Testing Linearity against Non-Signaling Strategies

Alessandro Chiesa
UC Berkeley, Berkeley (CA), USA
alexch@berkeley.edu

Peter Manohar
UC Berkeley, Berkeley (CA), USA
manohar@berkeley.edu

Igor Shinkar
UC Berkeley, Berkeley (CA), USA
igors@berkeley.edu

Abstract

Non-signaling strategies are collections of distributions with certain non-local correlations. They have been studied in Physics as a strict generalization of quantum strategies to understand the power and limitations of Nature's apparent non-locality. Recently, they have received attention in Theoretical Computer Science due to connections to Complexity and Cryptography.

We initiate the study of Property Testing against non-signaling strategies, focusing first on the classical problem of *linearity testing* (Blum, Luby, and Rubinfeld; JCSS 1993). We prove that any non-signaling strategy that passes the linearity test with high probability must be close to a *quasi-distribution* over linear functions.

Quasi-distributions generalize the notion of probability distributions over global objects (such as functions) by allowing negative probabilities, while at the same time requiring that "local views" follow standard distributions (with non-negative probabilities). Quasi-distributions arise naturally in the study of Quantum Mechanics as a tool to describe various non-local phenomena.

Our analysis of the linearity test relies on Fourier analytic techniques applied to quasi-distributions. Along the way, we also establish general equivalences between non-signaling strategies and quasi-distributions, which we believe will provide a useful perspective on the study of Property Testing against non-signaling strategies beyond linearity testing.

2012 ACM Subject Classification Theory of computation → Computational complexity and cryptography

Keywords and phrases property testing, linearity testing, non-signaling strategies, quasi-distributions

Digital Object Identifier 10.4230/LIPIcs.CCC.2018.17

Related Version Full version is available on the Electronic Colloquium on Computational Complexity as TR18-067, `https://eccc.weizmann.ac.il/report/2018/067/`.

Funding This work was supported by the UC Berkeley Center for Long-Term Cybersecurity.

Acknowledgements We are grateful to Aneesh Manohar for helpful discussions on the prior uses of quasi-distributions in quantum mechanics. We thank Tom Gur and Thomas Vidick for useful discussions and suggestions that have improved the presentation in this paper. We thank anonymous reviewers who brought [47, 42, 2] to our attention, encouraged us to also explore statements for non-signaling players, and provided other valuable feedback.

33rd Computational Complexity Conference (CCC 2018).
Editor: Rocco A. Servedio; Article No. 17; pp. 17:1–17:37

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Property Testing studies sublinear-time algorithms for approximate decision problems. A *tester* is an algorithm that receives oracle access to an input, samples a small number of locations, queries the input at these locations, and then decides whether to accept or reject. If the input has a certain property, the tester must accept with high probability; if instead the input is far from all inputs having this property, then the tester must reject with high probability.

Seminal works in Property Testing include those of Blum, Luby, and Rubinfeld [15], who studied the problem of deciding whether the input is the evaluation table of a linear function or is far from any such table, and of Rubinfeld and Sudan [45], who studied the analogous problem for low-degree functions. Property Testing for general decision problems was introduced in the foundational work of Goldreich, Goldwasser, and Ron [26].

We initiate the study of Property Testing when the input is a *non-signaling strategy* [35, 43, 40, 41], which means that the input belongs to a certain class of probabilistic oracles that answer a tester's queries by sampling from a distribution that may depend on all queries. This setting stands in stark contrast to the standard one, where each query's answer is *fixed* before queries are sampled. We provide a first analysis of linearity testing against non-signaling strategies, establishing general statements and techniques about non-signaling strategies along the way.

Non-signaling strategies have been studied in Physics for over 30 years as a strict generalization of quantum strategies, in order to understand the power and limitations of Nature's apparent non-locality.[1] Informally, Quantum Mechanics is a very accurate description of Nature but it may also be an incomplete one: it has not been successfully combined with General Relativity to get a quantum theory of gravity. Nevertheless, there is wide agreement that Nature forbids instantaneous communication despite its apparent non-locality, so this *non-signaling* property must be part of *any* ultimate theory of Nature. Non-signaling strategies exactly capture this minimal requirement, thus (purportedly) capturing any physically-realizable strategy.

Non-signaling strategies also have strong connections to Complexity Theory and Cryptography. Property Testing against non-signaling strategies is likely to strengthen these connections (see Section 4 for details), and thus we believe that it should be explicitly studied.

1.1 Linearity testing

A boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is *linear* if $f(x)+f(y) = f(x+y)$ for all $x, y \in \{0,1\}^n$, where bits are added modulo two and vectors are added component-wise. The problem of *linearity testing* is to decide whether a given arbitrary boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is linear or is far from all linear functions. Blum, Luby, and Rubinfeld [15] suggest a very simple 3-query tester: sample uniform and independent $x, y \in \{0,1\}^n$, and check that $f(x) + f(y) = f(x + y)$. Perhaps surprisingly, analyzing this tester is far from simple, and a tight characterization of its acceptance probability is still an open problem. Nevertheless, upper and lower bounds on the acceptance probability are known, which is sufficient for applications. Bellare, Coppersmith, Håstad, Kiwi, and Sudan [12] have shown that the acceptance probability is at most $1 - \Delta(f)$, where $\Delta(f)$ is the fractional Hamming distance

[1] "Non-locality" refers to correlations in Nature that appear non-local when interpreted using classical physics.

of f to the closest linear function. Many other works have studied this problem and closely related ones [50, 13, 14, 21]. Finally, Ito and Vidick [30, 52] analyzed linearity testing against quantum strategies. Fixed functions and quantum strategies are both special cases of non-signaling strategies, the subject of this work.

1.2 Non-signaling strategies

A non-signaling strategy is a collection of distributions, one per set of queries, that jointly satisfy certain restrictions. There are two distinct definitions, corresponding to whether the strategy is meant to represent a function or players in a game. Throughout most of this paper, we consider *non-signaling functions*, because the functional view fits better the setting of Property Testing; nevertheless, we also consider *non-signaling players*, and show that our results about non-signaling functions imply corresponding results about non-signaling players (see full version for details).

A *k-non-signaling function* $\mathcal{F}$ extends the notion of a function $f\colon \{0,1\}^n \to \{0,1\}$ as follows: it is a collection $\{\mathcal{F}_S\}_{S\subseteq\{0,1\}^n,|S|\leq k}$ where each $\mathcal{F}_S$ is a *distribution* over functions $f_S\colon S \to \{0,1\}$ and, for every two subsets S and T each of size at most k, the restrictions of $\mathcal{F}_S$ and $\mathcal{F}_T$ to $S\cap T$ are equal as distributions. We sometimes write "$\mathcal{F}(S) = \vec{b}$", for a subset $S \subseteq \{0,1\}^n$ and string $\vec{b} \in \{0,1\}^S$, to denote the event that the function sampled from $\mathcal{F}_S$ equals $\vec{b}$.

Observe that, given any $k \in \{1,\ldots,2^n\}$, every function $f\colon \{0,1\}^n \to \{0,1\}$ naturally induces a k-non-signaling function $\mathcal{F} = \{\mathcal{F}_S\}_{S\subseteq\{0,1\}^n,|S|\leq k}$, namely the one where each $\mathcal{F}_S$ equals the constant distribution that outputs the restriction of f to S with probability 1. More generally, every distribution over functions induces a corresponding k-non-signaling function in a similar way.

However, the set of non-signaling functions is richer, because consistency between local distributions need *not* imply a global distribution, as the following example shows. For $n = 2$ and $k = 2$, consider the non-signaling function $\{\mathcal{F}_S\}_{S\subseteq\{0,1\}^2,|S|\leq 2}$ defined as follows: $\mathcal{F}_{\{00,11\}}$ is uniform over the two functions $\left\{ \begin{smallmatrix} 00\to 0 \\ 11\to 1 \end{smallmatrix} , \begin{smallmatrix} 00\to 1 \\ 11\to 0 \end{smallmatrix} \right\}$ and, for every $\{x,y\} \neq \{00,11\}$, $\mathcal{F}_{\{x,y\}}$ is uniform over $\left\{ \begin{smallmatrix} x\to 0 \\ y\to 0 \end{smallmatrix} , \begin{smallmatrix} x\to 1 \\ y\to 1 \end{smallmatrix} \right\}$. No distribution over functions can explain the above strategy, as any f in the support of such a distribution would have to satisfy $f(00) \neq f(11)$ and $f(x) = f(y)$ for every $\{x,y\} \subseteq \{0,1\}^2 \setminus \{00,11\}$, which is impossible.

1.3 The problem and challenges

We study linearity testing against non-signaling functions, which is the following problem.

▶ **Question 1.1** (informal). *Let $\mathcal{F} = \{\mathcal{F}_S\}_{S\subseteq\{0,1\}^n,|S|\leq k}$ be a k-non-signaling function. Suppose that with probability at least $1-\varepsilon$ (for sufficiently small $\varepsilon \geq 0$) it holds that $f(x)+f(y) = f(x+y)$, where x and y are sampled uniformly and independently from $\{0,1\}^n$ and $f\colon \{x,y,x+y\} \to \{0,1\}$ is sampled from the distribution $\mathcal{F}_{\{x,y,x+y\}}$. Can we deduce any global properties about $\mathcal{F}$?*

In order to build intuition about this question, we temporarily put aside the case when $\varepsilon > 0$, and focus on the case $\varepsilon = 0$, which already turns out to be quite subtle. In other words, let us assume for now that for *every* $x, y \in \{0,1\}^n$ and *every* f in the support of $\mathcal{F}_{\{x,y,x+y\}}$ it holds that $f(x) + f(y) = f(x + y)$. What global properties, if any, can we deduce about $\mathcal{F}$?

Ideally, we would like to characterize the set of *all* non-signaling functions that pass the linearity test with probability 1 and say that this set is related to linear functions. If $\mathcal{F}$

is restricted to answer according to a single fixed function $f\colon \{0,1\}^n \to \{0,1\}$ (as in the standard setting) then f passing the linearity test with probability 1 is *equivalent* to f being linear by definition. On the other extreme, if $\mathcal{F}$ is allowed to answer queries arbitrarily without any non-signaling property then no interesting conclusion is possible. The case of $\mathcal{F}$ being a non-signaling function sits somewhere in between these two extremes: $\mathcal{F}$ is neither a fixed function nor completely arbitrary. We present two examples to highlight the challenges that arise when seeking an answer.

▶ **Example 1.2.** Consider the following 3-non-signaling function $\mathcal{F} = \{\mathcal{F}_S\}_{S\subseteq\{0,1\}^n,|S|\leq 3}$. For every subset $\{x, y, x+y\} \subseteq \{0,1\}^n \setminus \{0^n\}$, the random variable $f \leftarrow \mathcal{F}_{\{x,y,x+y\}}$ is such that $(f(x), f(y), f(x+y))$ is uniform over $\{(0,0,0),(1,1,0),(1,0,1),(0,1,1)\}$; for every subset $\{x, y, z\} \subseteq \{0,1\}^n \setminus \{0^n\}$ with $z \neq x+y$, the random variable $f \leftarrow \mathcal{F}_{\{x,y,z\}}$ is such that $(f(x), f(y), f(z))$ is uniform over $\{0,1\}^3$. For every set $S \subseteq \{0,1\}^n$ containing 0^n, $\mathcal{F}$ samples $f \leftarrow \mathcal{F}_{S\setminus\{0^n\}}$, and outputs the function g where $g(x) = f(x)$ for $x \in S \setminus \{0^n\}$ and $g(0^n) = 0$. Note that $\mathcal{F}$ is 3-non-signaling because for every $S \subseteq \{0,1\}^n \setminus \{0^n\}$ with $|S| = 3$ the restriction of $\mathcal{F}_S$ to any two coordinates $\{x, y\} \subseteq S$ induces a uniformly boolean random function over $f\colon \{x, y\} \to \{0,1\}$. In particular, for distinct $x, y \in \{0,1\}^n \setminus \{0^n\}$ it holds that $\mathcal{F}_{\{0^n,x,y\}}$ outputs 0 on 0^n, and random bits on x and y.

Clearly, $\mathcal{F}$ passes the linearity test with probability 1. Observe that we can alternatively describe its answers according to the following procedure: upon receiving a subset $S \subseteq \{0,1\}^n$, $\mathcal{F}$ samples a uniformly random *linear* function $f\colon \{0,1\}^n \to \{0,1\}$ (independent of S) and returns the restriction of f to S. We can thus explain $\mathcal{F}$ via the uniform distribution over linear functions.

Generalizing from the above example, any non-signaling function that is induced by sampling a linear function from *any* distribution (not just the uniform one) and answering accordingly will pass the linearity test with probability 1. Note that a distribution over linear functions is given by non-negative real numbers $(p_\alpha)_{\alpha\in\{0,1\}^n}$ such that $\sum_{\alpha\in\{0,1\}^n} p_\alpha = 1$, where p_α is the probability of sampling the function $\langle\alpha,\cdot\rangle$. If $\mathcal{F}$ answers according to $(p_\alpha)_{\alpha\in\{0,1\}^n}$, then $\Pr[\mathcal{F}(x) = b] = \sum_{\alpha:\langle\alpha,x\rangle=b} p_\alpha$ for every $x \in \{0,1\}^n$ and $b \in \{0,1\}$; a similar formula holds for more inputs.

The above discussion suggests a natural conjecture: every non-signaling function that passes the linearity test with probability 1 can be explained by some distribution over linear functions. In fact, this conjecture *is* true if the non-signaling strategy is restricted to be a quantum strategy [30, 52]. But the set of non-signaling strategies is strictly larger. Below we show that, perhaps surprisingly, these additional strategies make this conjecture false.

▶ **Example 1.3.** Consider the following 3-non-signaling function $\mathcal{F} = \{\mathcal{F}_S\}_{S\subseteq\{0,1\}^n,|S|\leq 3}$. For every subset $\{x, y, x+y\} \subseteq \{0,1\}^n \setminus \{0^n\}$, $\mathcal{F}_{\{x,y,x+y\}}$ is the following distribution

$$\Pr_{f\leftarrow\mathcal{F}_{\{x,y,x+y\}}}[f(x,y,x+y) = (a_1,a_2,a_3)] = \begin{cases} 1/7 \text{ if } (a_1,a_2,a_3) = (0,0,0) \\ 2/7 \text{ if } (a_1,a_2,a_3) \in \{(1,1,0),(1,0,1),(0,1,1)\} \end{cases}$$

for every subset $\{x, y, z\} \subseteq \{0,1\}^n \setminus \{0^n\}$ with $z \neq x+y$, $\mathcal{F}_{\{x,y,z\}}$ is the following distribution

$$\Pr_{f\leftarrow\mathcal{F}_{\{x,y,z\}}}[f(x,y,z) = (a_1,a_2,a_3)] = \begin{cases} 0 & \text{if } (a_1,a_2,a_3) = (0,0,0) \\ 1/7 & \text{if } (a_1,a_2,a_3) \neq (0,0,0) \end{cases}.$$

If an input set S contains 0^n, $\mathcal{F}_S$ assigns 0^n to 0 and answers the rest according to $\mathcal{F}_{S\setminus\{0^n\}}$. Note that $\mathcal{F}$ is 3-non-signaling because for distinct and non-zero x and y, the distribution of

$\mathcal{F}_{\{x,y\}}$ is

$$\Pr_{f \leftarrow \mathcal{F}_{\{x,y\}}}[f(x,y) = (a_1,a_2)] = \begin{cases} 1/7 & \text{if } (a_1,a_2) = (0,0) \\ 2/7 & \text{if } (a_1,a_2) \neq (0,0) \end{cases} .$$

In particular, for distinct and non-zero x and y, the distribution of $\mathcal{F}_{\{x,y,0^n\}}$ is

$$\Pr_{f \leftarrow \mathcal{F}_{\{x,y,0^n\}}}[f(x,y,0^n) = (a_1,a_2,0)] = \begin{cases} 1/7 & \text{if } (a_1,a_2) = (0,0) \\ 2/7 & \text{if } (a_1,a_2) \in \{(1,0),(0,1),(1,1)\} \end{cases} .$$

Observe that $\mathcal{F}$ passes the linearity test with probability 1. However, unlike before, a distribution over linear functions that explains $\mathcal{F}$ *does not exist*. Namely, there is no probability vector $(p_\alpha)_{\alpha \in \{0,1\}^n}$ with non-negative entries and $\sum_{\alpha \in \{0,1\}^n} p_\alpha = 1$ such that $\Pr[\mathcal{F}(x) = b] = \sum_{\alpha : \langle \alpha, x \rangle = b} p_\alpha$ for all $x \in \{0,1\}^n$ and $b \in \{0,1\}$. In fact, when trying to solve this linear system of equations with $(p_\alpha)_{\alpha \in \{0,1\}^n}$ as the variables, we obtain a solution vector in which some of the entries are *negative*.

The above example is problematic because it seems to suggest that a clean characterization of the set of all non-signaling functions passing the linearity test does not exist. Indeed, it shows that this set is strictly richer than the set of all distributions over linear functions.

1.4 Negative probabilities and quasi-distributions

In order to resolve the difficulty encountered in Example 1.3, we *embrace* negative probabilities (and probabilities greater than 1), and consider the notion of a *quasi-distribution* over boolean functions.

▶ **Definition 1.4** (informal). A *quasi-distribution* is defined as a vector of real numbers $\mathcal{Q} = \{q_f\}_{f \colon \{0,1\}^n \to \{0,1\}}$ such that $\sum_{f \colon \{0,1\}^n \to \{0,1\}} q_f = 1$. Similarly, a *quasi-distribution over linear functions* is a quasi-distribution $\mathcal{Q} = \{q_f\}_{f \colon \{0,1\}^n \to \{0,1\}}$ such that $q_f = 0$ for all f that are not linear functions; in this case, we also allow ourselves to represent the quasi-distribution by a vector $(q_\alpha)_{\alpha \in \{0,1\}^n}$, where each q_α is associated with the linear function $\langle \alpha, \cdot \rangle$.

A function f in a quasi-distribution $\mathcal{Q} = \{q_f\}_f$ is thus "sampled" with "probability" q_f, which means that for every subset $S \subseteq \{0,1\}^n$ and string $\vec{b} \in \{0,1\}^S$ the event "$\mathcal{Q}(S) = \vec{b}$" has *quasi-probability* given by $\widetilde{\Pr}[\,\mathcal{Q}(S) = \vec{b}\,] := \sum_{f \text{ s.t. } f(S) = \vec{b}} q_f$.

This may seem nonsensical, because quasi-probabilities are not restricted to be in $[0,1]$. But this shall soon make sense. In the words of Paul Dirac [22, p.8]: "*Negative energies and probabilities should not be considered as nonsense. They are well-defined concepts mathematically, like a negative sum of money, since the equations which express the important properties of energies and probabilities can still be used when they are negative. Thus negative energies and probabilities should be considered simply as things which do not appear in experimental results.*"

This viewpoint, which plays a central role in our work, is borrowed from Physics, where it is used to describe many physical phenomena [22, 25], including non-signaling ones [2].

While the non-signaling function $\mathcal{F}$ in Example 1.3 cannot be explained by any distribution over linear functions, it *can* be explained by a *quasi*-distribution over linear functions. Concretely, letting q_α represent the probability of "sampling" the function $\langle \alpha, \cdot \rangle$, we solve the following system of linear equations in the variables $(q_\alpha)_{\alpha \in \{0,1\}^n}$:

$$\sum_{\alpha \in \{0,1\}^n} q_\alpha = 1 \quad \text{and} \quad \forall\, x \in \{0,1\}^n \;\; \forall\, b \in \{0,1\} \sum_{\alpha : \langle \alpha, x \rangle = b} q_\alpha = \Pr[\mathcal{F}(x) = b] \;\; .$$

The solution to this system is $q_{\vec{0}} = 1 - \frac{8}{7}\frac{2^n-1}{2^n} < 0$ and $q_\alpha = \frac{8}{7} \cdot \frac{1}{2^n}$ for all $\alpha \neq \vec{0}$. We stress that the solution has a negative entry. One can then verify that the quasi-distribution obtained above not only matches $\mathcal{F}$ on events involving one input (which is by construction) but also on events involving two inputs: $\Pr[\mathcal{F}(x_1) = b_1\,,\, \mathcal{F}(x_2) = b_2] = \sum_{\alpha:\langle\alpha,x_1\rangle=b_1\,,\,\langle\alpha,x_2\rangle=b_2} q_\alpha$ for all $x_1, x_2 \in \{0,1\}^n$ and $b_1, b_2 \in \{0,1\}$. Similarly, the same holds for events involving three inputs.

Crucially, the quasi-probabilities of events that involve a small enough set of inputs "magically" add up to *non-negative* probabilities because, in particular, they describe distributions of $\mathcal{F}$. In other words, like in Dirac's observation above, the negative probabilities "do not appear in experimental results"; in our case the experiment is querying $\mathcal{F}$, and a quasi-distribution is merely a convenient mathematical abstraction to describe it.

The foregoing considerations directly lead to the following observation.

▶ **Observation 1.5.** *If $\mathcal{Q} = (q_f)_{f\colon \{0,1\}^n \to \{0,1\}}$ is a quasi-distribution that induces a probability distribution on every event of at most k inputs, then $\mathcal{Q}$ induces a k-non-signaling function.*

Furthermore, if $\mathcal{Q}$ is supported on linear functions only, then the corresponding k-non-signaling function passes the linearity test with probability 1*.*

The first part of the observation suggests using k as a measure of a quasi-distribution's locality: we say that a quasi-distribution $\mathcal{Q} = (q_f)_f$ is *k-local* if for every k inputs $x_1, \ldots, x_k \in \{0,1\}^n$ and k outputs $b_1, \ldots, b_k \in \{0,1\}$ it holds that $\sum_{f:f(x_1)=b_1,\ldots,f(x_k)=b_k} q_f \geq 0$. Thus $\mathcal{Q}$ behaves like a collection of (standard) distributions on all events that involve at most k inputs and, moreover, these distributions jointly satisfy the k-non-signaling property.

The second part of the observation shows the existence of a class of non-signaling functions that pass the linearity test with probability 1 that is *much richer* than the class of distribution over linear functions. Are there any other types of non-signaling functions that pass the linearity test with probability 1, or are these all of them? Moreover, how does this answer change when we merely require that a non-signaling function pass the linearity test with probability at least $1 - \varepsilon$? We now discuss our results, which will provide answers to these questions.

2 Our results

Quasi-distributions arose rather naturally when reasoning about non-signaling functions. First, we show that this is not a coincidence by proving that the two notions are equivalent.

▶ **Theorem 2.1** (informal)**.** *Local quasi-distributions and non-signaling functions are equivalent:*

1. *every k-local quasi-distribution induces a corresponding k-non-signaling function; conversely,*
2. *every k-non-signaling function has a k-local quasi-distribution that describes it. (In fact, this quasi-distribution is not unique: the set of all such quasi-distributions is an affine subspace.)*

See Section 8 (specifically, Theorem 8.1 and Theorem 8.2) for precise statements of the two items.

The first item is just Observation 1.5. The second item is proved via Fourier analytic techniques applied to a quasi-probability vector. Informally, the Fourier coefficients of quasi-probability vectors are indexed by subsets of $\{0,1\}^n$, and can be grouped into *levels* according to their size. We prove that the only coefficients that matter for the k-non-signaling

function are those in the levels for sizes at most k, while all others change the weights in the quasi-probability vector but do not affect the induced k-non-signaling function.

The foregoing equivalence can be viewed as the "functional analogue" of an equivalence proved in [2] for the (incomparable) case of non-signaling players. The Fourier analytic techniques that we use are novel and, moreover, can be adapted to the case of non-signaling players in order to strengthen [2]'s result to find *all* quasi-distributions (rather than just one) that describe a given set of non-signaling players (see full version for details). We believe that the mathematical structure uncovered by our Fourier analytic techniques is of independent interest.

Having established the equivalence of local quasi-distributions and non-signaling functions, we return to the problem of linearity testing against non-signaling functions. Our first theorem in this direction is a characterization of the set of non-signaling functions that pass the linearity test with probability 1: this set consists of local quasi-distributions over linear functions (essentially).

▶ **Theorem 2.2** (informal). *Let $\mathcal{F} = \{\mathcal{F}_S\}_{S \subseteq \{0,1\}^n, |S| \leq k}$ be a k-non-signaling function such that*

$$\Pr_{\substack{x,y \leftarrow \{0,1\}^n \\ f \leftarrow \mathcal{F}_{\{x,y,x+y\}}}} [f(x) + f(y) = f(x+y)] = 1 \ .$$

There is a unique $(k-1)$-local quasi-distribution $\mathcal{L}$ over linear functions describing $\mathcal{F}$ on all input sets of size $\leq k-1$ ($\mathcal{L}_S$ and $\mathcal{F}_S$ are equal as distributions for every set $S \subseteq \{0,1\}^n$ with $|S| \leq k-1$).

See Theorem 10.1 in Section 10 for the precise statement. (A minor technicality of the theorem is that $\mathcal{L}$ is only $(k-1)$-local and only matches $\mathcal{F}$ on at most $k-1$ inputs; the discussion after Theorem 10.1 explains why this is the best we can hope for.) To prove the theorem we define a quasi-distribution $\mathcal{L}$ over linear functions by solving a certain system of linear equations that ensures that $\mathcal{L}$ and $\mathcal{F}$ match on single inputs, i.e., that $\widetilde{\Pr}[\mathcal{L}(x) = b] = \Pr[\mathcal{F}(x) = b]$ for all $x \in \{0,1\}^n$ and $b \in \{0,1\}$. We then need to establish that $\mathcal{L}$ and $\mathcal{F}$ match on all sets of at most $k-1$ inputs. We do so in two steps: we first use linearity to show that $\mathcal{L}$ and $\mathcal{F}$ match on all parity events (i.e., $\widetilde{\Pr}\left[\sum_{i \in T} \mathcal{L}(x_i) = b\right] = \Pr\left[\sum_{i \in T} \mathcal{F}(x_i) = b\right]$ for all $x_1, \ldots, x_s \in \{0,1\}^n$ and $b \in \{0,1\}$ with $s \leq k-1$); then we use Fourier analysis to extend this claim to all allowed input sets.

We finally return to our original question (Question 1.1). Suppose that a non-signaling function $\mathcal{F}$ passes the linearity test with probability $1-\varepsilon$ for sufficiently small $\varepsilon \geq 0$ (possibly with $\varepsilon > 0$ so Theorem 2.2 does not apply). What can we learn about $\mathcal{F}$? Recall that if $\mathcal{F}$ answers according to a fixed function $f\colon \{0,1\}^n \to \{0,1\}$ (as in standard linearity testing), then we may conclude that f is ε-close to some linear function [15, 12]. The foregoing discussion for the case of $\varepsilon = 0$ leads to a natural conjecture: *non-signaling functions that pass the linearity test with high probability are local quasi-distributions over functions that are close to linear*. Our next theorem implies that this conjecture is true, but in a non-interesting way. That is, it holds even without the hypothesis: *every* k-non-signaling function can be expressed as a quasi-distribution over functions with support of size at most k (namely, over functions that are non-zero for at most k inputs).

▶ **Theorem 2.3** (informal). *Every k-non-signaling function $\mathcal{F}$ can be expressed as a k-local quasi-distribution $\mathcal{Q}$ over functions with support of size at most k.*

The above theorem is quite counterintuitive. On one hand, if $\mathcal{F}$ is described by a distribution over functions that are close to linear, then $\mathcal{F}$ passes the linearity test with high probability. But this simple fact does *not* extend to the case where $\mathcal{F}$ is a *quasi*-distribution over functions that are close to linear. For example, the all-ones function never passes the linearity test, yet Theorem 2.3 implies that it can be expressed as a quasi-distribution over functions with support of size at most k, i.e., functions that are $\frac{k}{2^n}$-close to the all-zeros function (a linear function)!

We prove Theorem 2.3 via a greedy approach: given the non-signaling function $\mathcal{F}$, we iteratively consider small-support functions from heaviest to lightest and, in each iteration, assign to these functions certain quasi-probabilities computed from $\mathcal{F}$. See Theorem 9.1 (in Section 9) for details.

Since our last conjecture turned out to be false, we again look for inspiration in the standard setting in order to formulate another conjecture. Taking a different view, linearity testing tells us that if a function $f\colon \{0,1\}^n \to \{0,1\}$ passes the linearity test with high probability then we know that there exists a linear function L such that for *every* $x \in \{0,1\}^n$ it holds that $L(x) = f(x+y) - f(y)$ with high probability over a random $y \in \{0,1\}^n$. Put another way, the answers to any given query (or, more generally, a set of queries) given by the self-correction of f and by L are close in statistical distance.

The foregoing observation suggests a conjecture: *if a non-signaling function passes the linearity test with high probability, then its self-correction is close to a quasi-distribution over linear functions.*

The self-correction $\hat{\mathcal{F}}$ of a non-signaling function $\mathcal{F}$ is naturally defined: on input $x \in \{0,1\}^n$, $\hat{\mathcal{F}}$ samples a random $y \in \{0,1\}^n$ and outputs $\mathcal{F}(x+y) - \mathcal{F}(y)$; a similar procedure applies if $\hat{\mathcal{F}}$ receives multiple inputs. Note that if $\mathcal{F}$ is k-non-signaling then $\hat{\mathcal{F}}$ is $\hat{k}$-non-signaling with $\hat{k} := \lfloor k/2 \rfloor$.

The notion of distance is also naturally defined: the distance between two non-signaling functions is the maximum statistical distance between the distributions induced on every subset S; the equivalence of non-signaling functions and quasi-distributions (Theorem 2.1) extends this definition to apply between two quasi-distributions, or between a non-signaling function and a quasi-distribution.

The following theorem shows that the conjecture above is in fact true.

▶ **Theorem 2.4** (informal). *Let $\mathcal{F} = \{\mathcal{F}_S\}_{S \subseteq \{0,1\}^n, |S| \le k}$ be a k-non-signaling function such that*

$$\Pr_{\substack{x,y \leftarrow \{0,1\}^n \\ f \leftarrow \mathcal{F}_{\{x,y,x+y\}}}} [f(x) + f(y) = f(x+y)] \ge 1 - \varepsilon \quad \text{for some } \varepsilon \ge 0 \ .$$

There is a $(\hat{k}-1)$-local quasi-distribution $\mathcal{L}$ over linear functions that is $O_{\hat{k}}(\varepsilon)$-close to $\hat{\mathcal{F}}$ on all input sets of size $\le \hat{k}-1$. That is, the maximum statistical distance between $\mathcal{L}_S$ and $\hat{\mathcal{F}}_S$, across all sets $S \subseteq \{0,1\}^n$ with $|S| \le \hat{k}-1$, is $O_{\hat{k}}(\varepsilon)$.

See Theorem 11.2 (in Section 11) for details. Our proof differs significantly from prior proofs of linearity testing in the standard setting. Informally, we start the proof by noting that $\hat{\mathcal{F}}$ satisfies $\Pr_{\hat{f} \leftarrow \hat{\mathcal{F}}_{\{x,y,x+y\}}} \Pr[\hat{f}(x) + \hat{f}(y) = \hat{f}(x+y)] \ge 1 - \hat{\varepsilon}$ for *every* $x, y \in \{0,1\}^n$ and $\hat{\varepsilon} := 4\varepsilon$. (By assumption, $\mathcal{F}$ merely satisfies such a statement for *random* $x, y \in \{0,1\}^n$.) The next step is similar to a step in the proof of Theorem 2.2: we define a quasi-distribution $\mathcal{L}$ over linear functions by solving a system of linear equations that ensures that $\mathcal{L}$ and $\hat{\mathcal{F}}$ match on single inputs, i.e., that $\Pr[\mathcal{L}(x) = b] = \Pr[\mathcal{F}(x) = b]$ for all $x \in \{0,1\}^n$ and $b \in \{0,1\}$.

We are left to argue that $\mathcal{L}$ and $\hat{\mathcal{F}}$ *almost* match on all sets of at most $\hat{k}-1$ inputs, i.e., that the distributions $\mathcal{L}_S$ and $\hat{\mathcal{F}}_S$ are statistically close for $|S| < \hat{k}$. As before, we do so in two steps: we first use linearity to show that $\mathcal{L}$ and $\hat{\mathcal{F}}$ *almost* match on all parity events (i.e., $\Pr[\sum_{i\in T} \hat{\mathcal{F}}(x_i) = b] \approx \widetilde{\Pr}[\sum_{i\in T} \mathcal{L}(x_i) = b]$ for all $x_1, \ldots, x_s \in \{0,1\}$ for $s \leq \hat{k}-1$), and then we use a quantitative Fourier analytic claim (Lemma 5.1) to extend this claim to the remaining query sets.

Finally, we use the foregoing results about non-signaling *functions* to prove analogous statements about non-signaling *players*.

Recall that a *k-non-signaling player* $\mathcal{P}$ extends the notion of k non-communicating players (possibly sharing randomness) as follows: it is a collection $(\mathcal{P}_{(x_1,\ldots,x_k)})_{(x_1,\ldots,x_k)\in\{0,1\}^{k\cdot n}}$ where each $\mathcal{P}_{(x_1,\ldots,x_k)}$ is a *distribution* over functions $f\colon [k] \to \{0,1\}$ (the players' k answers to the k inputs) and, for every two input vectors $(x_1,\ldots,x_k)$ and $(y_1,\ldots,y_k)$ that agree on a subset $I \subseteq [k]$ of entries, the restrictions of $\mathcal{P}_{(x_1,\ldots,x_k)}$ and $\mathcal{P}_{(y_1,\ldots,y_k)}$ to entries in I are equal as distributions. Non-signaling players are a richer class than non-communicating players (and quantum-entangled ones) [40].

Now the linearity test, given a k-non-signaling player $\mathcal{P} = (\mathcal{P}_{(x_1,\ldots,x_k)})_{(x_1,\ldots,x_k)\in\{0,1\}^{k\cdot n}}$, samples random vectors $x, y \in \{0,1\}^n$ and distinct players $i_1, i_2, i_3 \in [k]$, sends the three queries $x, y, x+y$ to the players $\mathcal{P}_{i_1}, \mathcal{P}_{i_2}, \mathcal{P}_{i_3}$, and checks that $\mathcal{P}_{i_1}(x) + \mathcal{P}_{i_2}(y) = \mathcal{P}_{i_3}(x+y)$.

▶ **Theorem 2.5** (informal). *Let $\mathcal{P} = (\mathcal{P}_{(x_1,\ldots,x_k)})_{(x_1,\ldots,x_k)\in\{0,1\}^{k\cdot n}}$ be a k-non-signaling player.*

1. *Suppose that*

$$\Pr_{\substack{x,y\leftarrow\{0,1\}^n \\ i_1,i_2,i_3\leftarrow[k] \\ \mathcal{P}}}[\mathcal{P}_{i_1}(x) + \mathcal{P}_{i_2}(y) = \mathcal{P}_{i_3}(x+y)] = 1 \ .$$

   There exists a $(k-2)$-local quasi-distribution $\mathcal{L}$ over linear functions that describes $\mathcal{P}$.

2. *Suppose that*

$$\Pr_{\substack{x,y\leftarrow\{0,1\}^n \\ i_1,i_2,i_3\leftarrow[k] \\ \mathcal{P}}}[\mathcal{P}_{i_1}(x) + \mathcal{P}_{i_2}(y) = \mathcal{P}_{i_3}(x+y)] \geq 1-\varepsilon \ .$$

   There exists a $(\hat{k}-1)$-local quasi-distribution $\mathcal{L}$ over linear functions that is $O_{\hat{k}}(\varepsilon)$-close to $\hat{\mathcal{P}}$, where $\hat{\mathcal{P}}$ is the (appropriately defined) self-correction of $\mathcal{P}$.

See full version for details. The proof of these theorems show how to reduce to the case of non-signaling functions, which we have already established (in Theorems 2.2 and 2.4 respectively).

We conclude this section via a brief comparison to the case of quantum strategies. Ito and Vidick [30, 52] show that any quantum strategy that passes the linearity test with high probability is close to a *distribution* over linear functions. Our results instead show that, in our setting, we can only hope for a conclusion involving a *quasi-distribution* over linear functions. This qualitative difference is due to the fact that non-signaling strategies are a richer class than quantum strategies.

3 Techniques

We highlight some of the techniques that we use by providing proof sketches of some of our results. We first discuss the ideas behind the equivalence between non-signaling functions and local quasi-distributions (Section 3.1) and then how we analyze the linearity test (Section 3.2). After that, we explain how we derive corresponding results about non-signaling players (Section 3.3).

3.1 Non-signaling functions and local quasi-distributions are equivalent

Our Theorem 2.1 states that non-signaling functions and local quasi-distributions are equivalent. One direction of this equivalence, namely that every k-local quasi-distribution induces a corresponding k-non-signaling function, is a simple observation. Below we focus on the other, more interesting direction, which is: given a k-non-signaling function $\mathcal{F} = \{\mathcal{F}_S\}_{S\subseteq\{0,1\}^n, |S|\leq k}$, how do we construct a quasi-distribution $\mathcal{Q} = \{q_f\}_{f\colon \{0,1\}^n\to\{0,1\}}$ that matches $\mathcal{F}$ on all sets of at most k queries?

We construct $\mathcal{Q}$ by specifying its *Fourier coefficients*. We view $\mathcal{Q} = (q_f)_{f\colon \{0,1\}^n\to\{0,1\}}$ as a function $q\colon \{0,1\}^{\{0,1\}^n} \to \mathbb{R}$ by setting $q(f) := q_f \in \mathbb{R}$, and then write $\mathcal{Q}$ via its Fourier expansion:

$$q(\cdot) = \sum_{T\subseteq\{0,1\}^n} \widehat{q}(T)\chi_T(\cdot) \quad \text{where} \quad \begin{cases} \chi_T(f) := (-1)^{\sum_{x\in T} f(x)} \\ \widehat{q}(T) := \langle q, \chi_T\rangle = \frac{1}{2^{2^n}} \sum_{f\colon D\to\{0,1\}} q(f)\chi_T(f) \end{cases} .$$

We set the 2^{2^n} Fourier coefficients as follows:

$$\widehat{q}(T) := \begin{cases} \frac{1}{2^{2^n}} & \text{if } T = \emptyset \\ \frac{2}{2^{2^n}} \left(\Pr[\sum_{x\in T} \mathcal{F}(x) = 0] - \frac{1}{2}\right) & \text{if } 1 \leq |T| \leq k \\ 0 & \text{if } |T| > k \end{cases} .$$

We have to argue that the above choice of $\mathcal{Q}$ does describe $\mathcal{F}$. First, we show that $\mathcal{F}$ and $\mathcal{Q}$ match on all *parity events* of size at most k, i.e., for all $S \subseteq \{0,1\}^n$ with $|S| \leq k$

$$\Pr\left[\sum_{x\in S} \mathcal{F}(x) = 1\right] = \sum_{f:\sum_{x\in S} f(x)=1} q_f = \widetilde{\Pr}\left[\sum_{x\in S} \mathcal{Q}(x) = 1\right] .$$

Recall (see Section 1.4) that $\widetilde{\Pr}[\cdot]$ denotes the quasi-probability for an event about a quasi-distribution.

Second, we prove that $\Pr[\mathcal{F}(S) \in E] = \widetilde{\Pr}[\mathcal{Q}(S) \in E]$ for every subset $S \subseteq \{0,1\}^n$ and event $E \subseteq \{0,1\}^S$. We build on the previous step by observing that any event can be expressed as a linear combination of parity events: there exist real numbers $\{c_T\}_T$ depending on E such that

$$\Pr[\mathcal{Q}(S) \in E] = \sum_{T\subseteq S} c_T \cdot \widetilde{\Pr}\left[\sum_{x\in T} \mathcal{Q}(x) = 0\right] . \tag{1}$$

In fact, the real numbers $\{c_T\}_T$ are closely related to the Fourier coefficients of the indicator function of E, and this relation is a consequence of the fact that the functions $\{\chi_T(\cdot)\}_T$ depend only on the parities of their inputs. See Lemma 5.1 for details.

The above is merely one quasi-distribution that explains $\mathcal{F}$. We can find other such quasi-distributions by noting that changing $\widehat{q}(T)$ for $|T| > k$ yields quasi-distributions that still match $\mathcal{F}$. Essentially, if $|T| > k$ then $\chi_T(\cdot)$ does not affect the induced distributions on sets of at most k inputs. We then argue that these are the only solutions possible. See Section 8 for details.

3.2 Testing linearity against non-signaling functions

We discuss the ideas behind our analysis of linearity test against non-signaling functions (that is, behind Theorem 2.2 and Theorem 2.4). We first explain why known proofs in the standard setting do not easily extend to our setting, and then we describe the approach that we took.

3.2.1 Difficulties of prior approaches

We begin with a helpful exercise for which difficulties do *not* arise: consider the task of analyzing the linearity test against a *distribution* $\mathcal{D}$ over boolean functions. Namely, if $\Pr[f(x)+f(y)=f(x+y)] \geq 1-\varepsilon$ for $f \leftarrow \mathcal{D}$ and $x, y \leftarrow \{0,1\}^n$ then what can we conclude about $\mathcal{D}$? This case is not hard to analyze: we separately apply known results on linearity testing to each function in the support of $\mathcal{D}$, and conclude that most of $\mathcal{D}$ is concentrated on nearly-linear functions. Indeed, by Markov's inequality, with probability $1-\sqrt{\varepsilon}$ over a choice of $f \leftarrow \mathcal{D}$ it holds that $\Pr_{x,y}[f(x)+f(y)=f(x+y)] \geq 1-\sqrt{\varepsilon}$ and thus that f is $\sqrt{\varepsilon}$-close to a linear function. This conclusion explains why $\mathcal{D}$ passes the linearity test with high probability.

However, when considering the linearity test against a non-signaling function, the situation changes significantly, as we now explain.

The Fourier analytic approach. One of the classical proofs of linearity testing in the standard setting follows a Fourier analytic approach [12]. Unfortunately, we do not see how to use this approach directly on a non-signaling function $\mathcal{F}$, because computing Fourier coefficients requires access to an entire function while $\mathcal{F}$ only provides local views. We could instead rely on the equivalence between non-signaling functions and local quasi-distributions, and apply Fourier analysis to the functions in a quasi-distribution $\mathcal{Q}=(q_f)_{f\colon \{0,1\}^n \to \{0,1\}}$ that describes $\mathcal{F}$. Namely, we could rewrite the probability $\Pr[\mathcal{F}(x)+\mathcal{F}(y)=\mathcal{F}(x+y)]$ as $\sum_f q_f \Pr[f(x)+f(y)=f(x+y)]$, and then reason about the Fourier coefficients of every f. We do not see how to make this work either, because the coefficients $\{q_f\}_f$ can be positive or negative (and even unbounded), which in particular forbids Markov-type arguments. It is also not clear what kind of conclusion we could expect about the Fourier coefficients about *all* functions.

The combinatorial approach. Another classical proof of linearity testing in the standard setting follows a combinatorial approach (e.g., [15, 13]): given the function $f\colon \{0,1\}^n \to \{0,1\}$, define its correction $g\colon \{0,1\}^n \to \{0,1\}$ to be $g(x) := \mathrm{maj}_{y\in\{0,1\}^n} f(x+y)-f(y)$, and show that it is close to f; then show that g is linear as, for every $x \in \{0,1\}^n$, a vast majority of y's yield $g(x)$. This approach also seems to fail in our setting: the foregoing correcting procedure relies on taking majority over *all* $y \in \{0,1\}^n$, but a non-signaling function only accepts up to k inputs at a time.

It is not surprising that prior approaches do not seem to apply to our setting: they were developed to show that a function f passing the linearity test with high probability is nearly-linear. But we already know that every non-signaling function can be described by a quasi-distribution over nearly-linear functions, so we are not interested in conclusions about nearly-linear functions. Instead, we aim to show that (the self-correction of) a non-signaling function passing the linearity test with high probability is close to a quasi-distribution over *linear* functions. We next discuss our approach to establish such a conclusion.

3.2.2 Our approach

Let us once more first focus on the case where a k-non-signaling function $\mathcal{F}$ passes the linearity test with probability 1, namely, $\Pr[\mathcal{F}(x)+\mathcal{F}(y)=\mathcal{F}(x+y)]=1$ for every $x, y \in \{0,1\}^n$. Our first step is to show that there exists a quasi-distribution $\mathcal{L}$ over linear functions that matches $\mathcal{F}$ on single inputs, namely, $\widetilde{\Pr}[\mathcal{L}(x)=b]=\Pr[\mathcal{F}(x)=b]$ for every $x \in \{0,1\}^n$ and

$b \in \{0,1\}$. Viewing $\mathcal{L}$ as a vector $(\ell_\alpha)_{\alpha\in\{0,1\}^n}$ where each α is associated with the linear function $\langle\alpha,\cdot\rangle$, we know that $\mathcal{L}$ must be a solution to the following system of linear equations:

$$\forall\, x \in \{0,1\}^n\,, \qquad \sum_{\alpha:\langle\alpha,x\rangle=0} \ell_\alpha = \Pr[\mathcal{F}(x)=0] \ .$$

Note that it suffices to consider constraints only involving $\Pr[\mathcal{F}(x)=0]$ because $\Pr[\mathcal{F}(x)=1] = 1-\Pr[\mathcal{F}(x)=0]$. Also, $\mathcal{L}$ is a quasi-distribution because $\sum_\alpha \ell_\alpha = \Pr[\mathcal{F}(0^n)=0] = \Pr_{x\leftarrow\{0,1\}^n}[\mathcal{F}(0^n)+\mathcal{F}(x)=\mathcal{F}(x)] = 1$ (as $\mathcal{F}$ always passes the linearity test). This system has a unique solution, which thus defines the quasi-distribution $\mathcal{L}$. We remark that it is no coincidence that quasi-distributions supported on LIN are uniquely defined by their probabilities on sets of size 1: a quasi-distribution is supported on LIN if and only if all of its Fourier coefficients are determined by the coefficients only for sets of size 1 (see full version for details).

Next, we need to argue that $\mathcal{L}$ and $\mathcal{F}$ match on larger sets of inputs. We first argue that they match on all parity events, similarly to the idea behind the equivalence between non-signaling functions and quasi-distributions discussed above (in Section 3.1). Specifically, we use the assumption on linearity to show that for every subset $S \subseteq \{0,1\}^n$ with $|S| < k$ it holds that

$$\widetilde{\Pr}\left[\sum_{x\in S}\mathcal{L}(x)=0\right] = \Pr\left[\sum_{x\in S}\mathcal{F}(x)=0\right] \ .$$

After that, using Eq. (1), we conclude that $\mathcal{L}$ and $\mathcal{F}$ match on all sets S of less than k inputs: we express each event $E \subseteq \{0,1\}^S$ as a linear combination of parity events for both $\mathcal{F}$ and $\mathcal{L}$,

$$\Pr\left[\mathcal{F}(S)\in E\right] = \sum_{T\subseteq S} c_T \cdot \Pr\left[\sum_{x\in T}\mathcal{F}(x)=0\right] \ ,$$

and similarly

$$\widetilde{\Pr}\left[\mathcal{L}(S)\in E\right] = \sum_{T\subseteq S} c_T \cdot \widetilde{\Pr}\left[\sum_{x\in T}\mathcal{L}(x)=0\right] \ .$$

The above shows that matching on parity events implies matching on all sets of less than k inputs.

Let us now relax the assumption that $\mathcal{F}$ passes the linearity test with probability 1 to merely that it passes the test with high probability, say at least $1-\varepsilon$ for $\varepsilon > 0$. We first consider $\hat{\mathcal{F}}$, which is the $\hat{k}$-non-signaling self-correction of $\mathcal{F}$ (with $\hat{k} := k/2$), and observe that there exists $\hat{\varepsilon} = 4\varepsilon$ such that $\hat{\mathcal{F}}$ satisfies, for *every* $x, y \in \{0,1\}^n$,

$$\Pr_{\hat{f}\leftarrow\hat{\mathcal{F}}_{\{x,y,x+y\}}}[\hat{f}(x)+\hat{f}(y)=\hat{f}(x+y)] \geq 1-\hat{\varepsilon} \ .$$

Note that, by assumption, $\mathcal{F}$ merely satisfies such a statement for *random* $x, y \in \{0,1\}^n$.

The next step is similar to the "$\varepsilon = 0$" case discussed above: we define a quasi-distribution $\mathcal{L}$ over linear functions by solving the system of linear equations that ensures that $\mathcal{L}$ and $\hat{\mathcal{F}}$ match on all single inputs, i.e., that $\widetilde{\Pr}[\mathcal{L}(x)=b] = \Pr[\hat{\mathcal{F}}(x)=b]$ for all $x \in \{0,1\}^n$ and $b \in \{0,1\}$.

We then argue that $\mathcal{L}$ and $\hat{\mathcal{F}}$ *almost* match on sets of less than $\hat{k}$ inputs, i.e., that the distributions $\mathcal{L}_S$ and $\hat{\mathcal{F}}_S$ are statistically close for every $S \subseteq \{0,1\}^n$ with $|S| < \hat{k}$. We do so,

again, in two steps. First, we use the almost linearity of $\hat{\mathcal{F}}$ to show that $\mathcal{L}$ and $\hat{\mathcal{F}}$ *almost* match on all parity events. Specifically, we show that for every subset $T \subseteq \{0,1\}^n$ with $|T| < \hat{k}$ and $b \in \{0,1\}$,

$$\left| \Pr\left[\sum_{x \in T} \hat{\mathcal{F}}(x) = b \right] - \widetilde{\Pr}\left[\sum_{x \in T} \mathcal{L}(x) = b \right] \right| < (|T| - 1)\hat{\varepsilon} \ .$$

Then, we use Eq. (1) to extend this claim to all events on these query sets: for every subset $S \subseteq \{0,1\}^n$ with $|S| < \hat{k}$ and event $E \subseteq \{0,1\}^S$

$$\left| \Pr[\hat{\mathcal{F}}(S) \in E] - \widetilde{\Pr}[\mathcal{L}(S) \in E] \right| < \sum_{T \subseteq S} |c_T| \cdot (|T| - 1) \cdot \hat{\varepsilon} \ .$$

Crucially, unlike the case of $\varepsilon = 0$, here we need *quantitative* bounds on the coefficients $\{c_T\}_T$ in order to derive an upper bound. We prove such bounds in Lemma 5.1.

Finally, while $\mathcal{L}$ is close to $\hat{\mathcal{F}}$ (see Definition 7.5 for how to extend the notion of statistical distance to our setting), it is possible that $\mathcal{L}$ does not induce a distribution on all subsets $S \subseteq \{0,1\}^n$ with $|S| < \hat{k}$, because it could be that $\widetilde{\Pr}[\mathcal{L}(S) \in E]$ is negative for some S and $E \subseteq \{0,1\}^S$. However, since $\Pr[\hat{\mathcal{F}}(S) \in E]$ is a probability (i.e., a number between 0 and 1), for all subsets $S \subseteq \{0,1\}^n$ with $|S| < \hat{k}$ it holds that $\widetilde{\Pr}[\mathcal{L}(S) \in E] \in [-\varepsilon', 1 + \varepsilon']$ for $\varepsilon' := (|S| - 1) \cdot \sqrt{|E|} \cdot \hat{\varepsilon}$. We then show that $\mathcal{L}$ can be corrected to obtain a $(\hat{k} - 1)$-local quasi-distribution $\mathcal{L}'$ that is close to $\mathcal{L}$ (see Corollary 7.9). By triangle inequality this implies that $\mathcal{L}'$ is also close to $\hat{\mathcal{F}}$.

See Section 11 for details.

3.3 Extending the analysis to non-signaling players

We make a "black-box" use of our results on testing linearity against non-signaling *functions* to derive corresponding results on testing linearity against non-signaling *players*. Recall that, given a k-non-signaling player $\mathcal{P} = (\mathcal{P}_{(x_1,\ldots,x_k)})_{(x_1,\ldots,x_k) \in \{0,1\}^{k \cdot n}}$, the linearity test is now as follows: sample $x, y \in \{0,1\}^n$ and (distinct) $i_1, i_2, i_3 \in [k]$ uniformly at random, send the three queries $x, y, x + y$ to the players $\mathcal{P}_{i_1}, \mathcal{P}_{i_2}, \mathcal{P}_{i_3}$ respectively, and check that $\mathcal{P}_{i_1}(x) + \mathcal{P}_{i_2}(y) = \mathcal{P}_{i_3}(x + y)$.

We prove that if $\mathcal{P}$ *always* passes the linearity test, then there exists a quasi-distribution $\mathcal{L}$ over linear functions that matches $\mathcal{P}$.

We first argue that $\mathcal{P}$ must be (almost) *symmetric*, that is, $\mathcal{P}$'s answers depend only on the set of asked queries but not also on which players answer these queries. In more detail, we show that, for every subset $I \subseteq [k]$ of $|I| = k - 1$ players, it holds that $\Pr[\mathcal{P}(\vec{x}) = \vec{b}] = \Pr[\mathcal{P}(\pi(\vec{x})) = \pi(\vec{b})]$ for every permutation $\pi \colon I \to I$, inputs $\vec{x} = (x_i)_{i \in I} \in (\{0,1\}^n)^I$, and answers $\vec{b} = (b_i)_{i \in I} \in \{0,1\}^I$.

We then define a $(k-1)$-non-signaling function $\mathcal{F}$ that matches $k - 1$ players of $\mathcal{P}$ in the natural way (we define $\Pr[\mathcal{F}(x_1) = b_1, \ldots, \mathcal{F}(x_{k-1}) = b_{k-1}]$ to be $\Pr[\mathcal{P}_1(x_1) = b_1, \ldots, \mathcal{P}_{k-1}(x_{k-1}) = b_k]$). By the aforementioned symmetry of $\mathcal{P}$, it does not matter which $k - 1$ players we use to define $\mathcal{F}$.

We then argue that $\mathcal{F}$ always passes the linearity test. Our earlier results imply that there exists a quasi-distribution $\mathcal{L}$ over linear functions that matches $\mathcal{F}$ on all subsets of at most $k - 2$ queries. By definition of $\mathcal{F}$ this implies that $\mathcal{L}$ *also* matches the players $\mathcal{P}_1, \ldots, \mathcal{P}_{k-2}$, and, using the symmetry of $\mathcal{P}$, we conclude that $\mathcal{L}$ also matches *every* subset of $k - 2$ players.

We now relax the assumption that $\mathcal{P}$ passes the linearity test with probability 1 to merely that it passes the test with probability $1 - \varepsilon$ for a small enough $\varepsilon > 0$.

Similarly to the case of non-signaling functions, we define a self-correction $\hat{\mathcal{P}}$ of $\mathcal{P}$ in the natural way: it is a $\hat{k}$-non-signaling player (for $\hat{k} := k/2$) that, given a query $(x_1, \ldots, x_{\hat{k}}) \in \{0,1\}^{\hat{k} \times n}$, samples $w_1, \ldots, w_{\hat{k}} \in \{0,1\}^n$ and a permutation $\pi\colon [k] \to [k]$ uniformly at random, and answers each x_i with $\mathcal{P}_{\pi(2i)}(x_i + w_i) + \mathcal{P}_{\pi(2i+1)}(w_i)$.

We show that $\hat{\mathcal{P}}$ is (fully) symmetric and that, for every $x, y \in \{0,1\}^n$ and distinct $i_1, i_2, i_3 \in [\hat{k}]$, $\Pr[\hat{\mathcal{P}}_{i_1}(x) + \hat{\mathcal{P}}_{i_2}(y) = \hat{\mathcal{P}}_{i_3}(x+y)] > 1 - \hat{\varepsilon}$ for $\hat{\varepsilon} := 4\varepsilon$. This is analogous to the average-case-to-worst-case statement that we showed for non-signaling functions. We define a $\hat{k}$-non-signaling function $\hat{\mathcal{F}}$ that matches $\hat{\mathcal{P}}$ similarly to the above (by letting $\Pr[\hat{\mathcal{F}}(x_1) = b_1, \ldots, \hat{\mathcal{F}}(x_{\hat{k}}) = b_{\hat{k}}] := \Pr[\hat{\mathcal{P}}_1(x_1) = b_1, \ldots, \hat{\mathcal{P}}_{\hat{k}}(x_{\hat{k}}) = b_{\hat{k}}]$), and show that it satisfies the analogous worst-case property, that is, $\Pr[\hat{\mathcal{F}}(x) + \hat{\mathcal{F}}(y) = \hat{\mathcal{F}}(x+y)] \geq 1 - \hat{\varepsilon}$ for every $x, y \in \{0,1\}^n$. Our earlier results imply that there exists a quasi-distribution $\mathcal{L}$ over linear functions that is close to $\hat{\mathcal{F}}$, and thus also close to $\hat{\mathcal{P}}$.

See full version for details.

4 Discussion and open problems

The study of non-signaling strategies in Physics is motivated by the goal of understanding the power and limitations of Nature's apparent non-locality [35, 43, 40, 41, 8]. Prior work has explored many topics, including the inter-convertibility between quantum strategies and non-signaling strategies [19, 11, 10, 31, 17]; communication complexity with non-signaling strategies [51, 16]; non-local computation [36]; using non-signaling strategies to achieve key distribution, oblivious transfer, and bit commitments [9, 53, 18, 49, 48]; and many others [38, 27, 20].

More recently, researchers have established connections with Complexity Theory and Cryptography. Property Testing against non-signaling strategies, the subject of our work, is likely to lead to a deeper understanding of these.

4.1 Powers and limitations of non-local strategies

Understanding the computational complexity of computing or approximating the value of certain classes of games is a fundamental problem in Complexity Theory. Games are typically phrased in terms of one or more *non-communicating* players that interact with a probabilistic polynomial-time Referee (with polynomial randomness), who decides at the end of the game if the players win or not. The complexity of these games is well-understood.

- Results on *Interactive Proofs* (IPs) [37, 46] imply that approximating the value of single-player games is PSPACE-complete, when given enough rounds.
- Results on *Multi-prover Interactive Proofs* (MIPs) [7] imply that approximating the value of multi-player games is NEXP-complete, even with only two players.
- Results on *Probabilistically Checkable Proofs* (PCPs) [6, 24, 4, 3] imply that, if the player's strategy is non-adaptive (the player merely answers queries from the Referee) then approximating the game's value is NEXP-complete, even if the Referee asks only a constant number of queries and receives answers over a constant-size alphabet.

However, if the players can use any non-signaling strategy to win the game, *much less is known.*

If there are only *two* players, then approximating the game's value is PSPACE-complete [29, 28]. If the game has k players then its value can be computed in time $\mathsf{poly}(2^{kr}, |\Sigma|^k)$, where r is the Referee's randomness complexity and Σ is each player's answer's alphabet [23], which means that this computation lies in EXP. This is *very unlike* the case of non-communicating players.

However, hardness results for this problem in the case of three or more players have been elusive. Recently, Kalai, Raz, and Rothblum [32, 34] established EXP-hardness for the case of polynomially-many provers, via a reduction from deterministic-time languages.

▶ **Theorem 4.1** ([32, 34]). *Let L be a language decidable in time $T\colon \mathbb{N} \to \mathbb{N}$. There exists a constant $c > 0$ such that, for any function $\lambda\colon \mathbb{N} \to \mathbb{N}$ with $\lambda \geq \log^c T$, L has a $(\lambda \log^c T)$-prover MIP with soundness error $2^{-\lambda}$ against non-signaling players. The verifier runs in time $n\lambda^2 \log^c T$ and the provers in time $\mathsf{poly}(T, \lambda)$; each query and answer consists of $\lambda \log^c T$ bits.*

The above theorem is proved by constructing a PCP verifier that is secure against non-signaling functions (Definition 6.1), which can then be compiled into an MIP verifier that is secure against non-signaling players. The proof is a technical tour-de-force showing that a modification of the "classical" PCP verifier in [7, 6] is secure against non-signaling functions.

The huge gap between the EXP-completeness for polynomially-many provers and the PSPACE-completeness for two provers motivates a natural question:

> *Is there a non-signaling analogue of the PCP Theorem?* I.e., does EXP have $O(1)$-query PCPs over a $O(1)$-size alphabet that are secure against non-signaling functions? (Equivalently, $O(1)$-prover MIPs over a $O(1)$-size alphabet that are secure against non-signaling players?)

We believe that initiating a study of Property Testing against non-signaling strategies will drive progress on this question. In particular, linearity testing is one of the ingredients of the (classical) PCP Theorem, and linearity testing against non-signaling strategies may be a good place to start.

We also believe that Property Testing against non-signaling strategies may play a significant *simplifying role*, which could itself drive progress on this and other questions. Indeed, the analysis of classical PCP constructions (including [7, 6]) is carried out in two conceptually simple steps: first argue soundness assuming that the PCP is a low-degree function, and then rely on low-degree testing and self-correction to ensure that the PCP is close to a low-degree function [45, 44, 5]. The study of this latter step as a standalone problem in the area of Property Testing has enabled much progress on PCP research. In contrast, while the analysis in [32, 34] does analyze low-degree tests by proving certain average-case-to-worst-case statements, it *does not prove any local-to-global phenomena* for the property of "low-degreeness".

We prove a first local-to-global phenomenon for Property Testing against non-signaling strategies. However, whether Property Testing is feasible beyond the case of linearity testing (our focus) and whether it plays a beneficial and simplifying role in PCP research are fascinating open problems.

4.2 Hardness of approximation

Feige et al. [24] showed a fundamental connection between MIPs/PCPs and the hardness of approximating values of constraint satisfaction problems. Kalai, Raz, and Regev [33] recently established a similar connection, this time between *non-signaling* MIPs/PCPs and the hardness of approximating values of *linear programs*. While the first connection considers approximation algorithms that are bounded in time, the second connection considers approximation algorithms that are bounded in *space*. We recall [33]'s result and its relation to our results.

▶ **Theorem 4.2** ([33]). *Let L be a language with a 1-round k-prover MIP with soundness error ϵ against non-signaling players in which:*

(i) *the verifier has time complexity T, space complexity S, and randomness complexity r;*
(ii) *the prover's answers are symbols in Σ.*

Then there is a family of polyhedra $\{H_n\}_{n\in\mathbb{N}}$ and a $\mathsf{poly}(2^{kr}, |\Sigma|^k, T)$-time $\mathsf{poly}(k, r, S)$-space reduction $\mathscr{R}$ such that:

(i) *For every instance $x \in \{0,1\}^*$, $\mathscr{R}(x)$ is a linear program with polyhedron $H_{|x|}$ and with $\mathsf{poly}(2^{kr}, |\Sigma|^k)$ variables and constraints.*
(ii) *If $x \in L$, then the value of the linear program $\mathscr{R}(x)$ is 1.*
(iii) *If $x \notin L$, then the value of the linear program $\mathscr{R}(x)$ at most ϵ.*

The above result, when combined with the non-signaling MIPs for deterministic-time languages of [32, 34] (see Section 4.1), implies that a $2^{\log^{o(1)}(n)}$-space approximation algorithm for linear programming is unlikely, even when given unbounded computation based on the polyhedron. (Since that would imply, in particular, that every problem in P can be solved in $2^{\log^{o(1)}(n)}$-space.)

The above conclusion, however, appears sub-optimal because both 2^{kr} and $|\Sigma|^k$ are super-polynomial in the construction of [32, 34]. Ideally, we would like a construction where $r = O(\log n)$ and $k = O(1)$, which again (as discussed in Section 4.1) leads to the question of whether there is a non-signaling analogue of the PCP Theorem. We conjecture that the study of Property Testing against non-signaling strategies is again very relevant.

4.3 One-round delegation of computation

Delegation of computation is a fundamental goal in Cryptography that involves designing protocols that enable a weak verifier to outsource expensive computations to a powerful but untrusted prover.

A key efficiency measure is round complexity (the number of back-and-forth messages between the verifier and prover). Aiello et al. [1] suggested a cryptographic method to transform any 1-round MIP into a 1-round delegation protocol, but did not provide a proof of security. Later on, Dwork et al. [23] showed that this method is not secure in the general case, by exhibiting a 1-round MIP for which the transformation yields a delegation protocol that can be fooled.

Nevertheless, Kalai, Raz, and Rothblum [32] proved that if the 1-round MIP used in the method is sound against non-signaling players then the resulting delegation protocol *cannot* be fooled (namely, is secure). More precisely, the 1-round MIP must be sound not only against all players that are non-signaling but also against all players that are *almost* non-signaling (see full version for details), where "almost" denotes a certain parameter that depends on the security reduction.

By invoking this method on the MIP of [32, 34] (which *is* secure against almost non-signaling players), one obtains a delegation protocol for all polynomial-time functions in which the prover runs in polynomial time and the verifier in polylogarithmic time.

Yet, the seemingly sub-optimal parameters of the MIP of [32, 34] suggest that there is room to improve efficiency by invoking the method on more efficient MIPs. For example:

> *Is there an <u>almost</u> non-signaling analogue of the PCP Theorem?*
> Namely, does EXP have $O(1)$-query PCPs (equivalently, $O(1)$-prover MIPs) over a $O(1)$-size alphabet that are secure against almost non-signaling strategies?

The study of Property Testing against almost non-signaling strategies is likely a first step, and our work establishes first results for *exact* non-signaling strategies.

▶ **Remark 4.3** (extension to almost non-signaling). *While almost non-signaling strategies are not our focus, in this paper we do show that almost non-signaling strategies are not outside the reach of tools that we use. Concretely, we show that every almost non-signaling function is "reasonably close" to a corresponding (exact) non-signaling function. The proof of this statement uses Fourier analysis, and the intuition behind it is similar to how almost-feasible solutions to Sherali–Adams relaxations are "smoothened"into feasible ones [42]. The generic lemma enables us, for example, to extend Theorem 2.4 to the case of almost non-signaling strategies. Whether a whitebox analysis of linearity testing against almost non-signaling strategies can improve upon such a blackbox extension remains an interesting open problem. See full version for details.*

5 Preliminaries

For a finite domain D, we denote by U_D the set of all boolean functions $f\colon D \to \{0,1\}$; when D is clear from context, we may omit the subscript in U_D. When $D = \{0,1\}^n$, a function $f \in U_{\{0,1\}^n}$ is *linear* if $f(x) + f(y) = f(x+y)$ for all $x, y \in \{0,1\}^n$; LIN is the set of all such linear functions.

5.1 Fourier analysis of boolean functions

We use standard notation for Fourier analysis of boolean functions (see [39] for more details). For a domain D of size N, we consider functions $f\colon \{0,1\}^D \to \mathbb{R}$. The inner product of two functions $g_1, g_2\colon \{0,1\}^D \to \mathbb{R}$ is $\langle g_1, g_2\rangle := \frac{1}{2^N}\sum_{x\in\{0,1\}^D} g_1(x)g_2(x)$. For a subset $T \subseteq D$, $\chi_T\colon \{0,1\}^D \to \mathbb{R}$ is the parity function $\chi_T(x) = (-1)^{\sum_{i\in T} x_i}$. It is not hard to verify that the set of functions $\{\chi_T\}_{T\subseteq D}$ is an orthonormal basis of the space of all functions from $\{0,1\}^D$ to $\mathbb{R}$. In particular, every function $f\colon \{0,1\}^D \to \mathbb{R}$ can be written as

$$f(\cdot) = \sum_{T\subseteq D} \widehat{f}(T)\chi_T(\cdot) \ ,$$

where $\widehat{f}(T) = \langle f, \chi_T\rangle = \frac{1}{2^N}\sum_{x\in\{0,1\}^D} f(x)\chi_T(x)$. In particular, by Parseval's identity for any two functions $f, g\colon \{0,1\}^D \to \mathbb{R}$ we have

$$\frac{1}{2^N}\sum_{x\in\{0,1\}^D} f(x)g(x) = \sum_{T\subseteq D} \widehat{f}(T)\widehat{g}(T) \ ,$$

which implies Plancherel's identity

$$\frac{1}{2^N}\sum_{x\in\{0,1\}^D} f(x)^2 = \sum_{T\subseteq D} \widehat{f}(T)^2 \ .$$

For a set $E \subseteq \{0,1\}^s$, its indicator function $\mathbf{1}_E\colon \{0,1\}^s \to \{0,1\}$ is defined as

$$\mathbf{1}_E = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{otherwise} \end{cases} \ .$$

Note that by Plancherel's identity we have $\sum_{T\subseteq[s]} \widehat{\mathbf{1}_E}(T)^2 = \mathbb{E}[\mathbf{1}_E] = \frac{|E|}{2^s}$. In particular, this implies $\|\widehat{\mathbf{1}_E}\|_1 = \sum_{T\subseteq[s]} |\widehat{\mathbf{1}_E}(T)| \le \sqrt{\sum_{T\subseteq[s]} \widehat{\mathbf{1}_E}(T)^2} \cdot \sqrt{\sum_{T\subseteq[s]} 1} \le \sqrt{\frac{|E|}{2^s}} \cdot 2^{s/2} = \sqrt{|E|}$.

5.2 Expressing boolean events as sums of parities

We state two lemmas that express the probability of certain events as probabilities about the *parities* of related events.

▶ **Lemma 5.1.** *Let $X_1, \ldots, X_s$ be boolean random variables. Then, for every event $E \subseteq \{0,1\}^s$ it holds that*

$$\Pr[(X_1 \ldots, X_s) \in E] = \sum_{T \subseteq [s]} c_T \cdot \Pr\left[\sum_{i \in T} X_i = 0\right] \ ,$$

where $c_\emptyset = 2 \cdot \widehat{\mathbf{1}_E}(\emptyset) - \mathbf{1}_E(\vec{0})$, and $c_T = 2 \cdot \widehat{\mathbf{1}_E}(T)$ for all $T \neq \emptyset$. In particular, c_T's depend only on E and $\sum_{T \subseteq [s]} |c_T| \leq 3\|\widehat{\mathbf{1}_E}\|_1 \leq 3\sqrt{|E|}$.

▶ **Corollary 5.2.** *Let $X_1, \ldots, X_s$ be boolean random variables. Then, for every $\vec{b} = (b_1, \ldots, b_s)$ in $\{0,1\}^s$ it holds that*

$$\Pr[X_1 = b_1, \ldots, X_s = b_s] = -1 + \frac{1}{2^{s-1}} \sum_{T \subseteq [s]} \Pr\left[\sum_{i \in T} X_i = \sum_{i \in T} b_i\right] \ .$$

Proof of Lemma 5.1. Define $p\colon \{0,1\}^s \to \mathbb{R}$ as $p(\vec{a}) = \Pr[X_1 = a_1, \ldots, X_s = a_s]$, and write $p = \sum_{T \subseteq [s]} \hat{p}(T) \cdot \chi_T$. We have

$$\begin{aligned}
&\hat{p}(T) = \mathbb{E}[p(\vec{a}) \cdot \chi_T(\vec{a})] \\
&= \frac{1}{2^s} \left(\sum_{\vec{a}: \sum_{i \in T} \vec{a}_i = 0} p(\vec{a}) - \sum_{\vec{a}: \sum_{i \in T} \vec{a}_i = 1} p(\vec{a}) \right) \\
&= \frac{1}{2^s} \left(2 \sum_{\vec{a}: \sum_{i \in T} \vec{a}_i = 0} p(\vec{a}) - 1 \right) \\
&= \frac{1}{2^s} \left(2 \Pr\left[\sum_{i \in T} X_i = 0\right] - 1 \right)
\end{aligned}$$

Let $E \subseteq \{0,1\}^s$ be an event, and let $\mathbf{1}_E\colon \{0,1\}^s \to \{0,1\}$ be its indicator function. Then, by Parseval's identity we have

$$\Pr[(X_1 \ldots, X_s) \in E] = \sum_{\vec{a} \in \{0,1\}^s} p(\vec{a}) \cdot \mathbf{1}_E(\vec{a}) = 2^s \cdot \sum_{T \subseteq [s]} \hat{p}(T) \cdot \widehat{\mathbf{1}_E}(T) \ .$$

By plugging in the formula $\hat{p}(T) = \frac{1}{2^s}\left(2\Pr[\sum_{i \in T} X_i = 0] - 1\right)$, and using $\Pr[\sum_{i \in \emptyset} X_i = 0] = 1$ we get

$$\Pr[(X_1 \ldots, X_s) \in E] = \sum_{T \subseteq [s]} \left(2 \Pr\left[\sum_{i \in T} X_i = 0\right] - 1 \right) \cdot \widehat{\mathbf{1}_E}(T) = \sum_{T \subseteq [s]} c_T \cdot \Pr\left[\sum_{i \in T} X_i = 0\right] \ ,$$

where $c_\emptyset = 2 \cdot \widehat{\mathbf{1}_E}(\emptyset) - \sum_{T \subseteq [s]} \widehat{\mathbf{1}_E}(T)$, and $c_T = 2 \cdot \widehat{\mathbf{1}_E}(T)$ for all $T \neq \emptyset$. Since $\mathbf{1}_E(\cdot) = \sum_{T \subseteq [s]} \widehat{\mathbf{1}_E}(T)\chi_T(\cdot)$, it follows that $\mathbf{1}_E(\vec{0}) = \sum_{T \subseteq [s]} \widehat{\mathbf{1}_E}(T)$, as required.

Thus, by the argument in Section 5.1 we have $\sum_{T \subseteq [s]} |c_T| \leq 3 \sum_{T \subseteq [s]} |\widehat{\mathbf{1}_E}(T)| \leq 3\sqrt{|E|}$. ◀

Proof of Corollary 5.2. Let $E = \{\vec{b}\}$ be the singleton event. It is easy to verify that $\widehat{\mathbf{1}_E}(T) = (-1)^{\sum_{i\in T} b_i} \cdot 2^{-s}$. Therefore, by Lemma 5.1 we have

$$\Pr[X_1 = b_1, \ldots, X_s = b_s] = \sum_{T\subseteq[s]} c_T \cdot \Pr\left[\sum_{i\in T} X_i = 0\right] \ ,$$

where $c_\emptyset = 2 \cdot \widehat{\mathbf{1}_E}(\emptyset) - \mathbf{1}_E(\vec{0}) = \frac{1}{2^{s-1}} - \mathbf{1}_E(\vec{0})$, and $c_T = (-1)^{\sum_{i\in T} b_i} \cdot 2^{-s+1}$ for all $T \neq \emptyset$. By substituting $\Pr\left[\sum_{i\in T} X_i = 0\right]$ with $1 - \Pr\left[\sum_{i\in T} X_i = 1\right]$ for all $T \subseteq [s]$ such that $\sum_{i\in T} b_i = 1$ we get

$$\begin{aligned}
&\Pr[X_1 = b_1, \ldots, X_s = b_s] = \sum_{T\subseteq[s]} c_T \cdot \Pr\left[\sum_{i\in T} X_i = 0\right] \\
&= \left(-\mathbf{1}_E(0) - \sum_{T:\sum_{i\in T} b_i = 1} \frac{1}{2^{s-1}}\right) + \frac{1}{2^{s-1}} \sum_{T\subseteq[s]} \Pr\left[\sum_{i\in T} X_i = \sum_{i\in T} b_i\right] \\
&= -1 + \frac{1}{2^{s-1}} \sum_{T\subseteq[s]} \Pr\left[\sum_{i\in T} X_i = \sum_{i\in T} b_i\right] \ ,
\end{aligned}$$

as required. ◀

5.3 A linear system

Below we prove that a certain linear system of equations, which we will use later, has a unique solution. This linear system is the inverse of the Hadamard–Walsh matrix.

▶ **Lemma 5.3.** *For every positive integer n and real numbers $\{c_\beta\}_{\beta\in\{0,1\}^n}$, the system of 2^n linear equations over $\mathbb{R}$ in 2^n variables $\{z_\alpha\}_{\alpha\in\{0,1\}^n}$ given by*

$$\left\{ \forall \beta \in \{0,1\}^n \sum_{\substack{\alpha\in\{0,1\}^n \\ \text{s.t. } \langle\alpha,\beta\rangle=0}} z_\alpha = c_\beta \right\}$$

has a unique solution.

Proof. Let A be the $2^n \times 2^n$ boolean matrix corresponding to the system of linear equations, that is, such that $Az = c$. Note that the (β,α)-th entry of A is equal to $1 - \langle\alpha,\beta\rangle$, and in particular, the row in A corresponding to $\beta = 0^n$ is the all-ones row. Define H to be the matrix obtained from A by performing the following elementary row operations: for every $\beta \neq 0^n$, multiply row β by 2 and then subtract the all-ones row (corresponding to $\beta = 0^n$).

Note that the (β,α)-th entry of H is equal to $(-1)^{\langle\alpha,\beta\rangle}$. (The matrix H is sometimes called the Hadamard–Walsh matrix.) Indeed, this holds trivially for the row $\beta = 0^n$ as $H_{\beta,\alpha} = (-1)^{\langle\alpha,0^n\rangle} = 1$, and for $\beta \neq 0^n$ we have $H_{\beta,\alpha} = 2(1 - \langle\alpha,\beta\rangle) - 1 = 1 - 2\langle\alpha,\beta\rangle = (-1)^{\langle\alpha,\beta\rangle}$. Since H was obtained from A by performing elementary row operations, A is invertible if and only if H is invertible. Observe that H is indeed invertible because the rows of H are mutually orthogonal since for every two distinct β and γ in $\{0,1\}^n$ it holds that

$$\langle \text{row } \beta, \text{row } \gamma\rangle = \sum_\alpha (-1)^{\langle\alpha,\beta\rangle}(-1)^{\langle\alpha,\gamma\rangle} = \sum_\alpha (-1)^{\langle\alpha,\beta\rangle+\langle\alpha,\gamma\rangle} = \sum_\alpha (-1)^{\langle\alpha,\beta+\gamma\rangle} = 0 \ ,$$

where the last equality holds because $\beta + \gamma \neq 0^n$. ◀

6 Non-signaling functions

We define *non-signaling functions*, introduce useful notation for them, and prove a simple lemma about them. The notions described here are used throughout the paper.

▶ **Definition 6.1** (non-signaling functions). A ***k-non-signaling (boolean) function*** over a finite domain D is a collection $\mathcal{F} = \{\mathcal{F}_S\}_{S\subseteq D, |S|\leq k}$ where

(i) each $\mathcal{F}_S$ is a distribution over functions $f\colon S \to \{0,1\}$, and

(ii) for every two subsets S and T each of size at most k, the restrictions of $\mathcal{F}_S$ and $\mathcal{F}_T$ to $S \cap T$ are equal as distributions.

(If $S = \emptyset$ then $\mathcal{F}_S$ always outputs the empty string.)

Given a set $S \subseteq D$ of size $|S| \leq k$ and a string $\vec{b} \in \{0,1\}^S$, we define

$$\Pr[\,\mathcal{F}(S) = \vec{b}\,] := \Pr_{f\leftarrow\mathcal{F}_S}[\,f(S) = \vec{b}\,] \ .$$

The non-signaling property in this notation is the following: for every two subsets $S, T \subseteq D$ of sizes $|S|, |T| \leq k$ and every string $\vec{b} \in \{0,1\}^{S\cap T}$, $\Pr[\,\mathcal{F}(S)|_{S\cap T} = \vec{b}\,] = \Pr[\,\mathcal{F}(T)|_{S\cap T} = \vec{b}\,]$.

Sometimes it is more convenient to consider a *vector* of inputs (rather than a *set*), and so we define notation for this case. Given a vector $\langle x_1, \ldots, x_s\rangle$ with entries in D and a vector $\langle b_1, \ldots, b_s\rangle$ with entries in $\{0,1\}$ (with $s \in \{1, \ldots, k\}$), we define $\Pr[\mathcal{F}(\langle x_1, \ldots, x_s\rangle) = \langle b_1, \ldots, b_s\rangle]$ and $\Pr[\mathcal{F}(x_1) = b_1, \ldots, \mathcal{F}(x_s) = b_s]$ to be the probability

$$\Pr_{f\leftarrow\mathcal{F}_{\{x_1,\ldots,x_s\}}}[\,f(x_1) = b_1, \ldots, f(x_s) = b_s\,] \ .$$

Note that $\{x_1, \ldots, x_s\}$ is an unordered set and its size may be less than s, because the entries of the vector $\langle x_1, \ldots, x_s\rangle$ may not be distinct. We abuse notation and still use symbols such as S and $\vec{b}$ to denote vectors as above. We stress that we use an ordering on S merely to match each element of S to the corresponding element in $\vec{b}$; the event remains unchanged if one permutes the entries of S and $\vec{b}$ according to the same permutation.

▶ **Remark 6.2** (Sherali–Adams hierarchy). *We note that k-non-signaling functions are solutions to the linear program arising from the k-relaxation in the Sherali–Adams hierarchy [47]. The variables are of the form $X_{S,\vec{b}}$ (for all $S \subseteq D$ of size at most k and $\vec{b} \in \{0,1\}^S$) and express $\Pr[\,\mathcal{F}(S) = \vec{b}\,]$. Consistency across subsets S and T is expressed using the natural linear constraints.*[2]

We conclude with a useful lemma.

▶ **Lemma 6.3.** *Let $\mathcal{F}$ be a k-non-signaling function over a domain D, let S_1, S_2 be subsets of D with $|S_1 \cup S_2| \leq k$, and let $g_1\colon \{0,1\}^{S_1} \to \{0,1\}^r$ and $g_2\colon \{0,1\}^{S_2} \to \{0,1\}^r$ be functions. If $\Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = g_2(\mathcal{F}(S_2))\,] \geq 1 - \varepsilon$, then for every $\vec{b} \in \{0,1\}^r$ it holds that*

$$\left|\Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = \vec{b}\,] - \Pr_{\mathcal{F}}[\,g_2(\mathcal{F}(S_2)) = \vec{b}\,]\right| \leq \varepsilon \ .$$

In particular, if $\varepsilon = 0$ then $\Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = \vec{b}\,] = \Pr_{\mathcal{F}}[\,g_2(\mathcal{F}(S_2)) = \vec{b}\,]$ for every $\vec{b} \in \{0,1\}^r$.

[2] In fact it suffices to only have variables of the form $X_{S,1^s}$ as all other probabilities can be computed from these.

Proof. By direct computation:

$$\begin{aligned}
&\left|\Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = \vec{b}\,] - \Pr_{\mathcal{F}}[\,g_2(\mathcal{F}(S_2)) = \vec{b}\,]\right| \\
&= \Big|\Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = \vec{b} \wedge g_2(\mathcal{F}(S_2)) = \vec{b}\,] + \Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = \vec{b} \wedge g_2(\mathcal{F}(S_2)) \neq \vec{b}\,] \\
&\quad - \Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = \vec{b} \wedge g_2(\mathcal{F}(S_2)) = \vec{b}\,] - \Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) \neq \vec{b} \wedge g_2(\mathcal{F}(S_2)) = \vec{b}\,]\Big| \\
&= \left|\Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = \vec{b} \wedge g_2(\mathcal{F}(S_2)) \neq \vec{b}\,] - \Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) \neq \vec{b} \wedge g_2(\mathcal{F}(S_2)) = \vec{b}\,]\right| \\
&\leq \Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) = \vec{b} \wedge g_2(\mathcal{F}(S_2)) \neq \vec{b}\,] + \Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) \neq \vec{b} \wedge g_2(\mathcal{F}(S_2)) = \vec{b}\,] \\
&\leq \Pr_{\mathcal{F}}[\,g_1(\mathcal{F}(S_1)) \neq g_2(\mathcal{F}(S_2))\,] \leq \varepsilon \ .
\end{aligned}$$

Note that we are implicitly using the fact that $|S_1 \cup S_2| \leq k$ whenever we have S_1 and S_2 in the same probability event because we are querying $\mathcal{F}$ on all inputs in $S_1 \cup S_2$ at once. ◀

7 Quasi-distributions

A quasi-distribution extends the notion of a probability distribution by allowing probabilities to be negative, and is the main tool that we use to analyze non-signaling functions.

▶ **Definition 7.1** (quasi-distributions)**.** Let D be a finite domain, and denote by U_D the set of all boolean functions of the form $f\colon D \to \{0,1\}$. A ***quasi-distribution*** $\mathcal{Q}$ over a subset $G \subseteq U_D$ is a set of real numbers $\{q_f\}_{f \in U_D}$ such that $\sum_{f \in U_D} q_f = 1$ and $q_f = 0$ for every $f \notin G$.

▶ **Definition 7.2** (quasi-probability)**.** Given a quasi-distribution $\mathcal{Q} = \{q_f\}_{f \in U_D}$, a subset $S \subseteq D$, and a string $\vec{b} \in \{0,1\}^S$, we define the ***quasi-probability*** of the event "$\mathcal{Q}(S) = \vec{b}$" to be the following (possibly negative) real number

$$\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}\,] := \sum_{f \in U_D \ s.t. \ f(S) = \vec{b}} q_f \ .$$

As in the case of non-signaling functions, it is sometimes more convenient to consider a *vector* of inputs rather than a *set*. Given a vector $\langle x_1, \ldots, x_s \rangle$ with entries in D and a vector $\langle b_1, \ldots, b_s \rangle$ with entries in $\{0,1\}$, we define $\Pr[\mathcal{Q}(\langle x_1, \ldots, x_s \rangle) = \langle b_1, \ldots, b_s \rangle]$ and $\Pr[\mathcal{Q}(x_1) = b_1, \ldots, \mathcal{Q}(x_s) = b_s]$ to be the (possibly negative) real number $\sum_{f \in U_D \ s.t. \ \forall i \ f(x_i) = b_i} q_f$. We abuse notation and still use symbols such as S and $\vec{b}$ to denote vectors as above.

Since a quasi-distribution $\mathcal{Q}$ is defined by its weights $q = (q_f)_{f \in U_D}$, we can view $\mathcal{Q}$ as a function from $\{0,1\}^D$ to $\mathbb{R}$, where we identify a function $f\colon D \to \{0,1\}$ with the corresponding vector in $\{0,1\}^D$ and $q(f)$ with q_f. In particular, we can write $q(\cdot) = \sum_{T \subseteq D} \widehat{q}(T)\chi_T(\cdot)$, where $\chi_T(f) = (-1)^{\sum_{x \in T} f(x)}$, and $\widehat{q}(T) = \langle q, \chi_T \rangle = \frac{1}{2^{|D|}} \sum_{f\colon D \to \{0,1\}} q(f)\chi_T(f)$.

The following lemma is an analogue of Lemma 5.1 for quasi-distributions.

▶ **Lemma 7.3.** *Let $\mathcal{Q} = (q_f)_f$ be a quasi-distribution, $S = \langle x_1, \ldots, x_s \rangle$ a vector with entries in $\{0,1\}^n$. Then, for every event $E \in \{0,1\}^s$ it holds that*

$$\sum_{f : f(S) \in E} q_f = \sum_{T \subseteq [s]} c_T \cdot \widetilde{\Pr}\left[\sum_{i \in T} \mathcal{Q}(x_i) = 0\right] = \sum_{T \subseteq [s]} c_T \cdot \left(\sum_{f : \sum_{i \in T} f(x_i) = 0} q_f\right) \ ,$$

where $c_\emptyset = 2 \cdot \widehat{\mathbf{1}_E}(\emptyset) - \mathbf{1}_E(\vec{0})$, and $c_T = 2 \cdot \widehat{\mathbf{1}_E}(T)$ for all $T \neq \emptyset$.

The proof of the lemma is immediate from the proof of Lemma 5.1, since the proof only uses the fact that probabilities add up to 1, which also holds for quasi-probabilities.

▶ **Definition 7.4** (locality)**.** Let D be a finite domain of size N. For $1 \leq \ell \leq N$ a quasi-distribution $\mathcal{Q}$ over U_D is ***ℓ-local*** if for every subset $S \subseteq D$ of size $|S| \leq \ell$ and string $\vec{b} \in \{0,1\}^S$,

$$\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] \in [0,1] \ .$$

For completeness, we also say that all quasi-distributions are 0-local.

If $\mathcal{Q}$ is ℓ-local, then for every subset $S \subseteq D$ of size $|S| \leq \ell$, we may view $\mathcal{Q}(S)$ as a probability distribution over $\{0,1\}^S$. If $\mathcal{Q}$ is ℓ-local then it is s-local for every $s \in \{0,1,\ldots,\ell\}$.

For $\mathcal{Q}$ to be ℓ-local, it suffices for all relevant $\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}]$ to be non-negative (as opposed to be in $[0,1]$). This is because $\sum_f q_f = 1$, so that $\sum_{\vec{b} \in \{0,1\}^S} \widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] = 1$ and, if all terms in this sum are non-negative, then we can deduce that $\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] \leq 1$ for every $\vec{b}$.

▶ **Definition 7.5** (statistical distance)**.** Given a finite domain D and an integer $\ell \in \{1,\ldots,|D|\}$, the ***$\Delta_\ell$-distance*** between two quasi-distributions $\mathcal{Q}$ and $\mathcal{Q}'$ is

$$\Delta_\ell(\mathcal{Q},\mathcal{Q}') := \max_{S \subseteq D\,,\,|S| \leq \ell} \Delta(\mathcal{Q}_S, \mathcal{Q}'_S) \ ,$$

where $\Delta(\mathcal{Q}_S,\mathcal{Q}'_S) := \max_{E \subseteq \{0,1\}^S} \left|\widetilde{\Pr}[\mathcal{Q}(S) \in E] - \widetilde{\Pr}[\mathcal{Q}'(S) \in E]\right|$.

We say that $\mathcal{Q}$ and $\mathcal{Q}'$ are ***ε-close in the Δ_ℓ-distance*** if $\Delta_\ell(\mathcal{Q},\mathcal{Q}') \leq \varepsilon$; else, they are ε-far.

▶ **Remark 7.6** (distance for non-signaling functions)**.** *The definition of Δ_ℓ-distance naturally extends to defining distances between k-non-signaling functions, as well as between quasi-distributions and k-non-signaling functions, provided that $\ell \leq k$.*

The notion above generalizes the standard notion of statistical (total variation) distance: if $\mathcal{Q}$ and $\mathcal{Q}'$ are *distributions* then their $\Delta_{|D|}$-distance equals their statistical distance. Also note that if $\mathcal{Q}$ and $\mathcal{Q}'$ are ℓ-local quasi-distributions then their Δ_ℓ-distance equals the maximum statistical distance, across all subsets $S \subseteq D$ with $|S| \leq \ell$, between the two *distributions* $\mathcal{Q}_S$ and $\mathcal{Q}'_S$ — in particular this means that any experiment that queries exactly one set of size at most ℓ cannot distinguish between the two quasi-distributions with probability greater than $\Delta_\ell(\mathcal{Q},\mathcal{Q}')$.

We stress that $\Delta_\ell(\mathcal{Q},\mathcal{Q}') = 0$ does *not* necessarily mean that $\mathcal{Q} = \mathcal{Q}'$! In fact, it is possible to have $\Delta_\ell(\mathcal{Q},\mathcal{Q}') = 0$ while $\sum_{f \in U} |q_f - q'_f|$ is arbitrarily large. We also remark that the Δ_ℓ-distance is not necessarily upper bounded by 1, and is in general unbounded.

▶ **Definition 7.7** (approximate locality)**.** Given a finite domain D, an integer $\ell \in \{1,\ldots,|D|\}$, and a real number $\varepsilon \geq 0$, a quasi-distribution $\mathcal{Q}$ over U_D is ***(ℓ,ε)-local*** if, for every subset $S \subseteq D$ of size $|S| \leq \ell$ and every event $E \subseteq \{0,1\}^S$,

$$\widetilde{\Pr}[\mathcal{Q}(S) \in E] \in [-\varepsilon, 1+\varepsilon] \ .$$

Approximate locality generalizes the notion of (exact) locality as in Definition 7.4. Indeed, note that in Definition 7.4 the condition is point-wise, i.e., $\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] \in [0,1]$ for each $\vec{b} \in \{0,1\}^S$. However, this is in fact equivalent to the event-wise definition, $\widetilde{\Pr}[\mathcal{Q}(S) \in E] \in [0,1]$ for all $E \subseteq \{0,1\}^S$, and hence every ℓ-local quasi-distribution $\mathcal{Q}$ is $(\ell,0)$-local.

Below we discuss the following questions. Given an approximately local quasi-distribution $\mathcal{Q}$, can we find a local quasi-distribution $\mathcal{Q}'$ close to it? Moreover, can we ensure that $\mathcal{Q}'$ "looks like" $\mathcal{Q}$? We show that if $\mathcal{Q}$ is (ℓ, ε)-local and is supported over a set G of functions that is nice in some precise way, then there is an ℓ-local $\mathcal{Q}'$ over G that is close to $\mathcal{Q}$. The proof idea is similar to that of "smoothening" almost-feasible solutions to Sherali–Adams relaxations into feasible ones [42].

▶ **Lemma 7.8.** *Let D be a finite domain, $\ell \in \{1, \ldots, |D|\}$ be an integer, and $\delta > 0$, $\varepsilon \geq 0$ be reals. Let $G \subseteq U_D$ be a set of functions $f \colon D \to \{0,1\}$ such that for all subsets $S \subseteq D$ of size $|S| \leq \ell$ and for all strings $\vec{b} \in \{0,1\}^S$ it holds that $\Pr_{f \leftarrow G}[f(S) = \vec{b}] \in \{0\} \cup [\delta, 1]$, where f is sampled uniformly at random from G. If $\mathcal{Q}$ is a (ℓ, ε)-local quasi-distribution over G, then there exists an ℓ-local quasi-distribution $\mathcal{Q}'$ over G such that $\Delta_\ell(\mathcal{Q}, \mathcal{Q}') \leq (1 + \varepsilon - \delta) \cdot \frac{\varepsilon}{\varepsilon + \delta}$.*

We highlight two notable special cases for the domain $D = \{0,1\}^n$. If $G = U_{\{0,1\}^n}$ (the set of *all* functions), then $\Pr_{f \leftarrow G}[f(S) = \vec{b}] = 2^{-\ell}$. Also, if $G = \mathsf{LIN}$ (the set of all *linear* functions), then for every subset $S \subseteq \{0,1\}^n$ of size at most ℓ and every string $\vec{b} \in \{0,1\}^S$ it holds that $\Pr_{f \leftarrow G}[f(S) = \vec{b}] = 0$ or $\Pr_{f \leftarrow G}[f(S) = \vec{b}] = 2^{-\dim(\mathrm{span}(S))} \geq 2^{-|S|} \geq 2^{-\ell}$. These two cases yield the following corollary.

▶ **Corollary 7.9.** *If $\mathcal{Q}$ is a (ℓ, ε)-local quasi-distribution over $U_{\{0,1\}^n}$ (resp., LIN), then there is an ℓ-local quasi-distribution $\mathcal{Q}'$ over $U_{\{0,1\}^n}$ (resp., LIN) such that $\Delta_\ell(\mathcal{Q}, \mathcal{Q}') \leq \frac{1+\varepsilon-2^{-\ell}}{1+2^\ell\varepsilon} \cdot 2^\ell\varepsilon < 2^\ell\varepsilon$.*

Proof. The hypothesis of Lemma 7.8 holds with $\delta = 2^{-\ell}$. So there exists an ℓ-local quasi-distribution $\mathcal{Q}'$ over $U_{\{0,1\}^n}$ (resp., LIN) such that $\Delta_\ell(\mathcal{Q}, \mathcal{Q}') \leq \varepsilon \cdot \frac{1+\varepsilon-\delta}{\varepsilon+\delta} = \frac{1+\varepsilon-2^{-\ell}}{\varepsilon+2^{-\ell}} \cdot \varepsilon = \frac{1+\varepsilon-2^{-\ell}}{1+2^\ell\varepsilon} \cdot 2^\ell\varepsilon$. Clearly the fraction is smaller than 1, and so the entire expression is at most $2^\ell\varepsilon$. ◀

We now prove the lemma.

Proof of Lemma 7.8. Let $\mathcal{U}_G$ be the uniform distribution over all functions in G. For $\varepsilon' := \frac{\varepsilon}{\varepsilon+\delta}$, define the quasi-distribution $\mathcal{Q}' := (1-\varepsilon')\mathcal{Q} + \varepsilon'\mathcal{U}_G$. Namely, if the vector of quasi-probabilities of $\mathcal{Q}$ is $(q_f)_{f \in G}$, then the the vector of quasi-probabilities of $\mathcal{Q}'$ is $(q'_f)_{f \in G}$ where $q'_f := (1-\varepsilon') \cdot q_f + \varepsilon'/|G|$.

First, we show that $\mathcal{Q}'$ is an ℓ-local quasi-distribution. That is, for all subsets $S \subseteq D$ of size at most ℓ and for every $\vec{b} \in \{0,1\}^S$ it holds that $\widetilde{\Pr}[\mathcal{Q}'(S) = \vec{b}] \geq 0$. Fix such an S and $\vec{b}$. If $\Pr_{f \in G}[f(S) = \vec{b}] = 0$, then there is no $f \in G$ such that $f(S) = \vec{b}$, and hence $\widetilde{\Pr}[\mathcal{Q}'(S) = \vec{b}] = 0$. Otherwise, $\Pr_{f \in G}[f(S) = \vec{b}] \geq \delta$, and hence,

$$\begin{aligned}
\widetilde{\Pr}[\mathcal{Q}'(S) = \vec{b}] &= \sum_{f \in G : f(S) = \vec{b}} q'_f \\
&= \left(\sum_{f \in G : f(S) = \vec{b}} (1-\varepsilon') q_f \right) + \varepsilon' \Pr_{f \in G}[f(S) = \vec{b}] \\
&\geq \left(\sum_{f \in G : f(S) = \vec{b}} (1-\varepsilon') q_f \right) + \varepsilon' \cdot \delta \\
&\geq -\varepsilon(1-\varepsilon') + \varepsilon' \cdot \delta \\
&= -\varepsilon \left(\frac{\delta}{\varepsilon+\delta} \right) + \frac{\varepsilon}{\varepsilon+\delta}\delta = 0 \ .
\end{aligned}$$

Second, we show that $\mathcal{Q}$ and $\mathcal{Q}'$ are close in the sense that $\Delta_\ell(\mathcal{Q},\mathcal{Q}') \le 2\varepsilon' \cdot (1+\varepsilon-\delta)$ (see Definition 7.5). Fix a subset $S \subseteq D$ of size at most ℓ, and let $E \subseteq \{0,1\}^S$. Then

$$\begin{aligned}\left|\widetilde{\Pr}[\mathcal{Q}(S)\in E] - \widetilde{\Pr}[\mathcal{Q}'(S)\in E]\right| &= \left|\left(\sum_{f\in G: f(S)\in E} \varepsilon' q_f\right) - \varepsilon' \Pr_{f\in G}[f(S)\in E]\right| \\ &= \left|\varepsilon'\widetilde{\Pr}[\mathcal{Q}(S)\in E] - \varepsilon' \Pr_{f\in G}[f(S)\in E]\right| \\ &\le \varepsilon'(1+\varepsilon-\delta)\ ,\end{aligned}$$

as required. ◀

8 Equivalence of non-signaling functions and local quasi-distributions

We establish an equivalence between non-signaling functions and local quasi-distributions. First, we show that every local quasi-distribution induces a non-signaling function. Second, we show that the converse is also true, namely, that every non-signaling function can be described by a local quasi-distribution. In fact, the set of quasi-distributions describing it is a real affine subspace.

▶ **Theorem 8.1** (from local quasi-distributions to non-signaling functions). *Let D be a finite domain. For every ℓ-local quasi-distribution $\mathcal{Q}$ over functions $f\colon D\to\{0,1\}$ there exists an ℓ-non-signaling function $\mathcal{F}$ over D such that for every subset $S\subseteq D$ of size $|S|\le\ell$ and string $\vec{b}\in\{0,1\}^S$, $\Pr[\mathcal{F}(S)=\vec{b}] = \widetilde{\Pr}[\mathcal{Q}(S)=\vec{b}]$.*

Proof. For every subset $S\subseteq D$ of size $|S|\le\ell$, define $\mathcal{F}_S$ to be the distribution over functions $f\colon S\to\{0,1\}$ where $\Pr[\mathcal{F}_S \text{ outputs } f] := \widetilde{\Pr}[\mathcal{Q}(S)=f(S)]$. Note that $\mathcal{F}_S$ is indeed a distribution because $\mathcal{Q}$ is ℓ-local, so the relevant probabilities are in $[0,1]$ and sum to 1. The definition immediately implies that $\Pr[\mathcal{F}(S)=\vec{b}] = \widetilde{\Pr}[\mathcal{Q}(S)=\vec{b}]$ for every string $\vec{b}\in\{0,1\}^S$. We are left to argue that $\mathcal{F}=\{\mathcal{F}_S\}_{S\subseteq D,|S|\le\ell}$ is ℓ-non-signaling.

Consider any two distinct subsets $S,T\subseteq D$ of size at most ℓ, and any string $\vec{b}\in\{0,1\}^{S\cap T}$. Let U_S denote the set of functions from $S\to\{0,1\}$. We have that

$$\begin{aligned}\Pr_{f\leftarrow\mathcal{F}_S}[f(S\cap T)=\vec{b}] &= \sum_{\substack{f\in U_S \text{ s.t.}\\ f(S\cap T)=\vec{b}}} \Pr[\mathcal{F}_S \text{ outputs } f] = \sum_{\substack{f\in U_S \text{ s.t.}\\ f(S\cap T)=\vec{b}}} \widetilde{\Pr}[\mathcal{Q}(S)=f(S)] \\ &= \sum_{\substack{f\in U_S \text{ s.t.}\\ f(S\cap T)=\vec{b}}} \sum_{\substack{g\in U \text{ s.t.}\\ g(S)=f(S)}} q_g = \sum_{\substack{g\in U \text{ s.t.}\\ g(S\cap T)=\vec{b}}} q_g = \widetilde{\Pr}[\mathcal{Q}(S\cap T)=\vec{b}]\end{aligned}$$

Similarly, we have that $\Pr_{f\leftarrow\mathcal{F}_T}[f(S\cap T)=\vec{b}] = \widetilde{\Pr}[\mathcal{Q}(S\cap T)=\vec{b}]$, and we conclude that $\Pr_{f\leftarrow\mathcal{F}_S}[f(S\cap T)=\vec{b}] = \Pr_{f\leftarrow\mathcal{F}_T}[f(S\cap T)=\vec{b}]$. Since S,T were arbitrary, $\mathcal{F}$ is ℓ-non-signaling. ◀

We now show that every k-non-signaling function $\mathcal{F}$ arises from a k-local quasi-distribution $\mathcal{Q}$. Moreover, the set of such quasi-distributions is an affine subspace of co-dimension $\binom{N}{\le k}$ in $\mathbb{R}^{2^N}$, where $N=|D|$ and $\binom{N}{\le k} := \sum_{i=0}^{k}\binom{N}{i}$. This converse is the interesting direction of the equivalence.

▶ **Theorem 8.2** (from non-signaling functions to local quasi-distributions). *For every k-non-signaling function $\mathcal{F}=\{\mathcal{F}_S\}_{S\subseteq D,|S|\le k}$ over a finite domain D of size N there exists a k-local*

quasi-distribution $\mathcal{Q}$ over functions $f\colon D \to \{0,1\}$ that describes $\mathcal{F}$ (for every subset $S \subseteq D$ of size $|S| \leq k$ and string $\vec{b} \in \{0,1\}^S$ it holds that $\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] = \Pr[\mathcal{F}(S) = \vec{b}]$).

Moreover, the set of such quasi-distributions (viewed as vectors in $\mathbb{R}^{2^N}$) is the affine subspace of co-dimension $\binom{N}{\leq k}$ given by $\mathcal{Q}_0 + span\{\chi_T : T \subseteq D, |T| > k\}$, where $\mathcal{Q}_0$ is any solution and $\chi_T\colon \{0,1\}^D \to \mathbb{R}$ is defined as $\chi_T(f) := (-1)^{\sum_{x\in T} f(x)}$.

Proof. We break the proof into three parts. First, we find one quasi-distribution that matches $\mathcal{F}$. Then, we find an affine space of such quasi-distributions. Finally, we prove that this affine space contains all possible solutions.

Finding one solution. We construct a k-local quasi-distribution $\mathcal{Q}$ that behaves like $\mathcal{F}$ on all sets of size at most k. Consider $q(\cdot) := \sum_{T:|T|\leq k} \widehat{q}(T)\chi_T(\cdot)$, where $\widehat{q}(T)$ is defined as follows.

$$\widehat{q}(T) := \begin{cases} \frac{1}{2^N} & \text{if } T = \emptyset \\ \frac{2}{2^N}\left(\Pr[\sum_{x\in T}\mathcal{F}(x) = 0] - \frac{1}{2}\right) & \text{if } 1 \leq |T| \leq k \\ 0 & \text{if } |T| > k \end{cases} .$$

Note that $\mathcal{Q}$ is a quasi-distribution because $\sum_f q_f = 2^N\langle q, \chi_\emptyset\rangle = 2^N\widehat{q}(\emptyset) = 1$. Now, for any subset $S = \langle x_1, \ldots, x_s\rangle$ with $|S| \leq k$,

$$\begin{aligned} \sum_{f:\sum_{x\in S} f(x)=0} q_f &= \sum_f q_f(-1)^{\sum_{x\in S} f(x)} + \sum_{f:\sum_{x\in S} f(x)=1} q_f \\ &= 2^N\langle q, \chi_S\rangle + \left(1 - \sum_{f:\sum_{x\in S} f(x)=0} q_f\right) \\ &= 2^N\frac{1}{2^{N-1}}\left(\Pr\left[\sum_{x\in S}\mathcal{F}(x) = 0\right] - \frac{1}{2}\right) + \left(1 - \sum_{f:\sum_{x\in S} f(x)=0} q_f\right) \\ &= 2\Pr\left[\sum_{x\in S}\mathcal{F}(x) = 0\right] - \sum_{f:\sum_{x\in S} f(x)=0} q_f \ , \end{aligned}$$

which implies that

$$\widetilde{\Pr}\left[\sum_{x\in S}\mathcal{Q}(x) = 0\right] = \sum_{f:\sum_{x\in S} f(x)=0} q_f = \Pr\left[\sum_{x\in S}\mathcal{F}(x) = 0\right] .$$

Therefore,

$$\widetilde{\Pr}\left[\sum_{x\in S}\mathcal{Q}(x) = 1\right] = 1 - \widetilde{\Pr}\left[\sum_{x\in S}\mathcal{Q}(x) = 0\right] = 1 - \Pr\left[\sum_{x\in S}\mathcal{F}(x) = 0\right] = \Pr\left[\sum_{x\in S}\mathcal{F}(x) = 1\right]$$

Thus, by Corollary 5.2 for any choice of bits $b_1, \ldots, b_s \in \{0,1\}$ we have

$$\begin{aligned}\Pr[\mathcal{F}(x_1) = b_1, \ldots, \mathcal{F}(x_s) = b_s] &= -1 + \frac{1}{2^{s-1}} \sum_{T \subseteq [s]} \Pr\left[\sum_{i \in T} \mathcal{F}(x_i) = \sum_{i \in T} b_i\right] \\ &= -1 + \frac{1}{2^{s-1}} \sum_{T \subseteq [s]} \widetilde{\Pr}\left[\sum_{i \in T} \mathcal{Q}(x_i) = \sum_{i \in T} b_i\right] \\ &= \widetilde{\Pr}[\mathcal{Q}(x_1) = b_1, \ldots, \mathcal{Q}(x_s) = b_s] \ .\end{aligned}$$

This shows that $\mathcal{Q}$ behaves like $\mathcal{F}$ on all sets of size at most k.

Finding more solutions. We argue that Fourier coefficients for subsets T of size greater than k do not affect the induced non-signaling function. Indeed, fix a subset $T \subseteq D$ of size greater than k, and let $\mathcal{Q}' = (q'_f)_f$ be the quasi-distribution obtained from $\mathcal{Q} = (q_f)_f$ by defining its weights as $q'_f := q_f + c\chi_T(f)$. Observe that for every ordered subset $S = \langle x_1, \ldots, x_s \rangle$ with $s \leq k$ and bits $b_1, \ldots, b_s$ it holds that

$$\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] = \sum_{f : f(S) = \vec{b}} q_f = \sum_{f : f(S) = \vec{b}} (q_f + c\chi_T(f)) = \widetilde{\Pr}[\mathcal{Q}'(S) = \vec{b}] \ .$$

To see that the middle equality holds, observe that there exists $y \in T \setminus S$, and thus

$$\begin{aligned}&\sum_{f : f(S) = \vec{b}} \chi_T(f) = \sum_{f(S) = \vec{b}} (-1)^{\sum_{x \in T} f(x)} \\ &= \sum_{\substack{f : f(S) = \vec{b} \\ f(y) = 0}} (-1)^{\sum_{x \in T \setminus \{y\}} f(x)} - \sum_{\substack{f : f(S) = \vec{b} \\ f(y) = 1}} (-1)^{\sum_{x \in T \setminus \{y\}} f(x)} = 0 \ .\end{aligned}$$

Therefore, $\mathcal{Q}'$ matches $\mathcal{Q}$ (and thus also $\mathcal{F}$) on all sets of size at most k. Since this holds for every T with $|T| > k$, we see that *every* q' in $q + \operatorname{span}\{\chi_T : T \subseteq D, |T| > k\}$ also matches $\mathcal{F}$ on all subsets of size at most k.

We found all solutions. Observe that if $\mathcal{Q}$ is a quasi-distribution, then for every subset $T \subseteq D$ with $1 \leq |T| \leq k$ it holds that

$$\begin{aligned}\widehat{q}(T) &= \frac{1}{2^N} \sum_f q_f (-1)^{\sum_{x \in T} f(x)} \\ &= \frac{1}{2^N} \left(\sum_{f : \sum_{x \in T} f(x) = 0} q_f - \sum_{f : \sum_{x \in T} f(x) = 1} q_f \right) \\ &= \frac{1}{2^N} \left(\widetilde{\Pr}\left[\sum_{x \in T} \mathcal{Q}(x) = 0\right] - \widetilde{\Pr}\left[\sum_{x \in T} \mathcal{Q}(x) = 1\right] \right) \\ &= \frac{1}{2^{N-1}} \left(\widetilde{\Pr}\left[\sum_{x \in T} \mathcal{Q}(x) = 0\right] - \frac{1}{2} \right) \ .\end{aligned}$$

If $\mathcal{Q}$ and $\mathcal{F}$ match on all input sets of size at most k, then they match on all parity events of size at most k, and so $\widehat{q}(T) = \frac{1}{2^{N-1}} \left(\widetilde{\Pr}[\sum_{x \in T} \mathcal{F}(x) = 0] - \frac{1}{2} \right)$. Since $\widehat{q}(\emptyset) = \frac{1}{2^N} \sum_f q_f = \frac{1}{2^N}$,

we see that exactly $\binom{N}{\leq k}$ Fourier coefficients are determined. Thus, the set of all solutions is contained in $q + \text{span}\{\chi_T : T \subseteq D, |T| > k\}$.

On the other hand, we have already shown that the affine space $q+\text{span}\{\chi_T : T \subseteq D, |T| > k\}$ contains only quasi-distributions that match $\mathcal{F}$ on all sets of size at most k. Thus, the affine space of *all* quasi-distributions that match $\mathcal{F}$ is precisely $q + \text{span}\{\chi_T : T \subseteq D, |T| > k\}$. ◀

9 Quasi-distributions over functions with small support

We show that *every* k-non-signaling function can be expressed as a quasi-distribution over functions with small support, namely, functions that evaluate to 1 for at most k inputs. For linearity testing, this implies that restricting a quasi-distribution to functions that are ε-close to linear is an empty condition, because all k-non-signaling functions can be expressed by such quasi-distributions for $\varepsilon = \frac{k}{2^n}$, regardless of whether they pass the linearity test with high or low probability.

For a finite domain D, we denote by U_D the set of all boolean functions $f\colon D \to \{0,1\}$ and, for $k \leq |D|$, denote by $U_{\leq k}$ the subset of U_D of all functions that evaluate to 1 for at most k values in D. We show that every k-non-signaling function $\mathcal{F}$ is described by a quasi-distribution over $U_{\leq k}$.

▶ **Theorem 9.1.** *Let D be a finite domain. For every k-non-signaling function $\mathcal{F}$ over D there exists a k-local quasi-distribution $\mathcal{Q}$ over D supported on $U_{\leq k}$ such that for every subset $S \subseteq D$ of size $|S| \leq k$ and string $\vec{b} \in \{0,1\}^S$ it holds that $\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] = \Pr[\mathcal{F}(S) = \vec{b}]$.*

The proof of Theorem 9.1 relies on the following claim.

▶ **Claim 9.2.** *Let $\mathcal{F}$ be a k-non-signaling function over a domain D, and let $\mathcal{Q}$ be a quasi-distribution over functions $f\colon D \to \{0,1\}$. If for every subset $S \subseteq D$ with $1 \leq |S| \leq k$ it holds that $\widetilde{\Pr}[\mathcal{Q}(S) = 1^{|S|}] = \Pr[\mathcal{F}(S) = 1^{|S|}]$ then for every subset $S \subseteq D$ with $|S| \leq k$ and string $\vec{b} \in \{0,1\}^S$ it holds that $\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] = \Pr[\mathcal{F}(S) = \vec{b}]$.*

We first prove Theorem 9.1 using the claim, and then prove the claim.

Proof of Theorem 9.1. By Claim 9.2 it suffices to prove that the following linear system of equations, in the variables $\{q_f\}_{f \in U_{\leq k}}$, has a solution:

$$\left\{ \begin{array}{rcl} \displaystyle\sum_{f \in U_{\leq k}} q_f & = & 1 \\ \displaystyle\sum_{\substack{f \in U_{\leq k} \text{ s.t.} \\ f(S) = 1^{|S|}}} q_f & = & \Pr[\mathcal{F}(S) = 1^{|S|}] \quad \forall\, S \subseteq D \text{ with } 1 \leq |S| \leq k \end{array} \right\} .$$

We do so by iteratively assigning values to the variables $\{q_f\}_{f \in U_{\leq k}}$, by considering all functions with support size k, then with support size $k-1$, and so on. At a high level, we shall use the fact that this system of linear equations corresponds to an upper triangular matrix (once variables are ordered according to support sizes), and thus can be solved via back substitution.

First, consider any $f \in U_{\leq k}$ such that $|\text{supp}(f)| = k$, and let $S := \text{supp}(f)$. Since f is the *only* function in $U_{\leq k}$ whose support equals S, we must assign

$$q_f := \Pr[\mathcal{F}(\text{supp}(f)) = 1^k] \ .$$

Next, we use induction on $s = k-1, \ldots, 1$ in decreasing order. Consider any $f \in U_{\leq k}$ such that $|\mathrm{supp}(f)| = s$, and set

$$q_f := \Pr[\mathcal{F}(\mathrm{supp}(f)) = 1^s] - \left(\sum_{\substack{f' \in U_{\leq k} \text{ s.t.} \\ \mathrm{supp}(f') \supsetneq \mathrm{supp}(f)}} q_{f'} \right) .$$

The above is well-defined since we first define q_f for all functions with larger support. Moreover, any choice of $q_{f''}$ for functions f'' whose support does not contain $\mathrm{supp}(f)$ does *not* affect the quasi-probability $\widetilde{\Pr}[\mathcal{Q}(\mathrm{supp}(f)) = 1^s]$, and so we may think of this assignment as q_f satisfying the constraint $\widetilde{\Pr}[\mathcal{Q}(\mathrm{supp}(f)) = 1^s] = \Pr[\mathcal{F}(\mathrm{supp}(f)) = 1^s]$.

Finally, if f is the all-zero function we define

$$q_f := 1 - \sum_{f' \neq f} q_{f'} \ ,$$

so that $\sum_{f \in U_{\leq k}} q_f = 1$. It is clear from the construction that the assignments to the variables $\{q_f\}_{f \in U_{\leq k}}$ above satisfy the necessary linear constraints, as desired. ◀

Proof of Claim 9.2. Fix any subset $S \subseteq D$ with $|S| \leq k$ and string $\vec{b} \in \{0,1\}^S$. We prove that $\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] = \Pr[\mathcal{F}(S) = \vec{b}]$, via induction on $|Z|$ where $Z := \{i \in S : b_i = 0\}$.

If $|Z| = 0$, then $\widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] = \Pr[\mathcal{F}(S) = \vec{b}]$ holds by the assumption of the claim.

Now suppose that $|Z| > 0$, and let $i^* \in S$ be any coordinate such that $b_{i^*} = 0$. Let $\vec{b}_{\neg i^*} \in \{0,1\}^S$ be the vector obtained from $\vec{b}$ by *flipping* the i^*-th coordinate to 1, and let $\vec{b}_{-i^*} \in \{0,1\}^{S \setminus \{i^*\}}$ be the vector obtained from $\vec{b}$ by *removing* the i^*-th coordinate. We deduce that

$$\begin{aligned} \Pr[\mathcal{F}(S) = \vec{b}] &= \Pr[\mathcal{F}(S \setminus \{i^*\}) = \vec{b}_{-i^*}] - \Pr[\mathcal{F}(S) = \vec{b}_{\neg i^*}] \ , \text{ and} \\ \widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}] &= \widetilde{\Pr}[\mathcal{Q}(S \setminus \{i^*\}) = \vec{b}_{-i^*}] - \widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}_{\neg i^*}] \ . \end{aligned}$$

The inductive hypothesis tells us that $\Pr[\mathcal{F}(S \setminus \{i^*\}) = \vec{b}_{-i^*}] = \widetilde{\Pr}[\mathcal{Q}(S \setminus \{i^*\}) = \vec{b}_{-i^*}]$ and $\Pr[\mathcal{F}(S) = \vec{b}_{\neg i^*}] = \widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}_{\neg i^*}]$, from which we obtain that $\Pr[\mathcal{F}(S) = \vec{b}] = \widetilde{\Pr}[\mathcal{Q}(S) = \vec{b}]$, as claimed. ◀

10 Exact local characterization of linear functions

We prove our results about non-signaling functions that always pass the linearity test. The theorem below states that the test passes with probability 1 if and only if the non-signaling function on sets of size at most $k-1$ can be described by a $(k-1)$-local quasi-distribution over linear functions.

▶ **Theorem 10.1** (exact local characterization). *Let $\mathcal{F}$ be a k-non-signaling function with $k \geq 4$. The following statements are equivalent.*

1. *The linearity test always accepts:* $\Pr_{x,y,\mathcal{F}}[\mathcal{F}(x) + \mathcal{F}(y) = \mathcal{F}(x+y)] = 1$.
2. *For all $x, y \in \{0,1\}^n$ it holds that* $\Pr_{\mathcal{F}}[\mathcal{F}(x) + \mathcal{F}(y) = \mathcal{F}(x+y)] = 1$.
3. *There exists a unique $(k-1)$-local quasi-distribution $\mathcal{L}$ over* LIN *such that for every set $S \subseteq \{0,1\}^n$ of size $|S| \leq k-1$ and vector $\vec{b} \in \{0,1\}^S$ it holds that* $\Pr[\mathcal{F}(S) = \vec{b}] = \widetilde{\Pr}[\mathcal{L}(S) = \vec{b}]$.

We comment on several aspects of the theorem.

- **The case of large k.** If $k \geq n+1$, then $\mathcal{L}$ in Item 3 is in fact a (standard) distribution over linear functions. *Explanation.* Let ℓ_α be the weight assigned to the linear function $\langle \alpha, \cdot \rangle$ by $\mathcal{L}$. Since $\mathcal{L}$ matches $\mathcal{F}$ on sets of size n, we see that each ℓ_α is non-negative:

$$\ell_\alpha = \sum_{\alpha' : \langle \alpha', e_i \rangle = \alpha_i \ \ 1 \leq i \leq n} \ell_{\alpha'} = \Pr[\mathcal{F}(e_1) = \alpha_1, \ldots, \mathcal{F}(e_n) = \alpha_n] \geq 0 \ .$$

- **Agreement on $k-1$ layers.** The fact that $|S| < k$ in Item 3 is necessary, because we can construct a k-non-signaling function $\mathcal{F}$ where $\Pr[\mathcal{F}(S) = \vec{b}] \neq \widetilde{\Pr}[\mathcal{L}(S) = \vec{b}]$ when $|S| = k$.

  Explanation. Let S_1 be the set of S such that $|S| < k$ or S is linearly dependent, and S_2 be the set of S such that $|S| = k$ and S is linearly independent. The non-signaling function $\mathcal{F}$ that answers according to a uniformly random linear function on all sets in S_1 and answers with uniformly random bits that sum to 0 on all sets in S_2 is k-non-signaling. Furthermore, the corresponding unique $\mathcal{L}$ is the uniform distribution over linear functions, and so $\Pr[\mathcal{F}(S) = \vec{b}] \neq \widetilde{\Pr}[\mathcal{L}(S) = \vec{b}]$ when $S \in S_2$.

- **The case of $k = 3$.** In the theorem it is necessary to have $k \geq 4$. This is because for $k = 3$ it is not true that Item 3 always implies Item 2: it is possible for Item 3 to hold while the linearity test passes with probability 0.

  Explanation. Let $\mathcal{L}$ be a uniform distribution over linear functions, and let $\mathcal{F}$ be a 3-non-signaling function that agrees with $\mathcal{L}$ on all query sets of size 2. For every subset $\{x, y, z\} \subseteq \{0,1\}^n \setminus \{0^n\}$ of size 3, the distribution of $\mathcal{F}$ is uniform over the set of tuples $\{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$. If the input set S contains 0^n, $\mathcal{F}_S$ assigns 0^n to 0 and answers the rest according to $\mathcal{F}_{S \setminus \{0^n\}}$. One can verify that $\mathcal{F}$ is indeed a 3-non-signaling function. Clearly, $\mathcal{F}$ satisfies Item 3, but passes the linearity test with probability 0, and hence does not satisfy Item 2.

Proof that 1 $\iff$ 2. The acceptance probability of the test can be re-written as

$$\Pr_{x,y \leftarrow \{0,1\}^n, \mathcal{F}}[\mathcal{F}(x) + \mathcal{F}(y) = \mathcal{F}(x+y)] = \frac{1}{2^{2n}} \sum_{x,y \in \{0,1\}^n} \Pr_{\mathcal{F}}[\mathcal{F}(x) + \mathcal{F}(y) = \mathcal{F}(x+y)] \ ,$$

and note that each of the probabilities in the sum lies in $[0,1]$. Therefore, the acceptance probability is 1 if and only if for *all* $x, y \in \{0,1\}^n$ it holds that $\Pr_{\mathcal{F}}[\mathcal{F}(x) + \mathcal{F}(y) = \mathcal{F}(x+y)] = 1$. ◀

Proof that 2 $\implies$ 3. We first argue that if $\mathcal{F}$ behaves linearly on sets of the form $\{x, y, x+y\}$, then it behaves linearly on all sets of size less than k. Let $s \in \{2, \ldots, k-1\}$, $x_1, \ldots, x_s \in \{0,1\}^n$, and $b \in \{0,1\}$, and define $S_i := \{\sum_{j=1}^{i} x_j, x_{i+1}, \ldots, x_s\}$ for every $i \in \{1, \ldots, s\}$. Note that $|S_i \cup S_{i+1}| = s - i + 2 \leq s + 1 \leq k$. Letting $\mathsf{add}(\cdot)$ be the addition function, the fact that the linearity test always passes implies that

$$\Pr\left[\mathsf{add}(\mathcal{F}(S_i)) = \mathsf{add}(\mathcal{F}(S_{i+1}))\right] = \Pr\left[\mathcal{F}\left(\sum_{j=1}^{i} x_j\right) + \mathcal{F}(x_{i+1}) = \mathcal{F}\left(\sum_{j=1}^{i+1} x_j\right)\right] = 1 \ .$$

This implies that $\Pr[\mathcal{F}(\sum_{i=1}^{s} x_i) = b] = \Pr[\sum_{i=1}^{s} \mathcal{F}(x_i) = b]$, via the following argument:

$$\begin{aligned}
&\left|\Pr\left[\sum_{i=1}^{s}\mathcal{F}(x_i)=b\right]-\Pr\left[\mathcal{F}\left(\sum_{i=1}^{s}x_i\right)=b\right]\right| \\
&=\left|\Pr[\mathsf{add}(\mathcal{F}(S_1))=b]-\Pr[\mathsf{add}(\mathcal{F}(S_s))=b]\right| \\
&=\left|\sum_{i=1}^{s-1}\Pr[\mathsf{add}(\mathcal{F}(S_i))=b]-\Pr[\mathsf{add}(\mathcal{F}(S_{i+1}))=b]\right| \\
&\leq\sum_{i=1}^{s-1}\left|\Pr[\mathsf{add}(\mathcal{F}(S_i))=b]-\Pr[\mathsf{add}(\mathcal{F}(S_{i+1}))=b]\right|=0\ ,
\end{aligned}$$

where the last equality is by Lemma 6.3, since $|S_i \cup S_{i+1}| \leq k$ for every i. Note that s must be strictly less than k because $|S_1 \cup S_2| = s + 1$.

We now construct $\mathcal{L}$, and argue that it has the desired properties. Define $(\ell_\alpha)_{\alpha\in\{0,1\}^n}$ to be the solution to the system of equations in Lemma 5.3 where $c_\beta := \Pr[\mathcal{F}(\beta) = 0]$ for each $\beta \in \{0,1\}^n$, and let $\mathcal{L}$ be the quasi-distribution over LIN that assigns weight ℓ_α to the linear function $\langle\alpha,\cdot\rangle$. That is, $(\ell_\alpha)_{\alpha\in\{0,1\}^n}$ satisfy the linear equations

$$\sum_{\alpha:\langle\alpha,x\rangle=0}\ell_\alpha=\Pr[\mathcal{F}(x)=0]$$

for all $x \in \{0,1\}^n$. Note that $\mathcal{L}$ is indeed a quasi-distribution, because $\sum_\alpha \ell_\alpha = \Pr[\mathcal{F}(0^n) = 0] = \Pr_{x\leftarrow\{0,1\}^n}[\mathcal{F}(0^n) + \mathcal{F}(x) = \mathcal{F}(x)] = 1$ (as $\mathcal{F}$ always passes the linearity test). We remark that every quasi-distribution supported on LIN is uniquely determined by its induced distributions on sets of size 1: a quasi-distribution is supported on LIN if and only if its distributions on sets of size 1 determine all of its Fourier coefficients (see full version for details).

Moreover, by definition of $(\ell_\alpha)_{\alpha\in\{0,1\}^n}$, for every $x \in \{0,1\}^n$ it holds that

$$\Pr[\mathcal{F}(x)=0]=\sum_{\alpha:\langle\alpha,x\rangle=0}\ell_\alpha=\widetilde{\Pr}[\mathcal{L}(x)=0]\ ,$$

which implies that for every $x \in \{0,1\}^n$ and bit $b \in \{0,1\}$ it holds that $\Pr[\mathcal{F}(x) = b] = \widetilde{\Pr}[\mathcal{L}(x) = b]$. In other words, $\mathcal{F}$ and $\mathcal{L}$ match on sets of size 1. This allows us to derive the same conclusion for all sets of size less than k, as follows.

For every $s \in \{1,\ldots,k-1\}$, $x_1,\ldots,x_s \in \{0,1\}^n$, and $b_1,\ldots,b_s \in \{0,1\}$,

$$\begin{aligned}
&\Pr[\mathcal{F}(x_1)=b_1,\ldots,\mathcal{F}(x_s)=b_s] \\
&=-1+\frac{1}{2^{s-1}}\sum_{T\subseteq[s]}\Pr\left[\sum_{i\in T}\mathcal{F}(x_i)=\sum_{i\in T}b_i\right]\ \text{(by Corollary 5.2)} \\
&=-1+\frac{1}{2^{s-1}}\sum_{T\subseteq[s]}\Pr\left[\mathcal{F}\left(\sum_{i\in T}x_i\right)=\sum_{i\in T}b_i\right]\ \text{(by linearity)} \\
&=-1+\frac{1}{2^{s-1}}\sum_{T\subseteq[s]}\widetilde{\Pr}\left[\mathcal{L}\left(\sum_{i\in T}x_i\right)=\sum_{i\in T}b_i\right] \\
&=-1+\frac{1}{2^{s-1}}\sum_{T\subseteq[s]}\widetilde{\Pr}\left[\sum_{i\in T}\mathcal{L}(x_i)=\sum_{i\in T}b_i\right]\ \text{(since supp}(\mathcal{L})\subseteq\mathsf{LIN}\text{)} \\
&=\widetilde{\Pr}[\mathcal{L}(x_1)=b_1,\ldots,\mathcal{L}(x_s)=b_s]\ .\ \text{(by Lemma 7.3)}
\end{aligned}$$

Finally, since $\mathcal{L}$ agrees with $\mathcal{F}$ on all subsets of size less than k, the quasi-probabilities must be in $[0,1]$, which means that $\mathcal{L}$ is $(k-1)$-local. ◀

Proof that 3 $\implies$ 2. Suppose that there exists a $(k-1)$-local quasi-distribution $\mathcal{L}$ over LIN such that for every $s \in \{1,\ldots,k-1\}$, $x_1,\ldots,x_s \in \{0,1\}^n$, and $b_1,\ldots,b_s \in \{0,1\}$ it holds that $\Pr[\mathcal{F}(x_1)=b_1,\ldots,\mathcal{F}(x_s)=b_s] = \widetilde{\Pr}[\mathcal{L}(x_1)=b_1,\ldots,\mathcal{L}(x_s)=b_s]$. For every $\alpha \in \{0,1\}^n$ denote by ℓ_α the weight assigned by $\mathcal{L}$ to the linear function $\langle \alpha, \cdot \rangle$. For every $x,y \in \{0,1\}^n$ it holds that

$$\begin{aligned}
\Pr[\mathcal{F}(x)+\mathcal{F}(y)=\mathcal{F}(x+y)] &= \sum_{b_1,b_2} \Pr[\mathcal{F}(x)=b_1, \mathcal{F}(y)=b_2, \mathcal{F}(x+y)=b_1+b_2] \\
&= \sum_{b_1,b_2} \widetilde{\Pr}[\mathcal{L}(x)=b_1, \mathcal{L}(y)=b_2, \mathcal{L}(x+y)=b_1+b_2] \\
&= \sum_{b_1,b_2} \sum_{\substack{\alpha:\langle\alpha,x\rangle=b_1 \\ \langle\alpha,y\rangle=b_2 \\ \langle\alpha,x+y\rangle=b_1+b_2}} \ell_\alpha = \sum_{b_1,b_2} \sum_{\substack{\alpha:\langle\alpha,x\rangle=b_1 \\ \langle\alpha,y\rangle=b_2}} \ell_\alpha = \sum_\alpha \ell_\alpha = 1 \ ,
\end{aligned}$$

as desired. Note that the equality on the second line uses the assumption that $k \geq 4$. This is because we need $\mathcal{L}$ to match $\mathcal{F}$ on sets of size 3, and we only know that $\mathcal{L}$ matches $\mathcal{F}$ on all sets of size at most $k-1$. ◀

11 Robust local characterization of linear functions

We prove our results about non-signaling functions that pass the linearity test with high probability. Given a k-non-signaling function $\mathcal{F}$, define its self-correction $\hat{\mathcal{F}}$ as follows. On an input $x \in \{0,1\}^n$ we sample from $\hat{\mathcal{F}}_{\{x\}}$ by drawing a uniform $w \in \{0,1\}^n$, sampling a function f from $\mathcal{F}_{\{x+w,w\}}$, and outputting $f(x+w)+f(w)$. We generalize this correction to larger input sets in the natural way.

▶ **Definition 11.1.** Given a k-non-signaling function $\mathcal{F}$, define the ***self-correction of*** $\mathcal{F}$ as follows. Given a set $S = \{x_1,\ldots,x_s\} \subseteq \{0,1\}^n$, we sample from $\hat{\mathcal{F}}_{\{x_1,\ldots,x_s\}}$ by drawing uniform and independent $w_1,\ldots,w_s \in \{0,1\}^n$, sampling a function f from the distribution $\mathcal{F}_{\{x_1+w_1,\ldots,x_s+w_s,w_1,\ldots,w_s\}}$, and outputting the function $\hat{f}$ that maps each x_i to $f(x_i+w_i)+f(w_i)$. That is, for every subset $S=\{x_1,\ldots,x_s\} \subseteq \{0,1\}^n$ of size at most $\hat{k}$ and $\vec{b} \in \{0,1\}^S$,

$$\Pr[\hat{\mathcal{F}}(S)=\vec{b}] := \Pr_{\substack{w_1,\ldots,w_s \leftarrow \{0,1\}^n \\ \mathcal{F}}} \begin{bmatrix} \mathcal{F}(x_1+w_1)+\mathcal{F}(w_1)=b_1 \\ \vdots \\ \mathcal{F}(x_s+w_s)+\mathcal{F}(w_s)=b_s \end{bmatrix} .$$

$\hat{\mathcal{F}}$ is a $\hat{k}$-non-signaling function for $\hat{k} \leq \lfloor k/2 \rfloor$. This follows immediately from the fact that the w_i's are random and independent, and the fact that $\mathcal{F}$ is k-non-signaling.

The following theorem says that, if a k-non-signaling function $\mathcal{F}$ passes the linearity test with high probability, then $\hat{\mathcal{F}}$ is close to a quasi-distribution over linear functions.

▶ **Theorem 11.2** (robust local characterization)**.** *Let $\mathcal{F}$ be a k-non-signaling function with $k \geq 7$, and let $\hat{\mathcal{F}}$ be its ($\hat{k}$-non-signaling) self-correction. Each of the following statements implies the next one.*

1. *The linearity test accepts with probability $1-\varepsilon$:* $\Pr_{x,y,\mathcal{F}}[\mathcal{F}(x)+\mathcal{F}(y)=\mathcal{F}(x+y)] \geq 1-\varepsilon$.

2. *For all* $x, y \in \{0,1\}^n$ *it holds that* $\Pr_{\hat{\mathcal{F}}}[\hat{\mathcal{F}}(x) + \hat{\mathcal{F}}(y) = \hat{\mathcal{F}}(x+y)] \geq 1 - \hat{\varepsilon}$ *with* $\hat{\varepsilon} := 4\varepsilon$*; moreover, it also holds that* $\Pr_{\hat{\mathcal{F}}}[\hat{\mathcal{F}}(0^n) = 0] = 1$.
3. *There exists a quasi-distribution* $\mathcal{L}$ *over* LIN *such that for every* $\ell \in \{1, \ldots, \hat{k}-1\}$ *it holds that* $\mathcal{L}$ *is* $(\ell, 2^{\ell/2}(\ell-1)\hat{\varepsilon})$*-local and, for every subset* $S \subseteq \{0,1\}^n$ *of size at most* ℓ *and every event* $E \subseteq \{0,1\}^S$, $\left|\Pr[\hat{\mathcal{F}}(S) \in E] - \widetilde{\Pr}[\mathcal{L}(S) \in E]\right| \leq (|S|-1)\cdot\|\widehat{\mathbf{1}_E}\|_1 \cdot \hat{\varepsilon} \leq (|S|-1)\cdot\sqrt{|E|}\cdot\hat{\varepsilon}$.
4. *For every* $\ell \in \{1, \ldots, \hat{k}-1\}$*, there exists an* ℓ*-local quasi-distribution* $\mathcal{L}'$ *over* LIN *such that* $\Delta_\ell(\hat{\mathcal{F}}, \mathcal{L}') \leq (2^\ell + 1) \cdot 2^{\ell/2}(\ell-1)\hat{\varepsilon}$.

We highlight some of the differences of Theorem 11.2 ($\varepsilon \geq 0$) from Theorem 10.1 ($\varepsilon = 0$).

- In Item 2, we now need to use the self-correction $\hat{\mathcal{F}}$ to ensure that $\Pr_{\hat{\mathcal{F}}}[\hat{\mathcal{F}}(x+y) = \hat{\mathcal{F}}(x) + \hat{\mathcal{F}}(y)]$ is large *for every* $x, y \in \{0,1\}^n$, as opposed to random $x, y \in \{0,1\}^n$. This is necessary because otherwise it is possible for $\Pr_{\mathcal{F}}[\mathcal{F}(x) + \mathcal{F}(y) = \mathcal{F}(x+y)]$ to be small for certain choices of x and y, and in this case a quasi-distribution supported only on linear functions has no hope of approximating $\mathcal{F}$ on sets containing $\{x, y, x+y\}$.
- In Item 3, we choose $\mathcal{L}$ to match $\hat{\mathcal{F}}$ exactly on all sets of size 1, as before. However, since the linearity condition only holds approximately, this means that we only get approximate matching on larger input sets, and this approximation deteriorates as the sets get larger.
- Since $\mathcal{L}$ only matches $\hat{\mathcal{F}}$ approximately, it is only an approximately ℓ-local distribution. Thus, we require the additional step of Item 4, where we correct $\mathcal{L}$ to an exactly ℓ-local distribution.

We now proceed to the proof of Theorem 11.2.

Proof that 1 $\implies$ 2. Fix $x, y \in \{0,1\}^n$. The definition of $\hat{\mathcal{F}}$ implies that

$$\begin{aligned}
&\Pr_{\hat{\mathcal{F}}}[\hat{\mathcal{F}}(x) + \hat{\mathcal{F}}(y) = \hat{\mathcal{F}}(x+y)] \\
&= \Pr_{\substack{w_x \\ w_y \\ w_{x+y} \\ \mathcal{F}}} [\mathcal{F}(x+w_x) + \mathcal{F}(w_x) + \mathcal{F}(y+w_y) + \mathcal{F}(w_y) = \mathcal{F}(x+y+w_{x+y}) + \mathcal{F}(w_{x+y})] \ .
\end{aligned}$$

Define

$$\begin{aligned}
S_1 &:= \{x+w_x, y+w_y, x+y+w_{x+y}, w_x, w_y, w_{x+y}\} \ , \\
S_2 &:= \{x+w_x+w_y, y+w_y, x+y+w_{x+y}, w_x, w_{x+y}\} \ , \\
S_3 &:= \{x+w_x+w_y, y+w_y+w_{x+y}, x+y+w_{x+y}, w_x\} \ , \\
S_4 &:= \{x+w_x+w_y, y+w_y+w_{x+y}, x+y+w_{x+y}+w_x\} \ .
\end{aligned}$$

Observe that $|S_i \cup S_{i+1}| \leq 7 \leq k$. Letting $\mathsf{add}(\cdot)$ be the addition function, the linearity test passing with probability at least $1-\varepsilon$ implies that

$$\begin{aligned}
&\Pr[\mathsf{add}(\mathcal{F}(S_1)) = \mathsf{add}(\mathcal{F}(S_2))] \\
&= \Pr_{\substack{w_x, w_y \\ \mathcal{F}}}[\mathcal{F}(x+w_x+w_y) = \mathcal{F}(x+w_x) + \mathcal{F}(w_y)] \geq 1-\varepsilon \ , \\
&\Pr[\mathsf{add}(\mathcal{F}(S_2)) = \mathsf{add}(\mathcal{F}(S_3))] \\
&= \Pr_{\substack{w_y, w_{x+y} \\ \mathcal{F}}}[\mathcal{F}(y+w_y+w_{x+y}) = \mathcal{F}(y+w_y) + \mathcal{F}(w_{x+y})] \geq 1-\varepsilon \ , \\
&\Pr[\mathsf{add}(\mathcal{F}(S_3)) = \mathsf{add}(\mathcal{F}(S_4))] \\
&= \Pr_{\substack{w_x, w_{x+y} \\ \mathcal{F}}}[\mathcal{F}(x+y+w_{x+y}+w_x) = \mathcal{F}(x+y+w_{x+y}) + \mathcal{F}(w_x)] \geq 1-\varepsilon \ , \\
&\Pr[\mathsf{add}(\mathcal{F}(S_4)) = 0] \\
&= \Pr_{\substack{w_x, w_y \\ w_{x+y} \\ \mathcal{F}}}[\mathcal{F}(x+w_x+w_y) + \mathcal{F}(y+w_y+w_{x+y}) = \mathcal{F}(x+y+w_{x+y}+w_x)] \geq 1-\varepsilon \ .
\end{aligned}$$

Therefore, by Lemma 6.3,

$$\begin{aligned}&|\Pr[\mathsf{add}(\mathcal{F}(S_1)) = 0] - \Pr[\mathsf{add}(\mathcal{F}(S_4)) = 0]| \\ &\leq \sum_{i=1}^{3} |\Pr[\mathsf{add}(\mathcal{F}(S_i)) = 0] - \Pr[\mathsf{add}(\mathcal{F}(S_{i+1})) = 0]| \leq 3\varepsilon \ .\end{aligned}$$

Since $\Pr[\mathsf{add}(\mathcal{F}(S_4)) = 0] \geq 1 - \varepsilon$, it follows that $\Pr_{\hat{\mathcal{F}}}[\hat{\mathcal{F}}(x) + \hat{\mathcal{F}}(y) = \hat{\mathcal{F}}(x+y)] = \Pr[\mathsf{add}(\mathcal{F}(S_1)) = 0] \geq 1 - 4\varepsilon = 1 - \hat{\varepsilon}$, as claimed. Finally, $\Pr[\hat{\mathcal{F}}(0^n) = 0] = \Pr_{w \in \{0,1\}^n}[\mathcal{F}(w + 0^n) + \mathcal{F}(w) = 0] = 1$. ◀

Proof that 2 $\implies$ 3. This proof generalizes the proof that 2 $\implies$ 3 in Theorem 10.1. We begin by arguing that $\hat{\mathcal{F}}$ behaves almost linearly on sets of size at most $\hat{k} - 1$. Let $s \in \{2, \ldots, \hat{k} - 1\}$, $x_1, \ldots, x_s \in \{0,1\}^n$, and $b \in \{0,1\}$, and define $S_i := \{\sum_{j=1}^{i} x_j, x_{i+1}, \ldots, x_s\}$ for every $i \in \{1, \ldots, s\}$. Note that $|S_i \cup S_{i+1}| = s - i + 2 \leq s + 1 \leq \hat{k}$. Letting $\mathsf{add}(\cdot)$ be the addition function, the fact that the linearity test passes with probability at least $1 - \hat{\varepsilon}$ implies that

$$\Pr\left[\mathsf{add}(\hat{\mathcal{F}}(S_i)) = \mathsf{add}(\hat{\mathcal{F}}(S_{i+1}))\right] = \Pr\left[\hat{\mathcal{F}}\left(\sum_{j=1}^{i} x_j\right) + \hat{\mathcal{F}}(x_{i+1}) = \hat{\mathcal{F}}\left(\sum_{j=1}^{i+1} x_j\right)\right] \geq 1 - \hat{\varepsilon} \ .$$

This implies that $\left|\Pr[\hat{\mathcal{F}}(\sum_{i=1}^{s} x_i) = b] - \Pr[\sum_{i=1}^{s} \hat{\mathcal{F}}(x_i) = b]\right| \leq (s-1)\hat{\varepsilon}$, via the following argument:

$$\begin{aligned}&\left|\Pr\left[\sum_{i=1}^{s} \hat{\mathcal{F}}(x_i) = b\right] - \Pr\left[\hat{\mathcal{F}}\left(\sum_{i=1}^{s} x_i\right) = b\right]\right| \\ =&\left|\Pr[\mathsf{add}(\hat{\mathcal{F}}(S_1)) = b] - \Pr[\mathsf{add}(\hat{\mathcal{F}}(S_s)) = b]\right| \\ =&\left|\sum_{i=1}^{s-1} \Pr[\mathsf{add}(\hat{\mathcal{F}}(S_i)) = b] - \Pr[\mathsf{add}(\hat{\mathcal{F}}(S_{i+1})) = b]\right| \\ \leq&\sum_{i=1}^{s-1} \left|\Pr[\mathsf{add}(\hat{\mathcal{F}}(S_i)) = b] - \Pr[\mathsf{add}(\hat{\mathcal{F}}(S_{i+1})) = b]\right| \\ \leq&(s-1)\hat{\varepsilon} \ .\end{aligned}$$

where the last inequality is by Lemma 6.3, since $|S_i \cup S_{i+1}| \leq \hat{k}$ for every i. Note that s must be strictly less than $\hat{k}$ because $|S_1 \cup S_2| = s + 1$.

We construct $\mathcal{L}$ as before. Define $(\ell_\alpha)_{\alpha \in \{0,1\}^n}$ to be the solution to the system of equations in Lemma 5.3 where $c_\beta := \Pr[\hat{\mathcal{F}}(\beta) = 0]$ for each $\beta \in \{0,1\}^n$, and let $\mathcal{L}$ be the quasi-distribution over LIN that assigns weight ℓ_α to the linear function $\langle \alpha, \cdot \rangle$. Note that $\mathcal{L}$ is indeed a quasi-distribution, because $\sum_\alpha \ell_\alpha = \Pr[\hat{\mathcal{F}}(0^n) = 0] = 1$.

Moreover, by definition of $(\ell_\alpha)_{\alpha \in \{0,1\}^n}$, for every $x \in \{0,1\}^n$ and $b \in \{0,1\}$ it holds that $\Pr[\hat{\mathcal{F}}(x) = b] = \widetilde{\Pr}[\mathcal{L}(x) = b]$. In other words, $\mathcal{F}$ and $\mathcal{L}$ match *exactly* on sets of size one. We now prove that $\mathcal{F}$ and $\mathcal{L}$ match *approximately* for sets of larger size (but still less than $\hat{k}$) with a guarantee that degrades with the set size.

Fix $s \in \{1, \ldots, k-1\}$, $S = \{x_1, \ldots, x_s\} \subseteq \{0,1\}^n$, and $E \subseteq \{0,1\}^S$. We use Lemma 5.1

to get real numbers $\{c_T\}_{T\subseteq[s]}$ that depend only on E such that

$$\begin{aligned}
&\left|\Pr[\hat{\mathcal{F}}(S)\in E]-\widetilde{\Pr}[\mathcal{L}(S)\in E]\right| \\
&=\left|\sum_{T\subseteq[s]} c_T\cdot\Pr\left[\sum_{i\in T}\hat{\mathcal{F}}(x_i)=0\right]-\sum_{T\subseteq[s]} c_T\cdot\widetilde{\Pr}\left[\sum_{i\in T}\mathcal{L}(x_i)=0\right]\right| \\
&=\left|\sum_{T\subseteq[s]} c_T\left(\Pr\left[\sum_{i\in T}\hat{\mathcal{F}}(x_i)=0\right]-\widetilde{\Pr}\left[\sum_{i\in T}\mathcal{L}(x_i)=0\right]\right)\right| \\
&=\left|\sum_{T\subseteq[s]} c_T\left(\Pr\left[\sum_{i\in T}\hat{\mathcal{F}}(x_i)=0\right]-\widetilde{\Pr}\left[\mathcal{L}\left(\sum_{i\in T}x_i\right)=0\right]\right)\right| \\
&=\left|\sum_{T\subseteq[s]} c_T\left(\Pr\left[\sum_{i\in T}\hat{\mathcal{F}}(x_i)=0\right]-\Pr\left[\hat{\mathcal{F}}\left(\sum_{i\in T}x_i\right)=0\right]\right)\right| \\
&\le\sum_{T\subseteq[s]}|c_T|\,(|T|-1)\,\hat{\varepsilon}\le\hat{\varepsilon}\cdot(s-1)\cdot\sum_{T\subseteq[s]}|c_T| \\
&\le\hat{\varepsilon}\cdot(s-1)\|\widehat{\mathbf{1}_E}\|_1\le\hat{\varepsilon}\cdot(s-1)\sqrt{|E|}\ .
\end{aligned}$$

Since $\hat{\mathcal{F}}$ defines probabilities in $[0,1]$, $\mathcal{L}$ is (ℓ,ε')-local with $\varepsilon'=(\ell-1)2^{\ell/2}\hat{\varepsilon}$ for any $\ell<\hat{k}$. ◀

Proof that 3 $\implies$ 4. Fix $\ell\in\{1,\ldots,\hat{k}-1\}$, and let $\mathcal{L}$ be the $(\ell,2^{\ell/2}(\ell-1)\hat{\varepsilon})$-local quasi-distribution $\mathcal{L}$ over LIN such that for every subset $S\subseteq\{0,1\}^n$ of size at most ℓ and event $E\subseteq\{0,1\}^S$ it holds that

$$\left|\Pr[\hat{\mathcal{F}}(S)\in E]-\widetilde{\Pr}[\mathcal{L}(S)\in E]\right|\le\sqrt{|E|}(|S|-1)\hat{\varepsilon}\le 2^{\ell/2}(\ell-1)\hat{\varepsilon}\ .$$

Thus, $\Delta_\ell(\hat{\mathcal{F}},\mathcal{L})\le 2^{\ell/2}(\ell-1)\hat{\varepsilon}$. By Corollary 7.9, there is an ℓ-local quasi-distribution $\mathcal{L}'$ such that $\Delta_\ell(\mathcal{L},\mathcal{L}')\le 2^\ell\cdot 2^{\ell/2}(\ell-1)\hat{\varepsilon}$. Therefore,

$$\Delta_\ell(\hat{\mathcal{F}},\mathcal{L}')\le\Delta_\ell(\hat{\mathcal{F}},\mathcal{L})+\Delta_\ell(\mathcal{L},\mathcal{L}')\le 2^{\ell/2}(\ell-1)\hat{\varepsilon}+2^\ell\cdot 2^{\ell/2}(\ell-1)\hat{\varepsilon}=(2^\ell+1)\cdot 2^{\ell/2}(\ell-1)\hat{\varepsilon}\ .\blacktriangleleft$$

References

1 William Aiello, Sandeep N. Bhatt, Rafail Ostrovsky, and Sivaramakrishnan Rajagopalan. Fast verification of any remote procedure call: Short witness-indistinguishable one-round proofs for NP. In *Proceedings of the 27th International Colloquium on Automata, Languages and Programming*, ICALP '00, pages 463–474, 2000.

2 Sabri W. Al-Safi and Anthony J. Short. Simulating all nonsignaling correlations via classical or quantum theory with negative probabilities. *Physical Review Letters*, 111:170403, 2013.

3 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. Preliminary version in FOCS '92.

4 Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. Preliminary version in FOCS '92.

5 Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. *Combinatorica*, 23(3):365–426, 2003. Preliminary version appeared in STOC '97.

6 László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *Proceedings of the 23rd ACM Symposium on Theory of Computing*, STOC '91, pages 21–32, 1991.

7 László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3–40, 1991. Preliminary version appeared in FOCS '90.

8 Jonathan Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75:032304, 2007.

9 Jonathan Barrett, Lucien Hardy, and Adrian Kent. No signaling and quantum key distribution. *Physical Review Letters*, 95:010503, 2005.

10 Jonathan Barrett, Noah Linden, Serge Massar, Stefano Pironio, Sandu Popescu, and David Roberts. Nonlocal correlations as an information-theoretic resource. *Physical Review Letters*, 71:022101, 2005.

11 Jonathan Barrett and Stefano Pironio. Popescu–Rohrlich correlations as a unit of nonlocality. *Physical Review Letters*, 95:140401, 2005.

12 Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos A. Kiwi, and Madhu Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.

13 Michael Ben-Or, Don Coppersmith, Mike Luby, and Ronitt Rubinfeld. Non-abelian homomorphism testing, and distributions close to their self-convolutions. *Random Structures and Algorithms*, 32(1):49–70, 2008.

14 Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. In *Proceedings of the 51st IEEE Symposium on Foundations of Computer Science*, FOCS '10, pages 488–497, 2010.

15 Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.

16 Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Méthot, Alain Tapp, and Falk Unger. Limit on nonlocality in any world in which communication complexity is not trivial. *Physical Review Letters*, 96:250401, 2006.

17 Anne Broadbent and André Allan Méthot. On the power of non-local boxes. *Theoretical Computer Science*, 358(1):3–14, 2006.

18 Harry Buhrman, Matthias Christandl, Falk Unger, Stephanie Wehner, and Andreas Winter. Implications of superstrong non-locality for cryptography. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 462(2071):1919–1932, 2006.

19 Nicolas J. Cerf, Nicolas Gisin, Serge Massar, and Sandu Popescu. Simulating maximal quantum entanglement without communication. *Physical Review Letters*, 94:220403, 2005.

20 Rui Chao and Ben W. Reichardt. Test to separate quantum theory from non-signaling theories. arXiv quant-ph/1706.02008, 2017.

21 Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. *SIAM Journal on Computing*, 46:1336–1369, 2017.

22 Paul A. M. Dirac. The physical interpretation of quantum mechanics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 180(980):1–40, 1942.

23 Cynthia Dwork, Michael Langberg, Moni Naor, Kobbi Nissim, and Omer Reingold. Succinct NP proofs and spooky interactions, December 2004. Available at `www.openu.ac.il/home/mikel/papers/spooky.ps`.

24 Uriel Feige, Shafi Goldwasser, Laszlo Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996. Preliminary version in FOCS '91.

25 Richard P. Feynman. Negative probability. In Basil J. Hiley and D. Peat, editors, *Quantum Implications: Essays in Honour of David Bohm*, pages 235–248. Law Book Co of Australasia, 1987.

26 Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.

27 Thomas Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(1):141–172, 2009. Preliminary version appeared in STOC '07.

28 Tsuyoshi Ito. Polynomial-space approximation of no-signaling provers. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming*, ICALP '10, pages 140–151, 2010.

29 Tsuyoshi Ito, Hirotada Kobayashi, and Keiji Matsumoto. Oracularization and two-prover one-round interactive proofs against nonlocal strategies. In *Proceedings of the 24th IEEE Annual Conference on Computational Complexity*, CCC '09, pages 217–228, 2009.

30 Tsuyoshi Ito and Thomas Vidick. A multi-prover interactive proof for NEXP sound against entangled provers. In *Proceedings of the 53rd IEEE Symposium on Foundations of Computer Science*, FOCS '12, pages 243–252, 2012.

31 Nick S. Jones and Lluís Masanes. Interconversion of nonlocal correlations. *Physical Review A*, 72:052312, 2005.

32 Yael Kalai, Ran Raz, and Ron Rothblum. Delegation for bounded space. In *Proceedings of the 45th ACM Symposium on the Theory of Computing*, STOC '13, pages 565–574, 2013.

33 Yael Tauman Kalai, Ran Raz, and Oded Regev. On the space complexity of linear programming with preprocessing. In *Proceedings of the 7th Innovations in Theoretical Computer Science Conference*, ITCS '16, pages 293–300, 2016.

34 Yael Tauman Kalai, Ran Raz, and Ron D. Rothblum. How to delegate computations: the power of no-signaling proofs. In *Proceedings of the 46th ACM Symposium on Theory of Computing*, STOC '14, pages 485–494, 2014. Full version available at `https://eccc.weizmann.ac.il/report/2013/183/`.

35 Leonid A. Khalfin and Boris S. Tsirelson. Quantum and quasi-classical analogs of Bell inequalities. *Symposium on the Foundations of Modern Physics*, pages 441–460, 1985.

36 Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. Quantum nonlocality and beyond: Limits from nonlocal computation. *Physical Review Letters*, 99:180502, 2007.

37 Carsten Lund, Lance Fortnow, Howard J. Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. *Journal of the ACM*, 39(4):859–868, 1992.

38 Lluís Masanes, Antonio Acín, and Nicolas Gisin. General properties of nonsignaling theories. *Physical Review A*, 73:012112, 2006.

39 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

40 Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

41 Sandu Popescu and Daniel Rohrlich. *Causality and Nonlocality as Axioms for Quantum Mechanics*, pages 383–389. Springer Netherlands, 1998.

42 Prasad Raghavendra and David Steurer. Integrality gaps for strong SDP relaxations of UNIQUE GAMES. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, FOCS '09, pages 575–585, 2009. Full version at `http://people.eecs.berkeley.edu/~prasad/Files/cspgaps.pdf`.

43 Peter Rastall. Locality, Bell's theorem, and quantum mechanics. *Foundations of Physics*, 15(9):963–972, 1985.

44 Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th ACM Symposium on Theory of Computing*, STOC '97, pages 475–484, 1997.

45 Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

46 Adi Shamir. IP = PSPACE. *Journal of the ACM*, 39(4):869–877, 1992.

47 Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.

48 Anthony J. Short, Nicolas Gisin, and Sandu Popescu. The physics of no-bit-commitment: Generalized quantum non-locality versus oblivious transfer. *Quantum Information Processing*, 5(2):131–138, 2006.

49 Anthony J. Short, Sandu Popescu, and Nicolas Gisin. Entanglement swapping for generalized nonlocal correlations. *Physical Review A*, 73:012101, 2006.

50 Amir Shpilka and Avi Wigderson. Derandomizing homomorphism testing in general groups. In *Proceedings of the 36th ACM Symposium on the Theory of Computing*, STOC '04, pages 427–435, 2004.

51 Wim van Dam. Implausible consequences of superstrong nonlocality. *Natural Computing*, 12:9–12, 2013.

52 Thomas Vidick. Linearity testing with entangled provers, 2014. `http://users.cms.caltech.edu/~vidick/linearity_test.pdf`.

53 Stefan Wolf and Jürg Wullschleger. Oblivious transfer and quantum non-locality. In *Proceedings of the 2005 International Symposium on Information Theory*, ISIT '05, pages 1745–1748, 2005.