2$ and $\mathcal{R}_m$ to $\mathcal{T}'_3$,

$$\mathcal{R}_t \triangleq \left(\mathrm{T}^{t+1,\dots,m}_{\bar\gamma_{t+1},\dots,\bar\gamma_m}\left(\mathrm{T}^{1,2,\dots,t}_{\bar\eta,\bar\gamma_2,\dots,\bar\gamma_t}\,\mathcal{T}_0\right)^{d\text{-proj-1}}\right)^K\,.$$

Formally, when applying the theorem to coordinate $t\in\Gamma\setminus\{1\}$, we have $A=\{1\}, B=\{t\}, C=\Gamma\setminus\{1,t\}, \gamma=\gamma$, and the distributions equal $\mathcal{P}=\mathcal{R}_{t-1}$ and $\mathcal{R}=\mathcal{R}_t$. The respective differences in expectation hence yield a bound on the difference in expectation between $\mathcal{T}'_2=\mathcal{R}_2$ and $\mathcal{T}'_3=\mathcal{R}_m$. According to the theorem, the difference in expectation for each application is bounded by $2\sqrt{\xi}+2\bar\gamma^J_t$. In effect, (2.9) $\le 6\sqrt{\xi}+6\bar\gamma^J_2$. $\square$

2.5 Proof of Proposition 2.7: Bounding $\mathbf{E}_{\mathcal{T}_3'}[\prod g]$

We limit ourselves to the hardest case: $\Gamma = \{2,3,4\}$ which corresponds to bounding

$$\mathbf{E}_{\pi,\mathcal{T}_3'}\left[g(y^{(2)})g(y^{(3)})g(y^{(4)})\right].$$

The other cases are bounded by the same line of argument which we comment on following the proof of the $\Gamma = \{2,3,4\}$ case.

Intuitively, due to noise the expression can be arbitrarily well approximated by low-degree expansions of g. Because of smoothness, low-degree terms in the Efron-Stein decomposition project to unique coordinates with probability arbitrarily close to one. For terms of functions with such projections, the expectation is the same for the distribution which draws arguments independently for every coordinate, as the one which draws $\mathbf{x}$ and adheres to the projections. Finally analyzing a distribution which is independent for every coordinate, we can simply bound by the expectation by the unprojected correlation between the three spaces which was shown in Lemma 2.11 to be at most $\sqrt{\delta}$. We recall again that $\mathcal{T}_3'$ is a noised version of $\mathcal{T}_0'$ and noise can not increase correlation.

In this subsection, let k be the largest integer less than $J/2$ where J is one of the smoothness parameters of our protocol. Recall that

$$\mathcal{T}_3' \triangleq \mathcal{T}_3^{d\text{-proj-1}\otimes K} \quad \text{and} \quad \mathcal{T}_3'' \triangleq \mathcal{T}_3^{\otimes L}.$$

In the following, g denotes g'^v as defined in the previous subsection.

Noised to low-degree functions

Lemma 2.27. *Let $\mathcal{D} = \mathcal{T}_3'$ or $\mathcal{T}_3''$. Then,*

$$\left|\mathbf{E}_{\mathcal{D}}\left[g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right] - \mathbf{E}_{\mathcal{D}}\left[g^{\le k}(\mathbf{y}^{(2)})g^{\le k}(\mathbf{y}^{(3)})g^{\le 2k}(\mathbf{y}^{(4)})\right]\right| \le 2\bar{\gamma}^k.$$

Proof. Let us denote by $g_t = g(\mathbf{y}^{(t)})$. Clearly,

$$\mathbf{E}[g_2g_3g_4] = \mathbf{E}\left[(g_2^{\le k} + g_2^{>k})(g_3^{\le k} + g_3^{>k})g_4\right]$$

and so,

$$\left|\mathbf{E}[g_2g_3g_4] - \mathbf{E}\left[g_2^{\le k}g_3^{\le k}g_4\right]\right| \le \left|\mathbf{E}\left[g_2^{>k}g_3g_4\right]\right| + \left|\mathbf{E}\left[g_2^{\le k}g_3^{>k}g_4\right]\right|.$$

We bound both terms on the RHS through Hölder's inequality. Let $h_2 \in \{g_2^{\le k}, g_2^{>k}\}$ and $h_3 \in \{g_3, g_3^{>k}\}$. Then,

$$|\mathbf{E}[h_2h_3g_4]| \le \|h_2\|_{\mathcal{D},2}\|h_3\|_{\mathcal{D},2}\|g_4\|_{\mathcal{D},\infty}.$$

As $\left\|g^{\le k}\right\|_2, \left\|g^{>k}\right\|_2 \le \|g\|_2$ for any function g with an Efron-Stein decomposition, and $\|g\|_{\mathcal{D},p} \le \|g\|_{\mathcal{T}_0',p} = 1$ for any integer $p \ge 1$, all three norms are bounded by 1. Additionally, if $h_2 = g_2^{>k}$, then

$$\|h_2\|_{\mathcal{D},2} = \sum_{S:|S|>k} \mathbf{E}_{\mathcal{D}}\left[g_S^2\right] = \sum_{S:|S|>k} \bar{\gamma}^{|S|}\,\mathbf{E}_{\mathcal{T}_0'}\left[g_S^2\right] \le \max\{\bar{\gamma}^{|S|} : |S| \ge k\}\sum_S \mathbf{E}\left[g_S^2\right] \le \bar{\gamma}^k.$$

Similarly, if $h_3 = g_3^{>k}$, then $\|h_3\|_{\mathcal{D},2} \le \bar{\gamma}^k$. Hence,

$$\left|\mathbf{E}[g_2 g_3 g_4] - \mathbf{E}\left[g_2^{\le k} g_3^{\le k} g_4\right]\right| \le \bar{\gamma}^k + \bar{\gamma}^k \le 2\bar{\gamma}^k .$$

To complete the lemma, we note that

$$\mathbf{E}\left[g_2^{\le k} g_3^{\le k} g_4\right] = \mathbf{E}\left[g_2^{\le k} g_3^{\le k} g_4^{\le 2k}\right] + \mathbf{E}\left[g_2^{\le k} g_3^{\le k} g_4^{> 2k}\right] ,$$

where the latter term equals

$$\begin{aligned}
&\sum_{\substack{S,T,U\subseteq L\\ |S|,|T|\le k; |U|>2k}} \mathbf{E}\left[g_S(\mathbf{y}^{(2)}) g_T(\mathbf{y}^{(3)}) g_U(\mathbf{y}^{(4)})\right] \\
&= \sum_{\substack{S,T,U\subseteq L\\ |S|,|T|\le k; |U|>2k}} \mathbf{E}\left[g_S(\mathbf{y}^{(2)}) g_T(\mathbf{y}^{(3)})\, \mathbf{E}_{\mathbf{y}^{(4)}_{S\cup T} \,\mid\, \mathbf{y}^{(2)}_S \mathbf{y}^{(3)}_T}\left[\mathbf{E}\left[g_U(\mathbf{y}^{(4)}) \mid \mathbf{y}^{(4)}_{S\cup T}\right]\right]\right]
\end{aligned}$$

which is 0 by properties of the Efron-Stein decomposition of g as $U - S - T$ is non-empty. □

Smooth low-degree to shattered functions via hypercontractivity The goal in this paragraph is to show that for smooth projections, low-degree functions essentially behave as their shattered parts. To this end, we employ the following well-known corollary of the Hypercontractivity Theorem, Lemma 1.7.

Lemma 2.28. *Let $\mathcal{D} = \mathcal{T}_3'$ or $\mathcal{T}_3''$. Then,*

$$\left|\mathbf{E}_{\pi,\mathcal{D}^\pi}\left[g_2^{\le k} g_3^{\le k} g_4^{\le 2k} - g_2^{\not\pi \le k} g_3^{\not\pi \le k} g_4^{\not\pi \le 2k}\right]\right| \le \sqrt{\xi}\, 3^{1.5k+1}.$$

Proof. It suffices to bound the terms

$$\begin{aligned}
&\left|\mathbf{E}\left[g_2^{\le k} g_3^{\le k} g_4^{\le 2k} - g_2^{\not\pi \le k} g_3^{\le k} g_4^{\le 2k}\right]\right| ,\\
&\left|\mathbf{E}\left[g_2^{\not\pi \le k} g_3^{\le k} g_4^{\le 2k} - g_2^{\not\pi \le k} g_3^{\not\pi \le k} g_4^{\le 2k}\right]\right|, \text{and}\\
&\left|\mathbf{E}\left[g_2^{\not\pi \le k} g_3^{\not\pi \le k} g_4^{\le 2k} - g_2^{\not\pi \le k} g_3^{\not\pi \le k} g_4^{\not\pi \le 2k}\right]\right| ,
\end{aligned}$$

where $g^{\not\pi}$ is the shattered part of g with respect to π, as defined in the preliminaries.

Assuming $2k \le J$, Lemma 2.20 bounds the first difference by

$$\sqrt{\xi\, \mathbf{Var}\left[g_2^{\le k}\right]}\left\|g_3^{\le k} g_4^{\le k}\right\|_2 .$$

Cauchy-Schwarz and using $\mathbf{Var}\left[g_2^{\le k}\right] = \left\|g_2^{\le k}\right\|_2^2$ yields the further bound

$$\sqrt{\xi}\left\|g^{\le k}\right\|_{\pi,\mathcal{D}^\pi,2}\left\|g^{\le k}\right\|_{\pi,\mathcal{D}^\pi,4}\left\|g^{\le 2k}\right\|_{\pi,\mathcal{D}^\pi,4} .$$

Employing Lemma 1.7 to the two l_4-norms, the first term is at most

$$\left|\mathbf{E}\left[g_2^{\le k} g_3^{\le k} g_4^{\le 2k} - g_2^{\not\pi \le k} g_3^{\le k} g_4^{\le 2k}\right]\right| \le \sqrt{\xi}\, 3^{1.5k}\left\|g^{\le k}\right\|_{\pi,\mathcal{D}^\pi,2}^2\left\|g^{\le 2k}\right\|_{\pi,\mathcal{D}^\pi,2} \le 3^{1.5k}\sqrt{\xi} .$$

The remaining two terms follow the same argument for a total error of

$$3^{1.5k}\sqrt{\xi} + 3^{1.5k}\sqrt{\xi} + 3^k\sqrt{\xi} \le 3^{1.5k+1}\sqrt{\xi} .$$

□

Shattered functions to independent coordinates

Lemma 2.29. *With respect to any projection π, the product of the shattered parts of the functions g_2, g_3, and g_4 have the same expectation for $\mathcal{T}_3'$ as if the coordinates in L were drawn independently. Formally,*

$$\mathbf{E}_{\mathcal{T}_3'}\left[g_2{}^{\not\pi\le k} g_3{}^{\not\pi\le k} g_4{}^{\not\pi\le 2k}\right] = \mathbf{E}_{\mathcal{T}_3''}\left[g_2{}^{\not\pi\le k} g_3{}^{\not\pi\le k} g_4{}^{\not\pi\le 2k}\right].$$

Proof. Consider the Efron-Stein decompositions of the functions in the two terms. Let Θ denote the set

$$\{(S,T,U) \subseteq L^3 : |S|, |T| \le k; |U| \le 2k; |\pi(S)| = |S|, |\pi(T)| = |T|, |\pi(U)| = |U|\},$$

i. e., the triplets of sets which are shattered and satisfy the degree restrictions; that is, $g_2{}^{\not\pi\le k} g_3{}^{\not\pi\le k} g_4{}^{\not\pi\le 2k}$ equals

$$\sum_{(S,T,U)\in\Theta} g_S(\mathbf{y}^{(2)}) g_T(\mathbf{y}^{(3)}) g_U(\mathbf{y}^{(3)}).$$

We argue that in fact each term, indexed by $(S,T,U) \in \Theta$, coincides in expectation for the two distributions. We note that since the terms are shattered, if $|\pi(S+T+U)| < |S+T+U|$, then one of the sets S, T, and U must contain a unique element and the term evaluates to 0 in expectation. Hence, we only need to consider the case when for each $i \in K$, there is at most one $j \in S+T+U$ projecting to i. As the two distributions have identical marginals, the Efron-Stein decompositions are identical, and it suffices to prove that the distribution over relevant argument is the same for the two distributions. This should be intuitively clear but we show it formally for completeness.

By linearity of expectation and properties of the decomposition, it suffices to show that

$$\mathbf{P}_{\mathcal{T}_3'}\Big(\mathbf{y}_S^{(2)}, \mathbf{y}_T^{(3)}, \mathbf{y}_U^{(4)}\Big) = \mathbf{P}_{\mathcal{T}_3''}\Big(\mathbf{y}_S^{(2)}, \mathbf{y}_T^{(3)}, \mathbf{y}_U^{(4)}\Big)$$

for any $(\mathbf{y}_S^{(2)}, \mathbf{y}_T^{(3)}, \mathbf{y}_U^{(4)}) \in \Omega_2^S \times \Omega_3^T \times \Omega_4^U$ such that $(S,T,U) \in \Theta$. With respect to the distribution $\mathcal{T}_3$, let $X, Y^{(2)}, Y^{(3)}$, and $Y^{(4)}$ denote the random variables taking values in $\Omega_1, \ldots, \Omega_4$, respectively. Furthermore, define A_j as the event "$Y^{(2)} = y_j^{(2)}$ if $j \in S, \ldots, Y^{(4)} = y_j^{(4)}$ if $j \in U$." Then,

$$\begin{aligned}\mathbf{P}_{\mathcal{T}_3'}\Big(\mathbf{y}_S^{(2)}, \mathbf{y}_T^{(3)}, \mathbf{y}_U^{(4)}\Big) &= \mathbf{E}_{\mathbf{x}_{\pi(S\cup T\cup U)}\sim\mathcal{T}_3'}\Big[\mathbf{P}_{\mathcal{T}_3'}\Big(\mathbf{y}_S^{(2)}, \mathbf{y}_T^{(3)}, \mathbf{y}_U^{(4)} \,|\, \mathbf{x}_{\pi(S\cup T\cup U)}\Big)\Big] \\ &= \prod_{i\in\pi(S\cup T\cup U)} \sum_{x_i} \mathbf{P}_{\mathcal{T}_3}(X = x_i) \prod_{j\in S\cup T\cup U\,:\,\pi(j)=i} \mathbf{P}_{\mathcal{T}_3}(\mathrm{A}_j \,|\, X = x_i). \qquad (2.10)\end{aligned}$$

We recall that the sets $(S,T,U) \in \Theta$ are shattered. Hence, for any $i \in \pi(S \cup T \cup U)$, there is exactly one $j \in S \cup T \cup U$ such that $\pi(j) = i$. It follows that (2.10) equals

$$\begin{aligned}&\prod_{i\in\pi(S\cup T\cup U)} \prod_{j\in S\cup T\cup U\,:\,\pi(j)=i} \sum_{x_i} \mathbf{P}_{\mathcal{T}_3}(X = x_i)\, \mathbf{P}_{\mathcal{T}_3}(\mathrm{A}_j \,|\, X = x_i) \\ &= \prod_{j\in S\cup T\cup U} \sum_{x_{\pi(j)}} \mathbf{P}_{\mathcal{T}_3}\big(X = x_{\pi(j)}\big)\, \mathbf{P}_{\mathcal{T}_3}\big(\mathrm{A}_j \,|\, X = x_{\pi(j)}\big) = \prod_{j\in S\cup T\cup U} \mathbf{P}(\mathrm{A}_j) = \mathbf{P}_{\mathcal{T}_3''}\Big(\mathbf{y}_S^{(2)}, \mathbf{y}_T^{(3)}, \mathbf{y}_U^{(4)}\Big). \qquad \square\end{aligned}$$

Putting it together

Proof of Proposition 2.7. Using the three preceding lemmas and $J/2-1 \le k \le J/2$,

$$\left|\mathbf{E}_{\mathcal{T}_3'}\left[g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right] - \mathbf{E}_{\mathcal{T}_3''}\left[g^{\not\rightarrowtail\pi \le k}(\mathbf{y}^{(2)})g^{\not\rightarrowtail\pi \le k}(\mathbf{y}^{(3)})g^{\not\rightarrowtail\pi \le 2k}(\mathbf{y}^{(4)})\right]\right| \le 2\bar{\gamma}^{J/2-1} + 3^{3J/4+1}\sqrt{\xi}$$

and

$$\left|\mathbf{E}_{\mathcal{T}_3''}\left[g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right] - \mathbf{E}_{\mathcal{T}_3''}\left[g^{\not\rightarrowtail\pi \le k}(\mathbf{y}^{(2)})g^{\not\rightarrowtail\pi \le k}(\mathbf{y}^{(3)})g^{\not\rightarrowtail\pi \le 2k}(\mathbf{y}^{(4)})\right]\right| \le 2\bar{\gamma}^{J/2-1} + 3^{3J/4+1}\sqrt{\xi}\,.$$

In effect,

$$\left|\mathbf{E}_{\mathcal{T}_3'}\left[g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right]\right| \le \left|\mathbf{E}_{\mathcal{T}_3''}\left[g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right]\right| + 4\bar{\gamma}^{J/2-1} + 6\cdot 3^{3J/4}\sqrt{\xi}\,.$$

It remains to bound

$$\left|\mathbf{E}_{\mathcal{T}_3''}\left[g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right]\right|. \tag{2.11}$$

By the definition of correlation,

$$(2.11) \le \rho(\Omega_2^L, \Omega_3^L \times \Omega_4^L;\ \mathcal{T}_3'')\left\|g(\mathbf{y}^{(2)})\right\|_{\mathcal{T}_3'',2}\left\|g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right\|_{\mathcal{T}_3'',2} \le \rho(\Omega_2^L, \Omega_3^L \times \Omega_4^L;\ \mathcal{T}_3'')\,.$$

For the distribution $\mathcal{T}_3''$, the coordinates L are independent and so by Lemma 2.15,

$$\rho(\Omega_2^L, \Omega_3^L \times \Omega_4^L;\ \mathcal{T}_3'') \le \max_{j \in L} \rho(\Omega_{2,j}, \Omega_{3,j} \times \Omega_{4,j};\ \mathcal{T}_3'') = \rho(\Omega_2, \Omega_3 \times \Omega_4;\ \mathcal{T}_3)\,.$$

In Lemma 2.11, we bounded this correlation by $\sqrt{\delta}$.

Consequently,

$$\left|\mathbf{E}_{\mathcal{T}_3'}\left[g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right]\right| \le 4\bar{\gamma}^{J/2-1} + 2\cdot 3^{3J/4+1}\sqrt{\xi} + \sqrt{\delta}\,.$$ □

Regarding terms $\Gamma \subsetneq \{2,3,4\}, |\Gamma| \ge 2$, we note that the same argument works. The corresponding bounds on high-degree terms only produce fewer terms and similarly with the step from low-degree terms to shattered low-degree terms; the argument that the expectation is the same as for independent coordinates is identical and finally the correlation between two sample spaces is no greater than between three.

2.6 Proof of Proposition 2.8: Bounding $\mathbf{E}_{\mathcal{T}_3'}[f\prod g]$

In this subsection, we bound mixed, i. e., $\mathbf{E}[f\prod g]$, terms. Our proof follows O'Donnell and Wu's [22] coordinate-wise distribution-substitution method although we analyze it immediately via Efron-Stein decompositions and avoid dependencies on d.

Proof of Proposition 2.8

Proof. Fix $E=\{u,v\}$ and π which yields f and g. We show for for $1\in\Gamma\subseteq[4], |\Gamma|\geq 3$ that

$$\psi_\Gamma(\mathcal{T}_3')\leq 8\gamma^{-1/2}\sqrt{\sum_{(i,j)\in\pi}\mathrm{Inf}_i^{(\bar{\eta})}(f)\mathrm{Inf}_j^{(\bar{\gamma})}(g)}\,.$$

For simplicity, we limit ourselves to the hardest case: $\Gamma=[4]$ which corresponds to

$$\mathbf{E}_{\mathcal{T}_3'}\left[f(\mathbf{x})g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right]$$

and the other cases are briefly addressed following this proof. Hereafter, let f and g, respectively, denote $\mathrm{T}_{\bar{\eta}}\, f$ and $\mathrm{T}_{\bar{\gamma}}\, g$ and we instead analyze

$$\mathbf{E}_{\mathcal{T}_0'}\left[f(\mathbf{x})g(\mathbf{y}^{(2)})g(\mathbf{y}^{(3)})g(\mathbf{y}^{(4)})\right]\,.$$

Since any one string $\mathbf{y}^{(t)}$ is independent of $\mathbf{x}$ for the test distribution $\mathcal{T}_0'$, the Efron-Stein decomposition of g with respect to L has the standard properties from Lemma 1.5. Furthermore, as the marginals coincide, the decomposition is the same for the three occurrences of g. Hence,

$$\mathbf{E}\left[f\prod g\right]=\sum_{S,\vec{T}}\mathbf{E}[f_S g_{T_2} g_{T_3} g_{T_4}]\,,$$

where the arguments have been dropped as they are implicit from the subscripts.

Let $\mathcal{H}_\delta'$ be the distribution which samples $\mathbf{x}$ and $(\mathbf{y}^{(2)},\mathbf{y}^{(3)},\mathbf{y}^{(4)})$ independently from $\mathcal{T}_0'$. Our goal is to show

$$\left|\sum_{S,\vec{T}}\mathbf{E}_{\mathcal{T}_0'}[f_S g_{T_2} g_{T_3} g_{T_4}]-\mathbf{E}_{\mathcal{H}_\delta'}[f_S g_{T_2} g_{T_3} g_{T_4}]\right|\leq 8\gamma^{-1/2}\sqrt{\sum_{(i,j)\in\pi}\mathrm{Inf}_i(f)\mathrm{Inf}_j(g)}\,.$$

This would complete the proof as $\mathbf{E}_{\mathcal{H}_\delta'}[f_S g_{T_2} g_{T_3} g_{T_4}]=\mathbf{E}_{\mathcal{T}_0'}[f]\,\mathbf{E}_{\mathcal{T}_0'}[\prod g]=0$.

We note first that for any term $(S,\vec{T})$ with $S=\emptyset$, the expectations are identical and we have a difference of 0. Similarly, any term $(S,\vec{T})$ with $\bigcup T_t=\emptyset$ corresponds to $\mathbf{E}[f_S g_\emptyset g_\emptyset g_\emptyset]=0$. For the remaining terms, the following are well defined:

$$i^*(S,\vec{T})=\max\{i\in S\},\ j^*(S,\vec{T})=\min\{j\in\bigcup T_t\mid\pi(j)=i^*\},W(S,\vec{T})=\{t\mid j^*\in T_t\}\,.$$

We further note that any term with $|W(S,\vec{T})|=1$ evaluates to 0 by the assumption that any single string $\mathbf{y}^{(t)}$ is independent of $\mathbf{x}$ and so by the properties of the Efron-Stein decomposition of g, $\mathbf{E}[f_S\prod g_{T_t}]=0$. Furthermore, for any remaining term, the expectation over $\mathcal{H}_\delta'$ is 0 as $\mathbf{E}[f_S]=0$ for any S.

Hence, it remains to bound

$$\left|\sum_{(i^*,j^*)\in\pi}\sum_{\substack{W\subseteq\{2,3,4\}\\|W|\geq 2}}\sum_{S,\vec{T}}\{\mathbf{E}_{\mathcal{T}_0'}[f_S g_{T_2} g_{T_3} g_{T_4}]\mid i^*(S,\vec{T})=i,j^*(S,\vec{T})=j,W(S,\vec{T})=W\}\right|\,. \tag{2.12}$$

We bound the inner sums separately. Fix arbitrary i^*, j^*, and W. Define J^* as

$$\{j \in L \mid \pi(j) \neq i \vee j \leq j^*\}.$$

Let $\mathcal{J}_1^*$ and $\mathcal{J}_0^*$ be all subsets of J^* containing, respectively not containing, j^*. The reader can convince eirself that the indices $(S, \vec{T})$ which satisfy the conditions on i^*, j^*, and W are precisely those where S is a subset of $[i^*]$ containing i^* and each T_t is a member of $\mathcal{J}_1^*$ or $\mathcal{J}_0^*$ depending on whether $t \in W$. Consequently, we rewrite the inner sum as

$$\sum_{S,\vec{T}} \{\mathbf{E}_{\mathcal{T}_0'}[f_S g_{T_2} g_{T_3} g_{T_4}] \mid i^*(S,\vec{T}) = i^*, j^*(S,\vec{T}) = j^*, W(S,\vec{T}) = W\}$$

$$= \sum \{\mathbf{E}_{\mathcal{T}_0'}[f_S g_{T_2} g_{T_3} g_{T_4}] \mid i^* \in S \subseteq [i^*], T_t \in \mathcal{J}_{[t \in W]}^*\} = \mathbf{E}_{\mathcal{T}_0'}\left[\sum_{i^* \in S \subseteq [i^*]} f_S \prod_t \sum_{T_t \in \mathcal{J}_{[t\in W]}^*} g_{T_t}\right].$$

Before we continue, consider an arbitrary function h with Efron-Stein decomposition $\{h_S\}_{S \subseteq [n]}$. By Lemma 1.10,

$$\sum_{S \subseteq T} h_S(\mathbf{x}) = \mathbf{E}[h(X) \mid X_T = \mathbf{x}_T].$$

Consequently, if $h \to [-1,1]$, then $\left|\sum_{S \subseteq T} h_S\right| \leq 1$. Similarly,

$$\left|\sum_{i \in S \subseteq T} h_S\right| \leq \left|\sum_{S \subseteq T} h_S\right| + \left|\sum_{S \subseteq T - \{i\}} h_S\right| \leq 2.$$

We also note that

$$\left\|\sum_{S \in \mathcal{S}} h_S\right\|_2 = \mathbf{E}\left[\left(\sum_{S \in \mathcal{S}} h_S\right)^2\right]^{1/2} = \left(\sum_{S \in \mathcal{S}} \mathbf{E}\left[h_S^2\right]\right)^{1/2}.$$

Additionally, if $\mathcal{S} \subseteq \mathcal{S}'$, then clearly

$$\sum_{S \in \mathcal{S}} \mathbf{E}\left[h_S^2\right] \leq \sum_{S \in \mathcal{S}'} \mathbf{E}\left[h_S^2\right].$$

Finally, by the definition of influences, $\sum_{S \ni i} \mathbf{E}\left[h_S^2\right] = \mathrm{Inf}_i(h)$.

From our preceding discussions, we consider an arbitrary W of cardinality at least 2. Consequently, let w_1, w_2 be two arbitrary members of W and w_3 be the unique remaining element in $\{2,3,4\}$. Applying Hölder's inequality to the last expression, we receive the bound

$$\left|\mathbf{E}_{\mathcal{T}_0'}\left[\sum_{i^* \in S \subseteq [i^*]} f_S \prod_t \sum_{T_t \in \mathcal{J}_{[t \in W]}^*} g_{T_t}\right]\right| = \left|\mathbf{E}_{\mathcal{T}_0'}\left[\left(\sum_{i^* \in S \subseteq [i^*]} f_S \sum_{j^* \in T_{w_1} \subseteq J^*} g_{T_{w_1}}\right)\left(\sum_{j^* \in T_{w_2} \subseteq J^*} g_{T_{w_2}}\right)\left(\sum_{T_t \in \mathcal{J}_{[w_3 \in W]}^*} g_{T_{w_3}}\right)\right]\right|$$

$$\leq \left\|\sum_{i^* \in S \subseteq [i^*]} f_S \sum_{j^* \in T_{w_1} \subseteq J^*} g_{T_{w_1}}\right\|_2 \left\|\sum_{j^* \in T_{w_2} \subseteq J^*} g_{T_{w_2}}\right\|_2 \left\|\sum_{T_{w_3} \in \mathcal{J}_{[w_3 \in W]}^*} g_{T_{w_3}}\right\|_\infty.$$

By independence between $\mathbf{x}$ and $\mathbf{y}^{(w_1)}$, the first factor is bounded by

$$\Big\| \sum_{i^* \in S \subseteq [i^*]} f_S \Big\|_2 \Big\| \sum_{j^* \in T_{w_1} \subseteq J^*} g_{T_{w_1}} \Big\|_2 .$$

The three two-norms are respectively bounded by $\mathrm{Inf}_{i^*}(f)$, $\mathrm{Inf}_{j^*}(g)$, and $\mathrm{Inf}_{j^*}(g)$. The third factor in the last expression is bounded by $\prod_{t \neq w_1, w_2} 2\|g\|_\infty \leq 2$ following the preceding discussion and noting that $\mathcal{J}_0^*$ are the subsets of $J^* - \{j^*\}$ and $\mathcal{J}_1^*$ are the sets T such that $j^* \in T \subseteq J^*$. Hence,

$$\Big\| \sum_{i^* \in S \subseteq [i^*]} f_S \sum_{j^* \in T_{w_1} \subseteq J^*} g_{T_{w_1}} \Big\|_2 \Big\| \sum_{j^* \in T_{w_2} \subseteq J^*} g_{T_{w_2}} \Big\|_2 \Big\| \sum_{T_{w_3} \in \mathcal{J}^*_{[w_3 \in W]}} g_{T_{w_3}} \Big\|_\infty \leq 2\sqrt{\mathrm{Inf}_{i^*}(f)\mathrm{Inf}_{j^*}(g)\mathrm{Inf}_{j^*}(g)} .$$

Summing over the choices of i^*, j^*, W, we get a bound on (2.12):

$$\sum_{(i^*,j^*) \in \pi} \sum_{W, |W| \geq 2} \mathbf{E}_{w_1 \neq w_2 \in W}\left[2\sqrt{\mathrm{Inf}_{i^*}(f)\mathrm{Inf}_{j^*}(g)\mathrm{Inf}_{j^*}(g)}\right] \quad \leq 8 \sum_{(i,j) \in \pi} \sqrt{\mathrm{Inf}_i(f)\mathrm{Inf}_j(g)\mathrm{Inf}_j(g)} . \tag{2.13}$$

Using Cauchy-Schwarz over (i, j), noting that π is a projection, (2.13) is bounded by

$$8\left(\sum_{(i,j) \in \pi} \mathrm{Inf}_i(f)\mathrm{Inf}_j(g)\right)^{1/2} \left(\sum_j \mathrm{Inf}_j(g)\right)^{1/2} . \tag{2.14}$$

Recalling that we set $f = \mathrm{T}_{\bar{\eta}} f$ and $g = \mathrm{T}_{\bar{\gamma}} g$, these are indeed noisy influences and consequently the total influence of g is at most γ^{-1}. That is, returning to our original definition,

$$(2.14) \leq 8\gamma^{-1/2} \sqrt{\sum_{(i,j) \in \pi} \mathrm{Inf}_i^{(\bar{\eta})}(f)\mathrm{Inf}_j^{(\bar{\gamma})}(g)} . \qquad \square$$

We have bounded the value of the Fourier term of P where all four arguments appear. For other mixed terms $\mathbf{E}[f \prod g]$, the same argument applies where one for arguments which do not appear can use the constant-one function instead. The fact that f appears, has expectation 0, and that the constant-one function has no influences offers the same upper bound.

3 Predicates of greater width

In the remainder of this treatise, we generalize the result of the preceding sections to predicates of width greater than four. The majority of the argument generalizes straightforwardly and is simply included for the sake of completeness. The first fundamental difference is that we use a test distribution that has a certain stronger independence property which is used to bound shattered unmixed terms. The second difference is that we prove and a stronger invariance-style theorem which we employ to directly bound mixed, i. e., $\mathbf{E}[f \prod g]$, terms.

Formally, we establish the following theorem.

Theorem 3.1. *Any predicate $P \subseteq \{0,1\}^m, m \geq 4$, strictly containing odd or even Parity is approximation resistant for satisfiable instances.*

Without loss of generality, we merely show for $m \geq 4$ the approximation resistance of the predicate containing all length-m binary strings of odd parity as well as the all-zero string. All other cases can be reduced from this setting; formally, supposing $P' \subseteq \{0,1\}^m$ contains either odd or even Parity plus a single disjoint element $\mathbf{r} \in \{0,1\}^m$, one reduces constraints "$(x_1+c_1,\ldots,x_m+c_m) \in P$" to "$(x_1+c_1+r_1,\ldots,x_m+c_m+r_m) \in P'$" where addition is taken in $\mathbb{F}_2$. One can verify that the former constraint is satisfied by an assignment if and only if the latter one is.

3.1 Generalized distribution

To bound unmixed terms—i. e., terms of the form $\mathbf{E}[\prod g]$—we use a test distribution which satisfies a certain independence condition. More specifically, over m number of $\{-1,1\}$-valued variables, which we call bits, we would like to have uniform marginals; the all-ones outcome to have strictly positive probability; and finally, conditioned on the outcome of the first bit and any partition $(A,\{2,\ldots,m\}\setminus A)$ of the remaining $m-1$ bits, independence among the variables indexed either by A or by $\{2,\ldots,m\}\setminus A$.

In the following, we show that such a distribution indeed exists. To begin with, we argue that for an odd number m of bits, there is an $(m-1)/2$-wise independent distribution with uniform marginals and weight on the all-ones outcome. This allows one to construct a distribution as follows: for even m, one can have $(m/2-1)$-wise independence, conditioned on the first bit; and for odd m, in effect, one can have independence among bits in at least one set of an arbitrary partition of the second to last bit, conditioned on the first. In particular, the distributions μ defined below satisfies for every $A \uplus A' \subseteq \{2,\ldots,m\}; A, A' \neq \emptyset$,

$$\mathbf{E}_{x_1\sim\mu}\left[\mathbf{E}_\mu\left[\prod_{i\in A} x_i \,\middle|\, x_1\right]\mathbf{E}_\mu\left[\prod_{i\in A'} x_i \,\middle|\, x_1\right]\right] = 0\,. \tag{3.1}$$

Definition 3.2. Let $P_m \subseteq \{-1,1\}^m$ consist of all vectors containing an odd number of -1's as well as the all-ones vector.

Lemma 3.3. *For odd $m \geq 1$, there exists an $(m-1)/2$-wise independent distribution on P_m with uniform marginals and strictly positive weight on the all-ones outcome.*

Proof. Let $f : \{-1,1\}^m \to \mathbb{R}$ be an arbitrary function with Fourier expansion $f(\mathbf{x}) = \sum_{S\subseteq[m]} \hat{f}_S \prod_{i\in S} x_i$. The intention is to show that there is a non-zero function f with support P_m satisfying $(m-1)/2$-wise independence, having uniform marginals, and being non-zero for the all-ones outcome $\mathbf{1}$. From this, one can construct a probability distribution with the desired properties by suitably adding to f a multiple of the $(m-1)$-wise independent distribution $\mathcal{D}$ on vectors of odd parity and scaling appropriately. More formally, without loss of generality, suppose $f(\mathbf{1}) > 0$ or else consider the function $-f$; we can form the claimed distribution by setting $\mu(\mathbf{x}) = f'(\mathbf{x})\left(\sum_{\mathbf{x}} f'(\mathbf{x})\right)^{-1}$ where $f'(\mathbf{x}) = f(\mathbf{x}) + \mathcal{D}(\mathbf{x}) \cdot \min_{\mathbf{x}} f(\mathbf{x})$ which equals $f(\mathbf{x})$ for $\mathbf{x} = \mathbf{1}$, $f(\mathbf{x}) + \min_{\mathbf{x}} f(\mathbf{x})$ for strings $\mathbf{x}$ of odd parity, and 0 otherwise.

We see the Fourier coefficients of f as variables and consider the homogeneous system of linear equations imposed by the balance, independence, and support conditions. First, for every $\mathbf{x} \in \{-1,1\}^m$ of even parity not equal to the all-ones vector, we require $f(\mathbf{x}) = 0$. Second, for every $S \subseteq [m]$ such that

$|S| \leq (m-1)/2$, we require $\hat{f}_S = 0$. Together, this yields a total of $2^m - 1$ equations. As the system is homogeneous and has 2^m variables, there is an infinite number of choices of f satisfying the conditions and in particular there is a non-zero solution.

Next, we argue that a non-zero choice of f implies that $f((1,\ldots,1)) \neq 0$. Consider the polynomial p of the average value of $f(\mathbf{x})$ as a function of the number of -1's in $\mathbf{x}$ times the $\{-1,1\}$-parity of $\mathbf{x}$; formally,

$$p(z) = \mathbf{E}_{\mathbf{x}:|\{i:x_i=-1\}|=z}\left[f(\mathbf{x})\prod_{i\in[m]} x_i\right] = (-1)^z \mathbf{E}_{\mathbf{x}}[f(\mathbf{x})] \,.$$

As f is zero outside P_m, $p(z)$ has $(m-1)/2$ zeroes at $z = 2,4,\ldots,m-1$. Next, we argue that $p(z)$ is of degree at most $(m-1)/2$ and in consequence $p(0) \neq 0$ for any non-zero p since a non-zero degree-d polynomial can have at most d zeroes. To see that p has degree at most $(m-1)/2$ is straightforward as $f(\mathbf{x})\prod_{i\in[m]} x_i$ does not depend on any terms involving $m-(m-1)/2$ or more variables. For the doubtful reader, we complete this argument formally.

By definition of the Fourier expansion, $f(\mathbf{x}) = \sum_{S\subseteq[m]} \hat{f}_S \prod_{i\in S} x_i = \sum_{S\subseteq[m]} \hat{f}_S \prod_{i\in S}(1-2y_i)$ where $y_i = 1$ if $x_i = -1$ and $y_i = 0$ if $x_i = 1$. With the slight change of notation,

$$\begin{aligned} p(z) = \mathbf{E}_{\mathbf{y}:\|\mathbf{y}\|_1=z}\left[\sum_{S\subseteq[m]} \hat{f}_{[m]\setminus S} \prod_{i\in S}(1-2y_i)\right] &= \sum_{S\subseteq[m]} \hat{f}_{[m]\setminus S} \sum_{T\subseteq S} (-2)^{|T|} \mathbf{E}_{\mathbf{y}:\|\mathbf{y}\|_1=z}\left[\prod_{i\in T} y_i\right] \\ &= \sum_{S\subseteq[m]} \hat{f}_{[m]\setminus S} \sum_{T\subseteq S} (-2)^{|T|} \binom{m}{|T|}^{-1} \binom{z}{|T|}. \end{aligned} \tag{3.2}$$

We recognize that if $\hat{f}_S = 0$ for $|S| \leq (m-1)/2$, then $\hat{f}_{[m]\setminus S} = 0$ for $|S| \geq (m-1)/2+1$ and factors of z^k which appear in (3.2) satisfy $k \leq (m-1)/2$. In effect, p has degree at most $(m-1)/2$. □

Corollary 3.4. *For even $m \geq 2$, there exists a distribution $\mu : P_m \to [0,1]$ with uniform marginals such that the all-ones outcome has non-zero probability and, for $\mathbf{x} \sim \mu$, conditioned on x_1, the outcome of $(x_2,\ldots,x_m)$ is $(m/2-1)$-wise independent.*

Proof. The distribution is formed by choosing x_1 uniformly at random from $\{-1,1\}$. If -1, $(x_2,\ldots,x_m)$ is drawn from the uniform distribution over vectors of even parity, which is a $(m-2)$-wise independent distribution with uniform marginals. If x_1 is set to 1, $(x_2,\ldots,x_m)$ is drawn according to the distribution which exists due to Lemma 3.3 and has the desired properties. □

Corollary 3.5. *For $m \geq 4$ there exists a distribution μ_m on P_m satisfying (3.1).*

Proof. For even m, the claim follows directly from Theorem 3.4. Consider an arbitrary odd $m \geq 5$. Again, we form a distribution by first sampling x_1 uniformly at random from $\{-1,1\}$. If $x_1 = -1$, draw $(x_2,\ldots,x_m)$ uniformly at random from vectors of even parity. If $x_1 = 1$, set $(x_2,\ldots,x_m)$ according to a distribution which exists by Theorem 3.4. Since the distribution has uniform marginals, we claim that—conditioned on x_1—for any $A \uplus A' \subseteq \{2,\ldots,m\}; A, A' \neq \emptyset$, either the variables $(x_i)_{i\in A}$ or the variables $(x_i)_{i\in A'}$ are independent and (3.1) is satisfied. To justify this; by Theorem 3.4, the variables $(x_3,\ldots,x_m)$ are $((m-1)/2-1)$-wise independent conditioned on x_2 and $x_1 = 1$. Hence, either

$$1 \leq \min\{|A|,|A'|\} \leq ((m-1)/2-1)$$

and one of the two sets have independent variables and satisfy (3.1), or

$$|A| = |A'| = (m-1)/2 \geq 2\,.$$

Without loss of generality, $2 \in A$ and $1 \leq |A \setminus \{2\}| < (m-1)/2$ and hence the bits indexed by $A \setminus \{2\}$ are independent conditioned on x_2 and $x_1 = 1$. In consequence the bits indexed by A are independent conditioned on $x_1 = 1$ and (3.1) is satisfied.

On the other hand, if x_1 is set to -1, the bits $(x_2, \ldots, x_m)$ are drawn from an $(m-2)$-wise independent distribution conditioned on $x_1 = -1$ and since $|A|, |A'| \leq m-2$, the bits are set independently and (3.1) is satisfied. □

3.2 Generalized protocol

Let $\mathcal{D}$ be the $(m-1)$-wise independent distribution which draws uniformly at random from Odd Parity, and let $\mathcal{E} = \mu_m$ be the distribution defined in the preceding subsection. The following PCP test is the simple analogue of the arity-four case.

1. Pick a random vertex $u \in U$ and a random neighbor $v \in V$. Sample $\pi = \pi^{\{u,v\}}$ as defined by the SMOOTH LABEL COVER instance and let $\bar{\pi}$ be an arbitrary bijection $L \leftrightarrow L$ such that for every $i, i' \in K$ and $r \in [d]$, $\pi(i,r) = i'$ iff $\exists_{r' \in [d]} \bar{\pi}(i,r) = (i', r')$.

2. Sample random folding constants $a, b \sim \{0,1\}$. Define $f_a(\mathbf{x}) = a \oplus f^u(a \oplus \mathbf{x})$ and $g_b(\mathbf{y}) = b \oplus g^v(b \oplus \mathbf{y} \circ \bar{\pi})$.

3. For each $i \in K$, independently choose x_i uniformly at random from $\{0,1\}$. For each $j \in L$, independently sample $(x_{\pi(j)}, y_j^{(2)}, \ldots, y_j^{(m)})$ conditioned on $x_{\pi(j)}$ from $\mathcal{D}$ with probability $\bar{\delta}$ and otherwise $\mathcal{E}$.

4. Accept iff $\left(f_a(\mathbf{x}), g_b(\mathbf{y}^{(2)}), \ldots, g_b(\mathbf{y}^{(m)})\right) \in P$.

The completeness and soundness claims are as follows.

Lemma 3.6. *The protocol has completeness 1. Said equivalently, if* $\mathrm{Val}(I) = 1$*, then* $\mathrm{Val}(\mathrm{R}_P(I)) = 1$.

Proof. The distributions $\mathcal{D}$ and μ_m both have support $P_m \subseteq P$ and hence setting functions to their dictators corresponding to a satisfying LABEL COVER solution is always accepted. □

Proposition 3.7. *The protocol has soundness* $2^{-m}|P| + \varepsilon_{CSP}$*. More specifically, if* $\mathrm{Val}(I) \leq \varepsilon_{LC} = \varepsilon_{LC}(\varepsilon_{CSP})$*, then* $\mathrm{Val}(\mathrm{R}_P(I)) \leq 2^{-m}|P| + \varepsilon_{CSP}$ *where* $\varepsilon_{CSP}(\varepsilon_{LC}) \to 0$ *as* $\varepsilon_{LC} \to 0$.

Constants Let $\alpha = \alpha(m)$ be the smallest strictly positive probability of any outcome of μ_m. Let $\delta \leq 2^{-2m-8}\varepsilon_{\mathrm{CSP}}^2$; define

$$\rho_0 \triangleq \sqrt{1/2 + 1/2(1-\alpha^2\delta^2/2)^2}$$

and choose $\gamma > 0$ sufficiently close to 0 such that $\sup_k \rho_0^k(1-\bar{\gamma}^k) \leq 2^{-m-4}\varepsilon_{\mathrm{CSP}}/m$.[3] Again, define

$$\rho_1 \triangleq \sqrt{1-\gamma^{m-1}}$$

and choose $\eta > 0$ sufficiently close to 0 such that $\sup_k \rho_1^k(1-\bar{\eta}^k) \leq 2^{-m-4}\varepsilon_{\mathrm{CSP}}/m$. Choose $d_2,\dots,d_m$ such that

$$2\bar{\gamma}^{d_2} \leq 2^{-m-4}/m \quad \text{and} \quad 2\bar{\gamma}^{d_r}(2r-m)^{\sum_2^{r-1} d_t/2} \leq 2^{-m-4}/m\,.$$

Finally, choose the smoothness parameters $J = d_m$, and $\xi \leq (2m)^{-md_m}2^{-2m-10}\varepsilon_{\mathrm{CSP}}^2$.

3.2.1 Soundness of the generalized protocol

Notation Redefine the generalized distributions as follows. The primary difference to the width-four case is that we use different degree constants $d_2 \leq \cdots \leq d_m$ in our proofs when bounding unmixed $\mathbf{E}[\prod g]$ terms.

$$\mathcal{T}_0 = \bar{\delta}\mathcal{D} + \delta\mathcal{E}, \quad \mathcal{T}'_0 = \mathcal{T}_0^{d\text{-proj-}1^{\otimes K}}, \quad \mathcal{T}'_1 = \left(\mathrm{T}^{2,\dots,m}_{\bar{\gamma},\dots,\bar{\gamma}}\,\mathcal{T}_0^{d\text{-proj-}1}\right)^{\otimes K},$$

$$\mathcal{T}'_2 = \left(\mathrm{T}^{2,\dots,m}_{\bar{\gamma},\dots,\bar{\gamma}}\left(\mathrm{T}^1_{\bar{\eta}}\,\mathcal{T}_0\right)^{d\text{-proj-}1}\right)^{\otimes K}, \quad \mathcal{T}_3 = \mathrm{T}^1_{\bar{\eta}}\,\mathrm{T}^{2,\dots,m}_{\bar{\gamma},\dots,\bar{\gamma}}\,\mathcal{T}_0, \quad \mathcal{T}'_3 = \mathcal{T}_3^{d\text{-proj-}1^{\otimes K}}, \qquad \mathcal{T}''_3 = \mathcal{T}_3^{\otimes L}.$$

The test distribution of the protocol corresponds to $\mathcal{T}'_0$. Intuitively, $\mathcal{T}'_1$ is the distribution where projected noise is applied to $\mathbf{y}^{(2)},\dots,\mathbf{y}^{(m)}$, i. e., all coordinates which share projection are changed by noise simultaneously. $\mathcal{T}'_2$ is the same distribution but with noise applied also to $\mathbf{x}$. We note that projected and non-projected (independent) noise are the same for $\mathbf{x}$ as it is defined on the smaller table. $\mathcal{T}'_3$ is the distribution all strings—$\mathbf{x},\mathbf{y}^{(2)},\dots,\mathbf{y}^{(m)}$—all have independent noise. Finally, $\mathcal{T}''_3$ is the same as $\mathcal{T}'_3$, we have noise for all strings, but $\mathbf{x}$ is defined on $\{-1,1\}^L$ and for each $j \in L$, the tuple $(x_j, y_j^{(2)},\dots,y_j^{(m)})$ is drawn independently; we note that this distribution will only be used for analyzing terms where $\mathbf{x}$ does not appear and equivalently one can see the strings $(\mathbf{y}^{(t)})_t$ as being drawn independent of $\mathbf{x}$. As for the width-four case, we define $f = \mathbf{E}_a[f_a]$ and $g = \mathbf{E}_b[g_b]$. Let the queries be $q_1 = f(\mathbf{x}), q_2 = g(\mathbf{y}^{(2)}),\dots,q_m = g(\mathbf{y}^{(m)})$. For an arbitrary $\Gamma \neq \emptyset$ and distribution $\mathcal{R}$, let us denote by $\psi_\Gamma(\mathcal{R}) = \mathbf{E}_{E,\mathcal{R}}[\chi_\Gamma(\mathbf{q})]$. Conceptually, we refer again to these terms as unmixed terms—$\mathbf{E}[\prod g]$—or mixed terms—$\mathbf{E}[f\prod g]$—for zero or more functions g.

As in preceding proofs, the aim is to show the following four generalized propositions from which the soundness follows.

Proposition 3.8. *$\psi_\Gamma(\mathcal{T}'_0) = 0$ for $\emptyset \neq \Gamma \subseteq [m], |\Gamma \cap \{2,\dots,m\}| \leq 1$.*

Proof. As the test distribution has uniform marginals, $\psi_{\{t\}}(\mathcal{T}'_0) = \mathbf{E}_{E,\mathcal{T}'_0}[q_t]$ which equals $\mathbf{E}_E[\mathbf{E}[f]]$ or $\mathbf{E}_E[\mathbf{E}[g]]$, both of which are 0 due to folding. Suppose $\Gamma = \{1,t\}$. Then $\psi_\Gamma(\mathcal{T}'_0) = \mathbf{E}_{E,\mathcal{T}'_0}[fg] = \mathbf{E}[f]\,\mathbf{E}[g] = 0$ since $\mathbf{y}^{(t)}$ is uniform and independent of $\mathbf{x}$ by Lemma 2.4, subsequently folding yields expectation 0. □

We note that the constants ρ_0 and ρ_1 appearing in the proposition are correlation bounds appearing in the proofs and are bounded away from 1 depending only on δ, γ, and m.

[3]In particular, setting $\bar{\gamma} = \rho_0^{2^{-m-4}\varepsilon_{\mathrm{CSP}}/m}$ suffices.

Proposition 3.9.

$$\left|\psi_\Gamma(\mathcal{T}_0') - \psi_\Gamma(\mathcal{T}_3')\right| \le (m-1)\sup_{k\ge 0}\rho_0^k(1-\bar{\gamma}^k) + \sup_{k\ge 0}\rho_1^k(1-\bar{\eta}^k) + 2m\sqrt{\xi} + 2m\bar{\gamma}^J \le \varepsilon_{CSP}/2^{-m-2}$$

for any $\Gamma \subseteq [m]$.

Proof. Appears in Section 3.4. □

Proposition 3.10.

$$\left|\psi_\Gamma(\mathcal{T}_3')\right| \le 2\bar{\gamma}^{d_2} + 2\sum_{r=3}^{m}\bar{\gamma}^{d_r}(2r-5)^{\sum_{t=2}^{r-1} d_t/2} + 2(m-1)(2m-5)^{\sum_{t=2}^{m} d_t/2}\sqrt{\xi} + \sqrt{\delta} \le \varepsilon_{CSP}/2^{-m-2}$$

for $1 \notin \Gamma \subseteq [m], |\Gamma| \ge 2$.

Proof. Appears in Section 3.5. □

Proposition 3.11. $|\psi_\Gamma(\mathcal{T}_3')| \le 2^{2m}\sqrt{\gamma^{-1}\,\mathbf{E}_E\left[\sum_{(i,j)\in\pi}\mathrm{Inf}_i^{(\bar{\eta})}(f)\mathrm{Inf}_j^{(\bar{\gamma})}(g)\right]}$ *for* $1 \in \Gamma \subseteq [m], |\Gamma| \ge 3$.

Proof. Appears in Section 3.7. □

3.3 Generalized correlation bounds

In this subsection, we establish generalized bounds on the correlation between strings in our test distribution. For preliminaries, we refer to the reader to Section 2.3.

The correlation bounds we aim to establish for the test distribution are the following analogues of the width-four case. The first lemma shows that for our test distribution $\mathcal{T}_0$, the correlation between arguments to g functions are bounded away from 1 independent of d. This in turn will enable us to introduce projected noise for g functions.

Lemma 3.12. *For any* $2 \le r \le m$,

$$\rho\left(\Omega_1 \times \Omega_{-1,-r}^d, \Omega_r^d;\ \mathcal{T}_0^{d\text{-proj-}1}\right) \le \sqrt{\frac{1}{2} + \frac{1}{2}\left(1 - \frac{\alpha^2\delta^2}{2}\right)^2},$$

where $\alpha = \alpha(m)$ *is the smallest strictly positive probability of any outcome for the distribution* $\mathcal{E} = \mu_m$ *as defined in Section 3.1.*

Proof. Lemma 2.4 implies that Ω_m^d is independent of Ω_1. Applying Theorem 2.19 with $A = \{1\}, B = \{2,\dots,m\}\setminus\{r\}, C = \{r\}$, we get

$$\rho\left(\Omega_1 \times \Omega_{-1,-r}^d, \Omega_r^d;\ \mathcal{T}_0^{d\text{-proj-}1}\right) \le \rho(\Omega_{-1,-r}, \Omega_r \mid \Omega_1;\ \mathcal{T}_0)$$

which by definition equals $\mathbf{E}_{\omega_1}\left[\rho(\Omega_{-1,-r}, \Omega_r;\ \mathcal{T}_0 \mid \omega_1)^2\right]^{1/2}$.

Switching to $\{0,1\}$ notation again, to establish that the considered correlation is bounded away from 1, it suffices that the conditioned correlation is bounded away from 1 for at least one of the cases $\omega_1 = 0$ and $\omega_1 = 1$. This is precisely what we do, we bound the latter by 1 and find a smaller bound for the case $\omega_1 = 0$.

As for the width-four case, we employ Lemma 2.16. The bipartite graph in question has left vertices $\Omega_{-1,-r}$, right vertices Ω_r, and an edge $\{\vec{\omega}_{-1,-r}, \omega_r\}$ whenever $(0, \vec{\omega}_{-1,-r}, \omega_r)$ has strictly positive probability; Lemma 2.16 states that if this graph connected, the correlation is bounded away from 1. To be specific, the correlation is at most $1-\alpha^2/2$ where α is the smallest strictly positive probability of any outcome. This graph is indeed connected. From $\omega_r = 1$, $\mathcal{D}$ connects to any $\vec{\omega}_{-1,-r}$ of even parity while from $\omega_r = 0$, $\mathcal{D}$ connects to any $\vec{\omega}_{-1,-r}$ of odd parity as well as the even-parity outcome $\vec{\omega}_{-1,-r} = \vec{0}$ due to the all-zero outcome of $\mathcal{E}$.

Defined above, let $\alpha = \alpha(m)$ be the minimum non-zero probability of any outcome of $\mathcal{E}$. The minimum strictly positive probability of any outcome for the distribution $\mathcal{T}_0 = \bar{\delta}\mathcal{D} + \delta\mathcal{E}$ conditioned on $\omega_1 = 0$ is then given by $\mathcal{E}$ as δ is close to 0, i. e., the minimum probability of any atom of $\bar{\delta}\mathcal{D} + \delta\mathcal{E}$ is $\alpha(m)\delta$. This implies that $\rho(\Omega_{-1,-r}, \Omega_r;\ \mathcal{T}_0 \mid \omega_1 = 0) \le 1 - \alpha(m)^2\delta^2/2$ and, as desired,

$$\rho\left(\Omega_1 \times \Omega^d_{-1,-r}, \Omega^d_r;\ \mathcal{T}_0^{d\text{-proj-1}}\right) \le \sqrt{\frac{1}{2}\cdot 1^2 + \frac{1}{2}\left(1 - \frac{\alpha^2\delta^2}{2}\right)^2}. \qquad \square$$

Along similar lines to the argument in Section 2.3, the second lemma implies that after we have introduced projected noise for all g functions, the argument to f has correlation bounded away from 1 independent of d. Again, this enables us to introduce noise for f.

Lemma 3.13.

$$\rho\left(\Omega_1, \Omega_2^d \times \cdots \times \Omega_m^d;\ \mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\, \mathcal{T}_0^{d\text{-proj-1}}\right) \le \sqrt{1-\gamma^{m-1}}\,.$$

Proof. We recall that the considered distribution is $\mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\, \mathcal{T}_0^{d\text{-proj-1}}$. With probability $\prod_{t=2}^m \gamma$, the outcome of $\Omega_2^d \times \cdots \times \Omega_m^d$ is independent of Ω_1 and the correlation is 0. Denote this event by A and let the correlations of the two possibilities be, respectively, $\rho_{\mathrm{A}} = 0$ and $\rho_{\bar{\mathrm{A}}} \le 1$. Then,

$$\rho\left(\Omega_1, \Omega_2^d \times \cdots \times \Omega_m^d;\ \mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\, \mathcal{T}_0^{d\text{-proj-1}}\right) \le \sqrt{\mathbf{P}(\mathrm{A})\,\rho_{\mathrm{A}}^2 + \mathbf{P}(\bar{\mathrm{A}})\,\rho_{\bar{\mathrm{A}}}^2}$$
$$\le \sqrt{\prod_{t=2}^m \gamma \cdot 0^2 + \left(1 - \prod_{t=2}^m \gamma\right)\cdot 1^2} = \sqrt{1-\gamma^{m-1}}\,. \qquad \square$$

The third and final lemma is used to show that a product of g-functions is always small if we do not have projections. This will be the final step when we bound terms of the form $\mathbf{E}[\prod g]$ after we have argued that the product behaves roughly as though there were unique projections.

Lemma 3.14.

$$\rho\left(\Omega_r, \Omega_{-1,-r};\ \mathcal{T}_0\right) \le \sqrt{\delta}\,.$$

Proof. With probability $\bar{\delta}$, Ω_r is drawn from $\mathcal{D}$ in which case it is independent of $\Omega_{-1,-r}$, yielding a correlation of 0. In the other event, the correlation is bounded by 1. Hence, using Theorem 2.18,

$$\rho\left(\Omega_r, \Omega_{-1,-r};\ \mathcal{T}_3^{d\text{-proj-1}}\right) \le \sqrt{\bar{\delta}\cdot 0^2 + \delta \cdot 1^2} \le \sqrt{\delta}\,. \qquad \square$$

3.4 Generalized noise introduction

To introduce noise, we prove the following generalizations of Lemmas 2.22 to 2.24.

Lemma 3.15. *Let $\Gamma \subseteq [m]$, $|\Gamma| \geq 2$ and define $\rho_0 = \sqrt{1/2 + 1/2(1-\alpha^2\delta^2/2)^2}$. Then,*

$$\left|\mathbf{E}\left[\psi_\Gamma(\mathcal{T}'_0) - \psi_\Gamma(\mathcal{T}'_1)\right]\right| \leq \sum_{t=2}^{m} \sup_k \rho_0^k (1-\bar{\gamma}^k) \leq (m-1) \sup_k \rho_0^k (1-\bar{\gamma}^k)\,.$$

Proof. Proved in Section 3.4. □

Lemma 3.16. *Let $\Gamma \subseteq [m]$, $|\Gamma| \geq 2$ and define $\rho_1 = \sqrt{1-\gamma^{m-1}}$. Then,*

$$\left|\mathbf{E}\left[\psi_\Gamma(\mathcal{T}'_1) - \psi_\Gamma(\mathcal{T}'_2)\right]\right| \leq \sup_k \rho_1^k (1-\bar{\eta}^k)\,.$$

Proof. Proved in Section 3.4. □

Lemma 3.17. *Let $\Gamma \subseteq [m]$, $|\Gamma| \geq 2$. Then,*

$$\left|\mathbf{E}\left[\psi_\Gamma(\mathcal{T}'_2) - \psi_\Gamma(\mathcal{T}'_3)\right]\right| \leq 2m\sqrt{\xi} + 2m\bar{\gamma}^J\,.$$

Proof. Proved in Section 3.4. □

Proof of Proposition 3.9. Follows directly from Lemmas 3.15, 2.23, and 3.17 by summing the respective differences. □

Proof of Lemma 3.15: Introducing projected noise for g functions We define $\Omega'_1 = \Omega_1, \Omega'_t = \Omega_t^d$ for $t \in \{2,\dots,m\}$, let $\overline{\mathbf{y}^{(t)}}$ be the lifted version of $\mathbf{y}^{(t)}$, and $\overline{\mathcal{T}}$ the lifted analogue of a distribution $\mathcal{T}$. Additionally, as we wish to claim simultaneously the lemmas for all $\Gamma \subseteq [m], |\Gamma| \geq 2$, let h_A for a subset $A \subseteq [m]$ denote

$$f^{[1\in A]} \prod_{t \in A\setminus 1} \overline{g^{(t)}}\,.$$

Proof. Let

$$\mathcal{D}_1 = \overline{\mathcal{T}_0^{d\text{-proj-}1}}, \quad \mathcal{D}_r = \mathbf{T}_{\bar{\gamma}}^{(r)} \mathcal{D}_{r-1}, \qquad r \in \{2,\dots,m\}\,.$$

We note that $\mathcal{D}_m^K$ is indeed the lifted analogue of $\mathcal{T}'_1$. Consequently, the lemma is proved by bounding the respective differences of expectations $\left|\psi_\Gamma(\mathcal{D}_r^K) - \psi_\Gamma(\mathcal{D}_{r-1}^K)\right|$ for $r \in \{2,\dots,m\}$.

Working out the notation, for $r \in \{2,\dots,m\}$,

$$\begin{aligned}
&\left|\psi_\Gamma(\mathcal{D}_{r-1}^K) - \psi_\Gamma(\mathcal{D}_r^K)\right| = \left|\mathbf{E}_{\mathcal{D}_{r-1}^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right] - \mathbf{E}_{\mathcal{D}_r^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right]\right| \\
&= \left|\mathbf{E}_{\mathcal{D}_{r-1}^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right] - \mathbf{E}_{(\mathbf{T}_{\bar{\gamma}}^{(r)} \mathcal{D}_{r-1})^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right]\right| = \left|\mathbf{E}_{\mathcal{D}_{r-1}^K}\left[\overline{g^{(r)}} h_{\Gamma\setminus r}\right] - \mathbf{E}_{\mathcal{D}_{r-1}^K}\left[(\mathbf{T}_{\bar{\gamma}} \overline{g^{(r)}}) h_{\Gamma\setminus r}\right]\right|.
\end{aligned} \tag{3.3}$$

This is the setting of Theorem 2.25.

By Lemma 3.12 and symmetry, the correlation $\rho(\Omega_1 \times \prod_{t\neq 1,r}\Omega_t^d, \Omega_r^d; \mathcal{T}_0^{d\text{-proj-1}})$, which equals $\rho(\Omega_1 \times \prod_{t\neq 1,r}\Omega'_t, \Omega'_r; \mathcal{D}_1)$, is bounded by

$$\rho_0 \triangleq \sqrt{1/2 + 1/2(1-\alpha^2\delta^2/2)^2}\,.$$

As noise can only decrease correlation, the same bound holds for $\mathcal{D}_{r-1}$. Similarly, this is a bound on any subset of sample spaces in the case $\Gamma \neq [m]$.

For $r \in \{2,\ldots,m\}$, if $r \notin \Gamma$, the difference (3.3) is 0. Otherwise, we bound using Theorem 2.25 with $\rho \leq \rho_0$; that is, (3.3) $\leq \sup_k \rho_0^k(1-\bar{\gamma}^k)$. In conclusion,

$$\left|\psi_\Gamma(\mathcal{T}'_0) - \psi_\Gamma(\mathcal{T}'_1)\right| = \left|\psi_\Gamma(\mathcal{D}_1^K) - \psi_\Gamma(\mathcal{D}_m^k)\right| \leq \sum_{t=2}^{m}\left|\psi_\Gamma(\mathcal{D}_{t-1}^K) - \psi_\Gamma(\mathcal{D}_t^K)\right| \leq (m-1)\sup_k \rho_0^k(1-\bar{\gamma}^k)\,. \quad \square$$

Proof of Lemma 3.16: Introducing noise for the f function

Proof. By Lemma 3.13,

$$\rho(\Omega_1, \Omega_2^d \times \cdots \times \Omega_m^d; \mathcal{T}_1^{d\text{-proj-1}}) \leq \rho_1 \triangleq \sqrt{1-\gamma^{m-1}}\,.$$

The same bound holds for $\rho(\Omega_1, \prod_{t\in\Gamma\setminus 1}\Omega'_t; \overline{\mathcal{T}_1^{d\text{-proj-1}}})$. Hence, using Theorem 2.25 again,

$$\begin{aligned}\left|\psi_\Gamma(\mathcal{T}'_1) - \psi_\Gamma(\mathcal{T}'_2)\right| &= \left|\mathbf{E}_{\left(\mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\mathcal{T}_0^{d\text{-proj-1}}\right)^K}\left[fh_{\Gamma\setminus 1}\right] - \mathbf{E}_{\left(\mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\left(\mathrm{T}^1_{\bar{\eta}}\,\mathcal{T}_0\right)^{d\text{-proj-1}}\right)^K}\left[fh_{\Gamma\setminus r}\right]\right| \\ &= \left|\mathbf{E}_{\left(\mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\mathcal{T}_0^{d\text{-proj-1}}\right)^K}\left[fh_{\Gamma\setminus r}\right] - \mathbf{E}_{\left(\mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\mathcal{T}_0^{d\text{-proj-1}}\right)^K}\left[(\mathrm{T}_{\bar{\eta}}\,f)h_{\Gamma\setminus r}\right]\right| \leq \sup_k \rho_1^k(1-\bar{\eta}^k)\,. \quad \square\end{aligned}$$

Proof of Lemma 3.17: From projected noise to independent noise

Proof. Just as for the width-four case, the lemma follows from repeated application of Theorem 2.21. To utilize the theorem, we again unravel the definition of g from the protocol:

$$g(\mathbf{y}) \triangleq \mathbf{E}_{b\sim\{0,1\}}[b \oplus g^v(b \oplus \mathbf{y}\circ\bar{\pi})]\,,$$

where $\bar{\pi}$ is an arbitrary bijection consistent with π. Define

$$g'^v(\mathbf{y}) = \mathbf{E}_{b\sim\{0,1\}}[b \oplus g^v(b \oplus \mathbf{y})]$$

and let $\mathcal{T}'^{\bar{\pi}}_2$ and $\mathcal{T}'^{\bar{\pi}}_3$ permute the coordinates of g' via $\bar{\pi}$. Also, for $F \subseteq [m]$, let

$$h_F^v = f^{[1\in F]}\prod_{t\in F\setminus 1} g'^v(\mathbf{y}^{(t)})\,.$$

We recall the definition of the following two distributions,

$$\mathcal{T}'_2 \triangleq \left(\mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\left(\mathrm{T}^1_{\bar{\eta}}\,\mathcal{T}_0\right)^{d\text{-proj-1}}\right)^K \quad \text{and} \quad \mathcal{T}'_3 \triangleq \left(\left(\mathrm{T}^{2,\ldots,m}_{\bar{\gamma},\ldots,\bar{\gamma}}\mathrm{T}^1_{\bar{\eta}}\,\mathcal{T}_0\right)^{d\text{-proj-1}}\right)^K.$$

The target difference equals

$$\left|\psi_\Gamma(\mathcal{T}_2') - \psi_\Gamma(\mathcal{T}_3')\right| = \left|\mathbf{E}_{u,v,\mathcal{T}_2'^{\pi\{u,v\}}}\left[f h^v_{\Gamma\setminus 1}\right] - \mathbf{E}_{u,v,\mathcal{T}_3'^{\pi\{u,v\}}}\left[f h^v_{\Gamma\setminus 1}\right]\right|. \tag{3.4}$$

We apply Theorem 2.21 up to $m-1$ times, once for each of the coordinates $2,\dots,m$ which appear in Γ. Introduce $\mathcal{R}_t$ as follows,

$$\mathcal{R}_t \triangleq \left(\mathbf{T}^{t+1,\dots,m}_{\bar{\gamma},\dots,\bar{\gamma}}\left(\mathbf{T}^{1,2,\dots,t}_{\bar{\eta},\bar{\gamma},\dots,\bar{\gamma}}\mathcal{T}_0\right)^{d\text{-proj-}1}\right)^K,$$

where we note that $\mathcal{R}_1$ corresponds to $\mathcal{T}_2'$ and $\mathcal{R}_m$ to $\mathcal{T}_3'$.

Formally, when applying it to coordinate $t \in \Gamma\setminus\{1\}$, we have $A=\{1\}, B=\{t\}, C=\Gamma\setminus\{1,t\}, \gamma=\gamma$, and the distributions equal $\mathcal{P}=\mathcal{R}_{t-1}$ and $\mathcal{R}=\mathcal{R}_t$. The respective differences in expectation hence yield a bound on the difference in expectation between $\mathcal{T}_2'=\mathcal{R}_2$ and $\mathcal{T}_3'=\mathcal{R}_m$. According to the theorem, the difference in expectation for coordinate t is bounded by $2\sqrt{\xi}+2\bar{\gamma}^J$. In effect,

$$(3.4) \le \sum_{t=2}^{m}(2\sqrt{\xi}+2\bar{\gamma}^J) \le 2m\sqrt{\xi}+2m\bar{\gamma}^J. \qquad \square$$

3.5 Proof of Proposition 2.7: Bounding $\mathbf{E}_{\mathcal{T}_3'}[\prod g]$, general case

We first limit ourselves to the hardest case: $\Gamma=\{2,\dots,m\}$ which corresponds to bounding

$$\mathbf{E}_{\pi,\mathcal{T}_3'}\left[g(\mathbf{y}^{(2)})\cdots g(\mathbf{y}^{(m)})\right],$$

and comment briefly on the other cases following the proof of the hardest case.

In the following, g denotes g'^v as defined in the previous subsection.

Noised to low-degree functions

Lemma 3.18. *Let $\mathcal{D}=\mathcal{T}_3'$ or $\mathcal{T}_3''$. Then,*

$$\left|\mathbf{E}_{\mathcal{D}}\left[g(\mathbf{y}^{(2)})\cdots g(\mathbf{y}^{(m)})\right] - \mathbf{E}_{\mathcal{D}}\left[g^{\le d_2}(\mathbf{y}^{(2)})\cdots g^{\le d_m}(\mathbf{y}^{(m)})\right]\right| \le \bar{\gamma}^{d_2} + \sum_{r=3}^{m}\bar{\gamma}^{d_r}(2r-5)^{\sum_{t=2}^{r-1}d_t/2}.$$

Proof. We considering the effect of removing the high-degree components one at a time and rewrite the treated difference,

$$\left|\sum_{r=2}^{m}\mathbf{E}_{\mathcal{D}}\left[\left(\prod_{t=2}^{r-1}g^{\le d_t}(\mathbf{y}^{(t)})\right)g^{>d_r}(\mathbf{y}^{(r)})\left(\prod_{t=r+1}^{m}g(\mathbf{y}^{(t)})\right)\right]\right|. \tag{3.5}$$

We bound in absolute value these terms separately. For $r=2$, Cauchy-Schwarz gives a bound of

$$\left\|g^{>d_2}(\mathbf{y}^{(2)})\right\|_2\cdot\left\|\prod_{t=3}^{m}g(\mathbf{y}^{(t)})\right\|_2 \le \bar{\gamma}^{d_2}.$$

For any term with $3 \le r \le m$, Hölder's inequality yields the bound

$$\prod_{t=2}^{r-1} \left\| g^{\le d_t}(\mathbf{y}^{(t)}) \right\|_{2r-4} \left\| g^{> d_r}(\mathbf{y}^{(r)}) \right\|_2 \prod_{t=r+1}^{m} \left\| g(\mathbf{y}^{(t)}) \right\|_\infty . \tag{3.6}$$

Regarding the first kind of factors, by Lemma 1.7,

$$\left\| g^{\le d_t}(\mathbf{y}^{(t)}) \right\|_{2r-4} \le (2r-5)^{d_t/2} \left\| g^{\le d_t}(\mathbf{y}^{(t)}) \right\|_2 \le (2r-5)^{d_t/2} .$$

Again, $\left\| g^{>d_r}(\mathbf{y}^{(r)}) \right\|_2 \le \bar{\gamma}^{d_r}$, while $\left\| g(\mathbf{y}^{(t)}) \right\|_\infty \le 1$. Hence, (3.6) is bounded by

$$\bar{\gamma}^{d_r} \prod_{t=2}^{r-1} (2r-5)^{d_t/2} = \bar{\gamma}^{d_r} (2r-5)^{\Sigma_2^{r-1} d_t/2}$$

and consequently (3.5) by

$$\bar{\gamma}^{d_2} + \sum_{r=3}^{m} \bar{\gamma}^{d_r} (2r-5)^{\Sigma_2^{r-1} d_t/2} . \qquad \square$$

Smooth low-degree to shattered functions via hypercontractivity We recall the term *shattered* denoting functions where every non-zero f_S satisfies $|\pi(S)| = |S|$. The goal is to show that for smooth projections, low-degree functions are essentially shattered.

Lemma 3.19. *Let $\mathcal{D} = \mathcal{T}_3'$ or $\mathcal{T}_3''$. Then,*

$$\left| \mathbf{E}_{\pi,\mathcal{D}^{\bar{\pi}}} \left[\prod_{t=2}^{m} g_t^{\le d_t}(\mathbf{y}^{(t)}) \right] - \mathbf{E}_{\pi,\mathcal{D}^{\bar{\pi}}} \left[\prod_{t=2}^{m} g_t^{\not\pi \le d_t}(\mathbf{y}^{(t)}) \right] \right| \le (2m-5)^{\Sigma_{t=2}^{m} d_t/2} (m-1) \sqrt{\xi} .$$

Proof. It suffices to bound for $2 \le r \le m$ the term

$$\left| \left(g^{\le d_r}(\mathbf{y}^{(r)}) - g^{\not\pi \le d_r}(\mathbf{y}^{(r)}) \right) \mathbf{E} \left[\left(\prod_{t=2}^{r-1} g^{\not\pi \le d_t}(\mathbf{y}^{(t)}) \right) \left(\prod_{t=r+1}^{m} g^{\le d_t}(\mathbf{y}^{(t)}) \right) \right] \right|$$

which via Hölder's is bounded by

$$\left\| g^{\le d_r}(\mathbf{y}^{(r)}) - g^{\not\pi \le d_r}(\mathbf{y}^{(r)}) \right\|_2 \prod_{t=2}^{r-1} \left\| g^{\not\pi \le d_t}(\mathbf{y}^{(t)}) \right\|_{2m-4} \prod_{t=r+1}^{m} \left\| g^{\le d_t}(\mathbf{y}^{(t)}) \right\|_{2m-4} .$$

By definition, $d_r \le J$ and hence Lemma 2.20 bounds the first factor by $\sqrt{\xi} \left\| g^{\le d_r}(\mathbf{y}^{(r)}) \right\|_2$ while Lemma 1.7 bounds the respective factors by $(2m-5)^{d_t/2} \|g\|_2$. As $2m-5 \ge 1$, this yields the desired bound, summing over the $m-1$ terms. $\square$

Shattered functions to independent coordinates

Lemma 3.20. *For shattered functions, $\mathcal{T}'_3$ and $\mathcal{T}''_3$ yield identical expectations, i. e., coordinates in L may for shattered functions be drawn independently. That is,*

$$\mathbf{E}_{\mathcal{T}'_3}\left[\prod_{t=2}^{m} g_t^{\rightarrowtail\pi\le d_t}\right] = \mathbf{E}_{\mathcal{T}''_3}\left[\prod_{t=2}^{m} g_t^{\rightarrowtail\pi\le d_t}\right].$$

Proof. Consider the Efron-Stein decompositions of the functions in the two terms. Similar to the width-four case, we argue that each term, indexed by $(S_t)_{t=2}^{m}$, coincides in expectation for the two distributions. First we claim that terms for which $|\pi(\bigcup S_t)| < |\bigcup S_t|$ evaluate to 0 due to our choice of test distribution. To see this, consider an $i \in K$ for which there are at least two different $j \neq j' \in \bigcup S_t$ projecting to i and let T and T' be the indices t for which $j \in S_t$ or $j' \in S_t$, respectively. Since we are considering shattered functions, the sets $(S_t)_t$ are shattered for non-zero terms and so T and T' are disjoint and non-empty. Since μ_m—and in extension $\bar{\delta}\mathcal{D} + \delta\mu_m$—satisfies (3.1), a factor of such terms has expectation zero.

Given $|\pi(\bigcup S_t)| = |\bigcup S_t|$, the argument is identical to that of the width-four case, Lemma 2.29. Namely, for each i, there is at most one j such that $(y_j^{(t)})_t$ depends on x_i and hence the distributions $\mathcal{T}'_3$ and $\mathcal{T}''_3$ coincide. □

Putting it together

Proof of Proposition 3.10. Using the three preceding lemmas and $d_t \le J$ for all t,

$$\begin{aligned}&\left|\mathbf{E}_{\mathcal{T}'_3}\left[g(\mathbf{y}^{(2)})\cdots g(\mathbf{y}^{(m)})\right] - \mathbf{E}_{\mathcal{T}''_3}\left[g^{\rightarrowtail\pi\le d_2}(\mathbf{y}^{(2)})\cdots g^{\rightarrowtail\pi\le d_m}(\mathbf{y}^{(m)})\right]\right| \\ &\le \bar{\gamma}^{d_2} + \sum_{r=3}^{m} \bar{\gamma}^{d_r}(2r-5)^{\sum_{t=2}^{r-1} d_t/2} + (m-1)(2m-5)^{\sum_{t=2}^{m} d_t/2}\sqrt{\xi}\,.\end{aligned}$$

and similarly,

$$\begin{aligned}&\left|\mathbf{E}_{\mathcal{T}''_3}\left[g(\mathbf{y}^{(2)})\cdots g(\mathbf{y}^{(m)})\right] - \mathbf{E}_{\mathcal{T}''_3}\left[g^{\rightarrowtail\pi\le d_2}(\mathbf{y}^{(2)})\cdots g^{\rightarrowtail\pi\le d_m}(\mathbf{y}^{(m)})\right]\right| \\ &\le \bar{\gamma}^{d_2} + \sum_{r=3}^{m} \bar{\gamma}^{d_r}(2r-5)^{\sum_{t=2}^{r-1} d_t/2} + (m-1)(2m-5)^{\sum_{t=2}^{m} d_t/2}\sqrt{\xi}\,.\end{aligned}$$

It follows that

$$\begin{aligned}\left|\mathbf{E}_{\mathcal{T}'_3}\left[g(\mathbf{y}^{(2)})\cdots g(\mathbf{y}^{(m)})\right]\right| \le &\left|\mathbf{E}_{\mathcal{T}''_3}\left[g(\mathbf{y}^{(2)})\cdots g(\mathbf{y}^{(m)})\right]\right| \\ &+ 2\bar{\gamma}^{d_2} + 2\sum_{r=3}^{m} \bar{\gamma}^{d_r}(2r-5)^{\sum_{t=2}^{r-1} d_t/2} + 2(m-1)(2m-5)^{\sum_{t=2}^{m} d_t/2}\sqrt{\xi}\end{aligned}$$

and it remains to bound

$$\left|\mathbf{E}_{\mathcal{T}''_3}\left[g(\mathbf{y}^{(2)})\cdots g(\mathbf{y}^{(m)})\right]\right|. \tag{3.7}$$

By the definition of correlation, (3.7) is no greater than

$$\rho\left(\Omega_2^L,\Omega_3^L\times\cdots\times\Omega_m^L;\ \mathcal{T}_3''\right)\left\|g(\mathbf{y}^{(2)})\right\|_{\mathcal{T}_3'',2}\left\|g(\mathbf{y}^{(3)})\cdots g(\mathbf{y}^{(m)})\right\|_{\mathcal{T}_3'',2}\le\rho\left(\Omega_2^L,\Omega_3^L\times\cdots\times\Omega_m^L;\ \mathcal{T}_3''\right).$$

For the distribution $\mathcal{T}_3''$, the coordinates L are independent and so by Lemma 2.15,

$$\rho\left(\Omega_2^L,\Omega_3^L\times\cdots\times\Omega_m^L;\ \mathcal{T}_3''\right)\le\max_{j\in L}\rho\left(\Omega_{2,j},\Omega_{3,j}\times\cdots\times\Omega_{m,j};\ \mathcal{T}_3''\right)=\rho\left(\Omega_2,\Omega_3\times\cdots\times\Omega_m;\ \mathcal{T}_3\right).$$

In Lemma 3.14, we bounded this correlation by $\sqrt{\delta}$.

Consequently,

$$\left|\mathbf{E}_{\mathcal{T}_3'}\left[g(\mathbf{y}^{(2)})\cdots g(\mathbf{y}^{(m)})\right]\right|\le 2\bar{\gamma}^{d_2}+2\sum_{r=3}^{m}\bar{\gamma}^{d_r}(2r-5)^{\sum_{t=2}^{r-1}d_t/2}+2(m-1)(2m-5)^{\sum_{t=2}^{m}d_t/2}\sqrt{\xi}+\sqrt{\delta}.\quad\square$$

Cases $\Gamma\subsetneq\{2,\dots,m\}$ As mentioned for the width-four case, terms $\Gamma\subsetneq\{2,\dots,m\},|\Gamma|\ge 2$, follow via the same arguments; bounds on high-degree terms only produce fewer terms and similarly with the step from low-degree terms to shattered low-degree terms; the argument that the expectation is the same as for independent coordinates is identical and finally the correlation between the same spaces indexed by Γ can not yield a better correlation than between the spaces $\Omega_2^L,\dots,\Omega_m^L$.

3.6 An invariance-style theorem

The following is essentially a variant of the second part of Theorem 1.14 in Mossel, 2010 [19], slightly generalized and without any dependence on α, the least probability of any atom in the relevant probability space. This permits the theorem to be used for analyzing reductions from LABEL COVER without restricting projection degrees. Our proofs are based on the the coordinate-wise substitution method of Lindeberg and more closely the analysis found in O'Donnell and Wu, 2009 [22]. In the following section, we apply this theorem to directly bound mixed, i. e., $\mathbf{E}[f\prod g]$, terms.

Theorem 3.21. *Consider functions $\{f^{(t)}\in\mathrm{L}^\infty(\Omega_t^n)\}_{t\in[m]}$ on a probability space $\mathcal{P}=(\prod_{t=1}^m\Omega_t,\mathrm{P})^{\otimes n}$ and a set $M\subsetneq[m]$. Furthermore, let $\mathcal{C}$ be the collection of minimal sets $C\subseteq[m],C\nsubseteq M$, such that the spaces $\{\Omega_t\}_{t\in C}$ are dependent. Then,*

$$\left|\mathbf{E}\left[\prod f^{(t)}\right]-\prod_{t\notin M}\mathbf{E}\left[f^{(t)}\right]\mathbf{E}\left[\prod_{t\in M}f^{(t)}\right]\right|\le 2^{2m}\max_{C\in\mathcal{C}}\sqrt{\min_{r\in C}\mathrm{TotInf}(f^{(r)})\sum_i\prod_{t\in C-r}\mathrm{Inf}_i(f^{(t)})}\prod_{t\notin C}\left\|f^{(t)}\right\|_\infty.$$

Proof. Let $\{f_S^{(t)}\}_{S\subseteq[n]}$ be the respective Efron-Stein decompositions of the functions. The LHS equals

$$\left|\sum_{\vec{S}\subseteq[n]^m}\left(\mathbf{E}\left[\prod f_{S_t}^{(t)}\right]-\prod_{t\notin M}\mathbf{E}\left[f_{S_t}^{(t)}\right]\mathbf{E}\left[\prod_{t\in M}f_{S_t}^{(t)}\right]\right)\right|.\tag{3.8}$$

For a vector $\vec{S}\subseteq[n]^m$, define $i^*(\vec{S})=\max\{\bigcup_{t\notin M}S_t\cup\{0\}\}$ and $T^*(\vec{S})=\{T\subseteq[m]\mid i^*(\vec{S})\in S_t\}$. Let $\{\mathcal{A}_{i,T}\}_{i\in\{0,\dots,n\},T\subseteq[m]}$ be the partition of $[n]^m$ by these two quantities and note that the choices of $\vec{S}\in\mathcal{A}_{i,T}$

correspond to vectors where S_t is a subset of $[i]$ if $t \notin M$ and else a subset of $[n]$; and $i \in S_t$ iff $t \in T$. Denote by $\mathcal{S}^{(i)}_{T,t}$ the respective choices of S_t, i. e., $\mathcal{A}_{i,T} = \prod \mathcal{S}^{(i)}_{T,t}$.

We note that for $\vec{S}$ such that $i^*(\vec{S}) = 0$, i. e., $S_t = \emptyset$ for every $t \notin M$, the two expectations in (3.8) coincide and the difference is 0. While whenever $\vec{S} \in \mathcal{A}_{i,T}$ for $T \not\subseteq M$, the coordinate i appears in a partition besides M and so by independence and properties of Efron-Stein decompositions, the right term evaluates to 0. Consequently,

$$(3.8) = \left| \sum_{\vec{S} \in \biguplus_{i,T} \mathcal{A}_{i,T}} \left(\mathbf{E}\left[\prod f^{(t)}_{S_t}\right] - \prod_{t \notin M} \mathbf{E}\left[f^{(t)}_{S_t}\right] \mathbf{E}\left[\prod_{t \in M} f^{(t)}_{S_t}\right]\right)\right| = \left| \sum_{i,T \not\subseteq M} \sum_{\vec{S} \in \mathcal{A}_{i,T}} \mathbf{E}\left[\prod f^{(t)}_{S_t}\right]\right|, \tag{3.9}$$

where the sum is over $i \in [n]$.

Similarly, whenever $T^*(\vec{S}) \not\subseteq M$ is a strict subset of some $C \in \mathcal{C}$, the expectation evaluates to 0 by independence and properties of the decompositions, i. e.,

$$(3.9) = \left| \sum_{T : \exists_{C \in \mathcal{C}} T \supseteq C} \sum_i \sum_{\vec{S} \in \mathcal{A}_{i,T}} \mathbf{E}\left[\prod f^{(t)}_{S_t}\right]\right|. \tag{3.10}$$

By the choices of $\vec{S} \in \mathcal{A}_{i,t}$ and noting that $T \supseteq C, C \in \mathcal{C}$ implies $T \not\subseteq M$,

$$(3.10) = \left| \sum_{T \supseteq C, C \in \mathcal{C}} \sum_i \mathbf{E}\left[\prod \left(\sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right)\right]\right| \leq \sum_{T \supseteq C, C \in \mathcal{C}} \sum_i \left| \mathbf{E}\left[\prod \left(\sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right)\right]\right|. \tag{3.11}$$

Let $r \in C$ be arbitrary such that $C - \{r\} \not\subseteq M$ which is well-defined as all $C \in \mathcal{C}$, being minimal dependent sets not contained in M, satisfy $|C - M| \geq 1$ and $|C| \geq 2$. Consider writing the expectation

$$\mathbf{E}\left[\prod_{t \in [m]} \left(\sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right)\right] = \mathbf{E}\left[\left(\sum_{S \in \mathcal{S}^{(i)}_{T,r}} f^{(r)}_S \right) \cdot \prod_{t \in C - \{r\}} \left(\sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right) \cdot \prod_{t \notin C} \left(\sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right)\right].$$

Applying Hölder's inequality to these three factors with respective parameters $2, 2$, and ∞,

$$(3.11) \leq \sum_{T \supseteq C, C \in \mathcal{C}} \min_{r \in C} \sum_i \left\| \sum_{S \in \mathcal{S}^{(i)}_{T,r}} f^{(r)}_S \right\|_2 \left\| \prod_{t \in C - \{r\}} \left(\sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right)\right\|_2 \prod_{t \notin C} \left\| \sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right\|_\infty. \tag{3.12}$$

As we assumed every $C \in \mathcal{C}$ was a minimal dependent set not contained in M and $C - \{r\}$ is not a subset of M, the spaces $\{\Omega_t\}_{t \in C - \{r\}}$ are independent and in effect,

$$(3.12) = \sum_{T \supseteq C, C \in \mathcal{C}} \sum_i \prod_{t \in C} \left\| \sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right\|_2 \prod_{t \notin C} \left\| \sum_{S \in \mathcal{S}^{(i)}_{T,t}} f^{(t)}_S \right\|_\infty. \tag{3.13}$$

Consider first the factor

$$\left\| \sum_{S \in \mathcal{S}_{T,t}^{(i)}} f_S^{(t)} \right\|_2$$

for some $t \notin M$ and $t \in C \subseteq T$. By construction, $\mathcal{S}_{T,t}^{(i)}$ consists of the subsets of $[i]$ containing i, i. e., the norm equals

$$\mathbf{E}\left[\left(\sum_{S \subseteq [i], i \in S} f_S^{(t)}\right)^2\right]^{1/2}$$

which by properties of the decomposition equals

$$\left(\sum_{S \subseteq [i], i \in S} \mathbf{E}\left[\left(f_S^{(t)}\right)^2\right]\right)^{1/2}.$$

Recalling the expression of influences as squares of Efron-Stein terms, this quantity is bounded from above by $\sqrt{\mathrm{Inf}_i(f^{(t)})}$. Similarly, whenever $t \in M$ and $t \in T$, the factors correspond exactly to expression of influences. Hence,

$$(3.13) \leq \sum_{T \supseteq C, C \in \mathcal{C}} \sum_i \sqrt{\prod_{t \in C} \mathrm{Inf}_i(f^{(t)})} \left\| \prod_{t \notin C} \sum_{S \in \mathcal{S}_{T,t}^{(i)}} f_S^{(t)} \right\|_\infty . \tag{3.14}$$

From Lemma 1.10, for every $t \in M$, $\sum_{S \in \mathcal{S}_{T,t}^i} f_S$ equals $f_{\subseteq [n]-\{i\}}$ if $t \notin T$ and otherwise $f_{\subseteq [n]} - f_{\subseteq [n]-\{i\}}$; either possibility bounded by $2\|f\|_\infty$. In light of this,

$$\prod_{t \notin C} \left\| \sum_{S \in \mathcal{S}_{T,t}^{(i)}} f_{S_t}^{(t)} \right\|_\infty \leq \prod_{t \notin C} 2 \left\| f^{(t)} \right\|_\infty \leq 2^m \prod_{t \notin C} \left\| f^{(t)} \right\|_\infty .$$

Returning to our expression,

$$(3.14) \leq 2^m \sum_{T \supseteq C, C \in \mathcal{C}} \sum_i \sqrt{\prod_{t \in C} \mathrm{Inf}_i(f^{(t)})} \prod_{t \notin C} \left\| f^{(t)} \right\|_\infty . \tag{3.15}$$

The square root of the influence of a coordinate for a function is bounded by the infinity norm of said function. Hence, the maximum of (3.15) is achieved for some $T = C, C \in \mathcal{C}$, and

$$\begin{aligned} (3.15) &\leq 2^{2m} \max_{C \in \mathcal{C}} \sum_i \sqrt{\prod_{t \in C} \mathrm{Inf}_i(f^{(t)})} \prod_{t \notin C} \left\| f^{(t)} \right\|_\infty \\ &= 2^{2m} \max_{C \in \mathcal{C}} \min_{r \in C} \sum_i \sqrt{\mathrm{Inf}_i(f^{(r)})} \sqrt{\prod_{t \in C \setminus \{r\}} \mathrm{Inf}_i(f^{(t)})} \prod_{t \notin C} \left\| f^{(t)} \right\|_\infty , \end{aligned} \tag{3.16}$$

where r is an arbitrary member of C. Applying Cauchy-Schwarz over i,

$$(3.16) \leq 2^{2m} \max_{C\in\mathcal{C}} \min_{r\in C} \left(\sum_i \mathrm{Inf}_i(f^{(r)})\right)^{1/2} \left(\sum_i \prod_{t\in C\setminus\{r\}} \mathrm{Inf}_i(f^{(t)})\right)^{1/2} \prod_{t\notin C} \left\|f^{(t)}\right\|_\infty$$
$$= 2^{2m} \max_{C\in\mathcal{C}} \sqrt{\min_{r\in C} \mathrm{TotInf}(f^{(r)}) \sum_i \prod_{t\in C\setminus\{r\}} \mathrm{Inf}_i(f^{(t)})} \prod_{t\notin C} \left\|f^{(t)}\right\|_\infty ,$$

as desired. □

Remark 3.22. Given a function $g : \Omega^{nd} \to \mathbb{R}$ and a projection $\pi : L \to K$ where $\overline{g}^\pi : (\Omega^d)^n \to \mathbb{R}$ is suitably defined, the influence of a coordinate $i \in K$ translates naturally to the sum of influences $j \in L$ which project to i. Namely, we have

$$\mathrm{Inf}_i(\overline{g}^\pi) = \mathrm{Inf}_{\pi^{-1}(i)}(g) \leq \sum_{j:\pi(j)=i} \mathrm{Inf}_j(g) .$$

This follows from the expression of influences in decompositions of g which equals $\sum_{T:i\in\pi(T)} \mathbf{E}\left[g_T^2\right]$ in the former two cases and $\sum_T |T\cap\pi^{-1}(i)|\,\mathbf{E}\left[g_T^2\right]$ in the third.

Corollary 3.23. *Let $\mathcal{P} = (\prod_{t=1}^m \Omega_t, \mathrm{P})$ be a probability space such that Ω_t is independent of Ω_1 for $t = 2,\ldots,m$. Consider label sets $K, L = K\times[d]$, and mean-zero functions $f : \Omega_1^K \to [-1,1]$ and $\{g^{(t)} : \Omega_t^L \to [-1,1]\}_{t=2}^m$. Finally, let $\gamma_1,\ldots,\gamma_m \in (0,1]$ be parameters and define*

$$\mathcal{P}' = \left(\mathrm{T}^{(1)}_{\bar{\gamma}_1} \left(\mathrm{T}^{(2)}_{\bar{\gamma}_2} \cdots \mathrm{T}^{(m)}_{\bar{\gamma}_m} \mathcal{P}\right)^{d\text{-proj-}1}\right)^{\otimes K} .$$

Then,

$$\left|\mathbf{E}_{\mathcal{P}'}\left[f \prod g^{(t)}\right]\right| \leq 2^{2m} \sqrt{\gamma^{-1} \max_r \sum_{i,j} \mathrm{Inf}_i^{(\bar{\gamma}_1)}(f)\, \mathrm{Inf}_{(i,j)}^{(\bar{\gamma}_r)}(g^{(r)})} ,$$

where $\gamma = \min_{t\neq 1} \gamma_t$.

Proof. We aim to employ Theorem 3.21 with the partition $[m] = M_1 \cup M_2, M_1 = \{1\}, M_2 = M = \{2,\ldots,m\}$. To this end, we define lifted functions $\overline{g^{(t)}} : (\Omega_t^d)^L \to [-1,1]$ as $\overline{g^{(t)}}(\overline{\mathbf{y}}) \triangleq g(\mathbf{y})$ where $\overline{y}_{i,r} = y_{(i,r)}$. Hence, as f has mean zero,

$$\left|\mathbf{E}_{\mathcal{P}'}\left[f\prod g^{(t)}\right]\right| = \left|\mathbf{E}_{\mathcal{P}'}\left[f\prod g^{(t)}\right] - \mathbf{E}_{\mathcal{P}'}[f]\,\mathbf{E}_{\mathcal{P}'}\left[\prod g^{(t)}\right]\right| = \left|\mathbf{E}_{\mathcal{P}'}\left[f\prod \overline{g^{(t)}}\right] - \mathbf{E}_{\mathcal{P}'}[f]\,\mathbf{E}_{\mathcal{P}'}\left[\prod \overline{g^{(t)}}\right]\right| . \tag{3.17}$$

As a consequence of the assumption that Ω_t is independent of Ω_1 for any $t \neq 1$, Ω_t^d is independent of Ω_1. In turn, the minimal dependent sets $C \not\subseteq M$ of $\mathcal{P}'$ are supersets of $\{\{1,r,r'\}\}_{r\neq r'\in[m]}$. Theorem 3.21, together with that the influence of a coordinate for a function is no greater than the squared infinity-norm of the function, implies that (3.17) is bounded by

$$2^{2m} \max_{2\le r,r'\le m;\ r\ne r'} \sqrt{\mathrm{TotInf}\left(\overline{g^{(r')}}\right) \sum_i \mathrm{Inf}_i(f)\mathrm{Inf}_i\left(\overline{g^{(r)}}\right)} \prod_{t\ne 1,r,r'} \left\|\overline{g^{(t)}}\right\|_\infty, \tag{3.18}$$

with respect to the distribution $\mathcal{P}'$. With respect to the uniform distribution, the expression equals

$$2^{2m} \max_{2\le r,r'\le m;\ r\ne r'} \sqrt{\mathrm{TotInf}\left(\overline{\mathrm{T}_{\bar\gamma_{r'}}\, g^{(r')}}\right) \sum_i \mathrm{Inf}_i(\mathrm{T}_{\bar\gamma_1} f)\mathrm{Inf}_i\left(\overline{\mathrm{T}_{\bar\gamma_r}\, g^{(r)}}\right)} \prod_{t\ne 1,r,r'} \left\|\overline{\mathrm{T}_{\bar\gamma_t}\, g^{(t)}}\right\|_\infty. \tag{3.19}$$

As earlier remarked, for a function $\bar g : (\Omega^d)^K \to \mathbb{R}$, $\mathrm{Inf}_i(\bar g) \le \sum_{j\in\pi^{-1}(i)} \mathrm{Inf}_j(g)$ and so $\mathrm{TotInf}(\bar g) \le \mathrm{TotInf}(g)$. For a γ-noised function, the total influence is bounded by γ^{-1} and hence

$$\mathrm{TotInf}\left(\overline{\mathrm{T}_{\bar\gamma_{r'}}\, g^{(r')}}\right) \le \gamma_{r'}^{-1} \le \gamma^{-1}$$

where $\gamma = \min_{t\ne 1} \gamma_t$. We also note that the infinity-norm of $\overline{\mathrm{T}_{\bar\gamma_t}\, g^{(t)}}$ is bounded by the maximum of $\left|g^{(t)}\right|$ which is at most one. Hence, the difference (3.19) is bounded by

$$\begin{aligned} 2^{2m} \max_{2\le r,r'\le m;\ r\ne r'} \sqrt{\gamma^{-1} \sum_i \mathrm{Inf}_i^{(\bar\gamma_1)}(f)\mathrm{Inf}_i\left(\overline{\mathrm{T}_{\bar\gamma_r}\, g^{(r)}}\right)} &\le 2^{2m} \sqrt{\gamma^{-1} \max_r \sum_i \mathrm{Inf}_i^{(\bar\gamma_1)}(f) \sum_j \mathrm{Inf}_{(i,j)}\left(\mathrm{T}_{\bar\gamma_r}\, g^{(r)}\right)} \\ &= 2^{2m} \sqrt{\gamma^{-1} \max_r \sum_{i,j} \mathrm{Inf}_i^{(\bar\gamma_1)}(f)\mathrm{Inf}_{(i,j)}^{(\bar\gamma_r)}\left(g^{(r)}\right)}. \end{aligned}$$

□

3.7 Proof of Proposition 3.11: Bounding $\mathbf{E}_{\mathcal{T}_3'}[f\prod g]$, general case

Bounding mixed, i. e., $\mathbf{E}[f\prod g]$, terms in the general case is straightforward with the invariance-style theorem of the preceding subsection

Proof. We apply Theorem 3.23 with $\gamma_1 = \eta, \gamma_2 = \gamma, \ldots, \gamma_m = \gamma$, $g^{(t)} = g$ and, without loss of generality, permuting coordinates such that $\pi(i,j) = i$. Applying the corollary, mixed terms are bounded as

$$\left|\psi_\Gamma(\mathcal{T}_3')\right| \le 2^{2|\Gamma|} \sqrt{\gamma^{-1} \sum_{i,j:\pi(j)=i} \mathrm{Inf}_i^{(\bar\eta)}(f) \max_{t\in\Gamma\setminus\{1\}} \mathrm{Inf}_j^{(\bar\gamma_t)}(g)},$$

where $\gamma = \min_{t\in\Gamma\setminus\{1\}} \gamma_t \ge \gamma_m$. Influence only decreases influence and the above is bounded by

$$2^{2m} \sqrt{\gamma^{-1} \sum_{i,j:\pi(j)=i} \mathrm{Inf}_i^{(\bar\eta)}(f) \max_{t\in\Gamma\setminus\{1\}} \mathrm{Inf}_j^{(\bar\gamma)}(g)},$$

yielding the desired bound. □

Acknowledgement The author would like to thank Johan Håstad for his invaluable advice, curious discussions, and intuitive explanations; Sangxia Huang for discussions and calling previous work to attention; and anonymous reviewers for a host of comments hopefully rendering this work more accessible.

References

[1] PER AUSTRIN AND JOHAN HÅSTAD: On the usefulness of predicates. *ACM Trans. Computation Theory*, 5(1):1, 2013. Preliminary version in CCC'12. [doi:10.1145/2462896.2462897] 716

[2] PER AUSTRIN AND ELCHANAN MOSSEL: Approximation resistant predicates from pairwise independence. *Comput. Complexity*, 18(2):249–271, 2009. Preliminary version in CCC'08. See also at ECCC. [doi:10.1007/s00037-009-0272-6] 706

[3] WILLIAM BECKNER: Inequalities in Fourier analysis. *Ann. of Math.*, 102(1):159–182, 1975. [doi:10.2307/1970980] 711

[4] ALINE BONAMI: Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970. NUMDAM. 711

[5] SIU ON CHAN: Approximation resistance from pairwise independent subgroups. In *Proc. 45th STOC*, pp. 447–456. ACM Press, 2013. See also at ECCC. [doi:10.1145/2488608.2488665] 706

[6] IRIT DINUR, ELCHANAN MOSSEL, AND ODED REGEV: Conditional hardness for approximate coloring. *SIAM J. Comput.*, 39(3):843–873, 2009. Preliminary version in STOC'06. See also at ECCC. [doi:10.1137/07068062X] 729

[7] BRADLEY EFRON AND CHARLES STEIN: The jackknife estimate of variance. *Ann. Stat.*, 9(3):586–596, 1981. [doi:10.1214/aos/1176345462] 711

[8] URIEL FEIGE: A threshold of ln n for approximating set cover. *J. ACM*, 45(4):634–652, 1998. Preliminary version in STOC'96. [doi:10.1145/285055.285059] 714

[9] VITALY FELDMAN, VENKATESAN GURUSWAMI, PRASAD RAGHAVENDRA, AND YI WU: Agnostic learning of monomials by halfspaces is hard. *SIAM J. Comput.*, 41(6):1558–1590, 2012. Preliminary version in FOCS'09. [doi:10.1137/120865094] 708

[10] VENKATESAN GURUSWAMI, PRASAD RAGHAVENDRA, RISHI SAKET, AND YI WU: Bypassing UGC from some optimal geometric inapproximability results. In *Proc. 23rd Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'12)*, pp. 699–717. ACM Press, 2012. [ACM:2095174] 706, 708

[11] JOHAN HÅSTAD: Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. Preliminary version in STOC'97. [doi:10.1145/502090.502098] 704, 715

[12] JOHAN HÅSTAD: On the approximation resistance of a random predicate. *Comput. Complexity*, 18(3):413–434, 2009. Preliminary version in APPROX'07. [doi:10.1007/s00037-009-0262-8] 704

[13] JOHAN HÅSTAD: On the NP-hardness of Max-Not-2. In *Proc. 15th Internat. Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'12)*, pp. 170–181. Springer, 2012. [doi:10.1007/978-3-642-32512-0_15] 707, 711

[14] SANGXIA HUANG: Approximation resistance on satisfiable instances for predicates strictly dominating PARITY. *Electron. Colloq. on Comput. Complexity (ECCC)*, 19:40, 2012. ECCC. 705

[15] SANGXIA HUANG: Approximation resistance on satisfiable instances for predicates with few accepting inputs. In *Proc. 45th STOC*, pp. 457–466. ACM Press, 2013. [doi:10.1145/2488608.2488666] 706

[16] SUBHASH KHOT: Hardness results for coloring 3-colorable 3-uniform hypergraphs. In *Proc. 43rd FOCS*, pp. 23–32. IEEE Comp. Soc. Press, 2002. [doi:10.1109/SFCS.2002.1181879] 708, 714

[17] SUBHASH KHOT: On the power of unique 2-prover 1-round games. In *Proc. 34th STOC*, pp. 767–775. ACM Press, 2002. [doi:10.1145/509907.510017] 705

[18] SUBHASH KHOT AND RISHI SAKET: A 3-query non-adaptive PCP with perfect completeness. In *Proc. 21st IEEE Conf. on Computational Complexity (CCC'06)*, pp. 159–169. IEEE Comp. Soc. Press, 2006. [doi:10.1109/CCC.2006.5] 708

[19] ELCHANAN MOSSEL: Gaussian bounds for noise correlation of functions. *Geom. Funct. Anal.*, 19(6):1713–1756, 2010. Preliminary version in FOCS'08. [doi:10.1007/s00039-010-0047-x] 706, 707, 708, 711, 720, 723, 726, 728, 729, 750

[20] ELCHANAN MOSSEL, RYAN O'DONNELL, AND KRZYSZTOF OLESZKIEWICZ: Noise stability of functions with low influences: invariance and optimality. *Annals of Math.*, 171(1):295–341, 2010. Preliminary version in FOCS'05, see also in CoRR. [doi:10.4007/annals.2010.171.295] 707

[21] RYAN O'DONNELL AND JOHN WRIGHT: A new point of NP-hardness for Unique Games. In *Proc. 44th STOC*, pp. 289–306. ACM Press, 2012. [doi:10.1145/2213977.2214005] 708

[22] RYAN O'DONNELL AND YI WU: Conditional hardness for satisfiable 3-CSPs. In *Proc. 41st STOC*, pp. 493–502. ACM Press, 2009. [doi:10.1145/1536414.1536482] 705, 720, 735, 750

[23] CHRISTOS H. PAPADIMITRIOU AND MIHALIS YANNAKAKIS: Optimization, approximation, and complexity classes. *J. Comput. System Sci.*, 43(3):425–440, 1991. Preliminary version in STOC'88. [doi:10.1016/0022-0000(91)90023-X] 714

[24] RAN RAZ: A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998. Preliminary version in STOC'95. [doi:10.1137/S0097539795280895] 714

[25] SUGURU TAMAKI AND YUICHI YOSHIDA: A query efficient non-adaptive Long Code test with perfect completeness. In *Proc. 14th Internat. Workshop on Randomization and Computation (RANDOM'10)*, pp. 738–751. Springer, 2010. See also at ECCC. [doi:10.1007/978-3-642-15369-3_55] 706

[26] LINQING TANG: Conditional hardness of approximating satisfiable Max 3CSP-q. In *Internat. Symp. Algorithms and Computation (ISAAC'09)*, pp. 923–932. Springer, 2009. [doi:10.1007/978-3-642-10631-6_93] 706

[27] CENNY WENNER: Circumventing d-to-1 for approximation resistance of satisfiable predicates strictly containing parity of width at least four. *Electron. Colloq. on Comput. Complexity (ECCC)*, 19:145, 2012. ECCC. 703

[28] CENNY WENNER: Circumventing d-to-1 for approximation resistance of satisfiable predicates strictly containing parity of width four (Extended Abstract). In *Proc. 15th Internat. Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'12)*, pp. 325–337. Springer, 2012. See also at ECCC 19:145, 2012. [doi:10.1007/978-3-642-32512-0_28] 703

AUTHOR

Cenny Wenner
Ph. D. student
Stockholm University, Stockholm, Sweden
cenny@cwenner.net
http://www.cwenner.net

ABOUT THE AUTHOR

CENNY WENNER is a Ph. D. student of the Theory Group at KTH Royal Institute of Technology and Stockholm University advised by professors Johan Håstad and Viggo Kann. His thesis topic is the hardness of approximating NP-hard optimization problems with a particular inclination towards studying the leverage points of modern techniques and whether the area really needs pesky unproven conjectures. In the author's spare time, he enjoys most geeky activities such as playing Go, Mahjong, programming, volunteering, and daydreaming about making a difference in the world when he grows up.