Lower Bounds on Same-Set Inner Product in Correlated Spaces*

Jan Hązła†[1], Thomas Holenstein[2], and Elchanan Mossel‡[3]

1 **ETH Zürich, Department of Computer Science, Zurich, Switzerland**
jan.hazla@inf.ethz.ch
2 **Google, Zurich, Switzerland**
thomas.holenstein@gmail.com
3 **MIT, Cambridge MA, USA**
elmos@mit.edu

Abstract

Let $\mathcal{P}$ be a probability distribution over a finite alphabet Ω^ℓ with all ℓ marginals equal. Let $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}, \underline{X}^{(j)} = (X_1^{(j)}, \ldots, X_n^{(j)})$ be random vectors such that for every coordinate $i \in [n]$ the tuples $(X_i^{(1)}, \ldots, X_i^{(\ell)})$ are i.i.d. according to $\mathcal{P}$.

The question we address is: does there exist a function $c_{\mathcal{P}}()$ independent of n such that for every $f : \Omega^n \to [0,1]$ with $\mathrm{E}[f(\underline{X}^{(1)})] = \mu > 0$:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c_{\mathcal{P}}(\mu) > 0 \ ?$$

We settle the question for $\ell = 2$ and when $\ell > 2$ and $\mathcal{P}$ has bounded correlation $\rho(\mathcal{P}) < 1$.

1998 ACM Subject Classification G.2.1 Combinatorics, G.3 Probabilistic algorithms

Keywords and phrases same set hitting, product spaces, correlation, lower bounds

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2016.34

1 Introduction

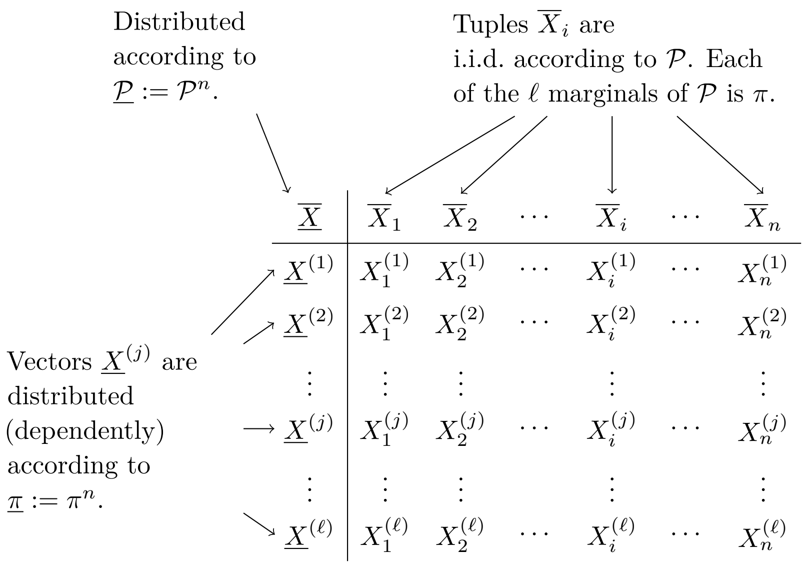
1.1 Basic example

To introduce the problem we are studying, consider the following example. Let $S \subseteq \{0,1,2\}^n$ be a non-empty set of density $\mu = \frac{|S|}{3^n}$. We pick a random vector $\underline{X} = (X_1, \ldots, X_n)$ uniformly from $\{0,1,2\}^n$, and then sample another vector $\underline{Y} = (Y_1, \ldots, Y_n)$ such that for each i independently, coordinate Y_i is picked uniformly in $\{X_i, X_i + 1 \bmod 3\}$. Our goal is to show that:

$$\Pr[\underline{X} \in S \wedge \underline{Y} \in S] \geq c(\mu) > 0 \ .$$

In other words, we want to bound away the probability from 0 by an expression which only depends on μ and not on n.

* Part of this work was done while T. H. and E. M. were at the Simons Institute.
† J. H was supported by the Swiss National Science Foundation (SNF), project no. 200021-132508.
‡ E. M. was supported by NSF grant DMS-1106999, NSF Grant CCF 1320105 and DOD ONR grant N000141110140 and grant 328025 from the Simons foundation.



$$\alpha(\mathcal{P}) := \min_{x\in\Omega} \mathcal{P}(x, x, \ldots, x)$$

$$\rho(\mathcal{P}) : \text{See Definition 10}$$

$$\begin{aligned} X_i^{(j)} &\in \Omega \\ \underline{X}^{(j)} &\in \underline{\Omega} := \Omega^n \\ \overline{X}_i &\in \overline{\Omega} := \Omega^\ell \\ \overline{\underline{X}} &\in \overline{\underline{\Omega}} := \Omega^{n\cdot\ell} \\ S &\subseteq \underline{\Omega} \end{aligned}$$

Figure 1 Naming of the random variables in the general case. The columns $\overline{X}_i$ are distributed *i.i.d* according to $\mathcal{P}$. Each $X_i^{(j)}$ is distributed according to π. The overall distribution of $\overline{\underline{X}}$ is $\underline{\mathcal{P}}$.

1.2 Our results

More generally, let Ω be a finite alphabet and assume we are given a probability distribution $\mathcal{P}$ over Ω^ℓ for some $\ell \geq 2$ – we will call it an *ℓ-step probability distribution over* Ω.

Furthermore, assume we are given $n \in \mathbb{N}$. We consider ℓ vectors $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}$, $\underline{X}^{(j)} = (X_1^{(j)}, \ldots, X_n^{(j)})$ such that for every $i \in [n]$, the ℓ-tuple $(X_i^{(1)}, \ldots, X_i^{(\ell)})$ is sampled according to $\mathcal{P}$, independently of the other coordinates $i' \neq i$ (see Figure 1 for an overview of the notation).

▶ **Definition 1.** Let $\mu, \delta \in (0, 1]$. We say that a distribution $\mathcal{P}$ *is* (μ, δ)*-same-set hitting*, if, whenever a function $f : \Omega^n \to [0, 1]$ satisfies $\mathrm{E}[f(\underline{X}^{(j)})] \geq \mu$ for every $j \in [\ell] := \{1, \ldots, \ell\}$, we have

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq \delta\ .$$

We call $\mathcal{P}$ *same-set hitting* if for every $\mu \in (0, 1]$ there exists $\delta \in (0, 1]$ such that $\mathcal{P}$ is (μ, δ)-same-set hitting.

In this paper we address the question: which distributions $\mathcal{P}$ are same-set hitting? We achieve full characterisation for $\ell = 2$ and answer the question affirmatively for a large class of distributions with $\ell > 2$.

To explain related work and our results, we introduce a stronger notion:

▶ **Definition 2.** Let $\mu, \delta \in (0, 1]$. We say that a distribution $\mathcal{P}$ *is* (μ, δ)*-set hitting*, if, whenever functions $f^{(1)}, \dots, f^{(\ell)} : \Omega^n \to [0, 1]$ satisfy $\mathrm{E}[f^{(j)}(\underline{X}^{(j)})] \ge \mu$ for every $j \in [\ell]$, we have

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \ge \delta \; . \tag{1}$$

We call $\mathcal{P}$ *set hitting* if for every $\mu \in (0, 1]$ there exists $\delta \in (0, 1]$ such that $\mathcal{P}$ is (μ, δ)-set hitting.

The full classification of set hitting distributions can be deduced from a paper on reverse hypercontractivity[1] by Mossel, Oleszkiewicz and Sen [10]:

▶ **Theorem 3** (follows from [10]). *A probability space $\mathcal{P}$ is set hitting if and only if:*

$$\min_{\substack{x^{(1)} \in \mathrm{supp}(X_i^{(1)}),\\ \dots,\\ x^{(\ell)} \in \mathrm{supp}(X_i^{(\ell)})}} \mathcal{P}(x^{(1)}, \dots, x^{(\ell)}) > 0 \; . \tag{2}$$

To state our results, we need to introduce the following properties of $\mathcal{P}$:

▶ **Definition 4.** We say that $\mathcal{P}$ *has equal marginals* if for every $j \in [\ell]$ and every $x \in \Omega$:

$$\Pr[X_i^{(1)} = x] = \ldots = \Pr[X_i^{(j)} = x] = \ldots = \Pr[X_i^{(\ell)} = x] \; .$$

As explained in Section 3.4.2, the same-set hitting is interesting only for distributions with equal marginals. Whenever we discuss such distributions, we assume w.l.o.g that Ω is equal to the support of the marginal.

▶ **Definition 5.** We define:

$$\alpha(\mathcal{P}) := \min_{x \in \Omega} \mathcal{P}(x, \dots, x) \; ,$$
$$\beta(\mathcal{P}) := \min_{x^{(1)}, \dots, x^{(\ell)} \in \Omega} \mathcal{P}(x^{(1)}, \dots, x^{(\ell)}) \; .$$

1.2.1 The case of two steps

In case of $\ell = 2$ we establish the following theorem:

▶ **Theorem 6** (cf. Theorem 12). *A two-step probability distribution with equal marginals $\mathcal{P}$ is same-set hitting if and only if $\alpha(\mathcal{P}) > 0$.*

Of course, if $\beta(\mathcal{P}) > 0$, then Theorem 6 follows from Theorem 3. However, we are not aware of any previous work in case $\beta(\mathcal{P}) = 0$, i.e., when the distribution is same-set hitting but not set hitting, in particular for the probability space from Section 1.1.

[1] That $\mathcal{P}$ is set hitting if (2) holds is a consequence of Lemma 8.3 in [10]. If (2) does not hold, an appropriate combination of dictators establishes a counterexample.

1.2.2 More than two steps

In a general case of an ℓ-step distribution with equal marginals, it is still clear that $\alpha(\mathcal{P}) > 0$ is necessary. However, it remains open if it is sufficient.

We provide the following partial results. Firstly, by a simple inductive argument based on Theorem 12, we show that multi-step probability spaces induced by Markov chains are same-set hitting.

Secondly, we show that $\mathcal{P}$ is same-set hitting if $\alpha(\mathcal{P}) > 0$ and its *correlation* $\rho(\mathcal{P})$ is smaller than 1. The opposite condition $\rho(\mathcal{P}) = 1$ is equivalent to the following: There exist $j \in [\ell]$, $S \subseteq \Omega$, $T \subseteq \Omega^{\ell-1}$ such that $0 < |S| < |\Omega|$ and:

$$X_i^{(j)} \in S \iff \left(X_i^{(1)}, \dots, X_i^{(j-1)}, X_i^{(j+1)}, \dots, X_i^{(\ell)}\right) \in T .$$

For the full definition of $\rho(\mathcal{P})$, cf. Definition 10.

▶ **Theorem 7** (cf. Theorem 13). *Let $\mathcal{P}$ be a probability distribution with equal marginals. If $\alpha(\mathcal{P}) > 0$ and $\rho(\mathcal{P}) < 1$, then $\mathcal{P}$ is same-set hitting.*

We are not aware of any general results in case $\rho(\mathcal{P}) = 1$. In particular, let $\mathcal{P}$ be a three-step distribution over $\Omega = \{0, 1, 2\}$ such that $X_i^{(1)}, X_i^{(2)}, X_i^{(3)}$ is uniform over $\{000, 111, 222, 012, 120, 201\}$. To the best of our knowledge, it is an open question whether this distribution $\mathcal{P}$ is same-set hitting.

1.2.3 Set hitting for functions with no large Fourier coefficients

The methods developed here also allow to obtain lower bounds on the probability of hitting multiple sets. In fact, we show that if $\rho(\mathcal{P}) < 1$, then such lower bounds exist in terms of ρ, the measures of the sets and the largest non-empty Fourier coefficient.

▶ **Theorem 8** (Informal, cf. Theorem 14). *Let $\mathcal{P}$ be a probability distribution with $\rho(\mathcal{P}) < 1$. Then, $\mathcal{P}$ is set-hitting for functions $f^{(1)}, \dots, f^{(\ell)} : \underline{\Omega} \to [0, 1]$ that have both:*

- *Noticeable expectations, i.e.,* $\mathrm{E}[f^{(j)}(\underline{X}^{(j)})] \geq \Omega(1)$.
- *No large Fourier coefficients, i.e.,* $\max_\sigma \left|\hat{f}^{(j)}(\sigma)\right| \leq o(1)$.

1.3 Background and related work

The question of understanding which sets are hit often by dynamical systems is central to ergodic theory and additive combinatorics.

The Poincaré recurrence theorem states that measure preserving maps satisfy the property that for each set U of positive measure, and for almost every point x of U, the dynamical system started at x will hit U infinitely often, see, e.g., [12].

Much of the ergodic theory deals with quantifying this phenomena of repeatedly hitting the set U. The ergodic theorem, for example, implies that ergodic measure preserving dynamical systems satisfy that for almost all starting points x, the set U will be hit in limiting frequency which is equal to its measure.

Understanding set hitting by a number of consecutive steps of a process has been of intense study in additive combinatorics (where almost always $\rho = 1$).

For example, a well-studied case are random arithmetic progressions. Let Z be a finite additive group and $\ell \in \mathbb{N}$. Then, we can define a distribution $\mathcal{P}_{Z,\ell}$ of random ℓ-step arithmetic progressions in Z. Specifically, for every $x, r \in Z$ we set:

$$\mathcal{P}_{Z,\ell}(x, x + r, x + 2r, \dots, x + (\ell - 1)r) := 1/|Z|^2 .$$

Some of the distributions $\mathcal{P}_{Z,\ell}$ can be shown to be same-set hitting using the hypergraph regularity lemma:

▶ **Theorem 9** ([14], [15], [6], cf. Theorem 11.27, Proposition 11.28 and Exercise 11.6.3 in [17]). *If $|Z|$ is coprime to $(\ell-1)!$, then $\mathcal{P}_{Z,\ell}$ is same-set hitting.*

This follows a long line of work, starting by Szemerédi lemma [16], its proof by Furstenberg using the ergodic theorem [3] as well as the finite group and multi-dimensional versions, see, e.g., [13, 4, 5].

One might conjecture that $\alpha(\mathcal{P}) > 0$ is the sole sufficient condition for same-set hitting. Unfortunately, the techniques used to prove Theorem 9 do not seem to extend easily to less symmetric spaces. This suggests that proving the conjecture fully in $\rho = 1$ case might be a difficult undertaking.

The case of $\rho < 1$ has also been studied in the context of extremal combinatorics and hardness of approximation. In particular, Mossel [9] uses the invariance principle to prove that if $\rho(\mathcal{P}) < 1$, then $\mathcal{P}$ is set hitting for low-influence functions. We use this result to establish Theorem 7. Additionally, Theorem 8 can be seen as a strengthening of [9].

Furthermore, Austrin and Mossel [1] establish the result equivalent to Theorem 8 assuming in addition to $\rho(\mathcal{P}) < 1$ also that $\mathcal{P}$ is pairwise independent (they also prove results for the case $\rho(\mathcal{P}) = 1$ with pairwise independence but these involve only bounded degree functions).

Finally we note that another relevant paper in the case of $\ell = 2$ and symmetric $\mathcal{P}$ is by Dinur, Friedgut and Regev [2], who give a characterization of non-hitting sets. However, due to a different framework, their results are not directly comparable to ours.

We hope that our work might turn out to be useful in inapproximability. For example, our theorem is related to the proof of hardness for rainbow colorings of hypergraphs by Guruswami and Lee [7]. In particular, it is connected to their Theorem 4.3 and partially answers their Questions C.4 and C.6.

1.4 Outline of the paper

The rest of the paper is organised as follows: the notation is introduced in Section 2, Section 3 contains full statements of our theorems and Section 4 sketches the proof of our main theorem.

The full proofs along with he modified proof of the low-influence theorem from [9] can be found in the Arxiv version of the paper [8].

2 Notation and Preliminaries

2.1 Notation

We will now introduce our setting and notation. We refer the reader to Figure 1 for an overview.

We always assume that we have n independent coordinates. In each coordinate i we pick ℓ values $X_i^{(j)}$ for $j \in [\ell] = \{1, \ldots, \ell\}$ at random using some distribution. Each value $X_i^{(j)}$ is chosen from the same fixed set Ω, and the distribution of the tuple $\overline{X}_i = (X_i^{(1)}, \ldots, X_i^{(\ell)})$ of values from Ω^ℓ is given by a distribution $\mathcal{P}$.

This gives us values $X_i^{(j)}$ for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, \ell\}$. Thus, we have ℓ vectors $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}$, where $\underline{X}^{(j)} = (X_1^{(j)}, \ldots, X_n^{(j)})$ represents the j-th step of the random process. In case $\ell = 2$, we might call our two vectors $\underline{X}$ and $\underline{Y}$ instead.

For reasons outlined in Section 3.4.2 we assume that all of $X_i^{(1)}, \ldots, X_i^{(\ell)}$ have the same marginal distribution, which we call π. We assume that Ω is the support of π.

Even though it is not necessary, for clarity of the presentation we assume that each coordinate $\overline{X}_i = (X_i^{(1)}, \dots, X_i^{(j)}, \dots, X_i^{(\ell)})$ has the same distribution $\mathcal{P}$.

We consistently use index i to index over the coordinates (from $[n]$) and j to index over the steps (from $[\ell]$).

As visible in Figure 1, we denote the aggregation across the coordinates by the underline and the aggregation across the steps by the overline. For example, we write $\underline{\Omega} = \Omega^n$, $\overline{\Omega} = \Omega^\ell$, $\underline{\mathcal{P}} = \mathcal{P}^n$ and $\underline{\overline{X}} = (\overline{X}_1, \dots, \overline{X}_n) = (\underline{X}^{(1)}, \dots, \underline{X}^{(\ell)})$.

We sometimes call $\underline{\mathcal{P}}$ a tensorized, multi-step probability distribution as opposed to a tensorized, single-step distribution $\underline{\pi}$ and single-coordinate, multi-step distribution $\mathcal{P}$.

Furthermore, we extend the index notation to subsets of indices or steps. For example, for $S \subseteq [\ell]$ we define $X^{(S)}$ to be the collection of random variables $\left\{X^{(j)} : j \in S\right\}$.

We also use the set difference symbol to mark vectors with one element missing, e.g., $\overline{X}^{\setminus j} := (X^{(1)}, \dots, X^{(j-1)}, X^{(j+1)}, \dots, X^{(\ell)})$.

One should think of ℓ and $|\Omega|$ as constants and of n as large. We aim to get bounds which are independent of n.

2.2 Correlation

In case $\ell > 2$, the bound we obtain will depend on the *correlation* of the distribution $\mathcal{P}$. This concept was used before in [9].

▶ **Definition 10.** Let $\mathcal{P}$ be a single-coordinate distribution and let $S, T \subseteq [\ell]$. We define the *correlation*:

$$\rho(\mathcal{P}, S, T) := \sup\Big\{\mathrm{Cov}[f(X^{(S)}), g(X^{(T)})] \;\Big|\; f : \Omega^{(S)} \to \mathbb{R}, g : \Omega^{(T)} \to \mathbb{R},$$
$$\mathrm{Var}[f(X^{(S)})] = \mathrm{Var}[g(X^{(T)})] = 1\Big\} .$$

The correlation of $\mathcal{P}$ is $\rho(\mathcal{P}) := \max_{j \in [\ell]} \rho\left(\mathcal{P}, \{j\}, [\ell] \setminus \{j\}\right)$.

2.3 Influence

A crucial notion in the proof of Theorem 7 is the *influence* of a function. It expresses the average variance of a function, given that all but one of its n inputs have been fixed to random values:

▶ **Definition 11.** Let $\underline{X}$ be a random vector over alphabet $\underline{\Omega}$ and $f : \underline{\Omega} \to \mathbb{R}$ be a function and $i \in [n]$. The *influence of f on the i-th coordinate* is:

$$\mathrm{Inf}_i(f(\underline{X})) := \mathrm{E}\left[\mathrm{Var}\left[f(\underline{X}) \mid \underline{X}_{\setminus i}\right]\right] .$$

The *(total) influence of f* is $\mathrm{Inf}(f(\underline{X})) := \sum_{i=1}^n \mathrm{Inf}_i(f(\underline{X}))$.

Note that the influence depends both on the function f and the distribution of the vector $\underline{X}$.

3 Our Results

Here we give precise statements of our results presented in the introduction.

3.1 The case of $\ell = 2$

▶ **Theorem 12.** *Let Ω be a finite set and $\mathcal{P}$ a probability distribution over Ω^2 with equal marginals π. Let pairs (X_i, Y_i) be i.i.d. according to $\mathcal{P}$ for $i \in \{1, \ldots, n\}$.*

Then, for every $f : \Omega^n \to [0,1]$ with $\mathrm{E}[f(\underline{X})] = \mu > 0$:

$$\mathrm{E}[f(\underline{X})f(\underline{Y})] \geq c\,(\alpha(\mathcal{P}), \mu) \ , \tag{3}$$

where the function $c()$ is positive whenever $\alpha(\mathcal{P}) > 0$.

We remark that Theorem 12 does not depend on $\rho(\mathcal{P})$ in any way. This is in contrast to the case $\ell > 2$. It is possible to obtain a polynomially large explicit bound $c()$ for symmetric two-step spaces.

To prove Theorem 12 we make a convex decomposition argument and then apply the multi-step Theorem 13. For completeness, we provide a proof of Theorem 6 assuming Theorem 12.

Proof of Theorem 6. The "if" part follows from Theorem 12. The "only if" can be seen by taking f to be an appropriate dictator. ◀

3.2 The general case

▶ **Theorem 13.** *Let Ω be a finite set and $\mathcal{P}$ a distribution over Ω^ℓ in which all marginals are equal. Let tuples $\overline{X}_i = (X_i^{(1)}, \ldots, X_i^{(\ell)})$ be i.i.d. according to $\mathcal{P}$ for $i \in \{1, \ldots, n\}$.*

Then, for every function $f : \Omega^n \to [0,1]$ with $\mathrm{E}[f(\underline{X}^{(j)})] = \mu > 0$:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c\,(\alpha(\mathcal{P}), \rho(\mathcal{P}), \ell, \mu) \ , \tag{4}$$

where the function $c()$ is positive whenever $\alpha(\mathcal{P}) > 0$ and $\rho(\mathcal{P}) < 1$.

Furthermore, there exists some $D(\mathcal{P}) > 0$ (more precisely, D depends on α, ρ and ℓ) such that if $\mu \in (0, 0.99]$, one can take:

$$c(\alpha, \rho, \ell, \mu) := 1/\exp\left(\exp\left(\exp\left((1/\mu)^D\right)\right)\right) \ . \tag{5}$$

Note that this bound *does* depend on $\rho(\mathcal{P})$. We also obtain a bound that does not depend on $\rho(\mathcal{P})$ for multi-step probability spaces generated by Markov chains.

3.3 Hitting of different sets by uniform functions

Finally, we state the generalization of low-influence theorem from [9]. We assume that the reader is familiar with Fourier coefficients $\hat{f}(\sigma)$ and the basics of discrete function analysis, for details see, e.g., Chapter 8 of [11]. For the proof see the full version of the paper [8].

▶ **Theorem 14.** *Let $\overline{X}$ be a random vector distributed according to an ℓ-step distribution $\mathcal{P}$ with $\rho(\mathcal{P}) \leq \rho < 1$ and let $\mu^{(1)}, \ldots, \mu^{(\ell)} \in (0,1]$.*

There exist $k \in \mathbb{N}$ and $\gamma > 0$ (both depending only on $\mathcal{P}$ and $\mu^{(1)}, \ldots, \mu^{(\ell)}$) such that for all functions $f^{(1)}, \ldots, f^{(\ell)} : \underline{\Omega} \to [0,1]$, if $\mathrm{E}[f^{(j)}(\underline{X}^{(j)})] = \mu^{(j)}$ and $\max_{\sigma: 0 < |\sigma| \leq k} |\hat{f}^{(j)}(\sigma)| \leq \gamma$, then

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq c(\mathcal{P}, \mu^{(1)}, \ldots, \mu^{(\ell)}) > 0 \ . \tag{6}$$

3.4 Assumptions of the theorems

3.4.1 Equal distributions: unnecessary

In Theorems 12, 13 and 14 we assumed that the tuples $(X_i^{(1)}, \dots, X_i^{(\ell)})$ are distributed identically for each i. It is natural to ask if it is indeed necessary.

This is not the case. Instead, we made this assumption for simplicity of notation and presentation. If one is interested in statements which are valid where coordinate i is distributed according to $\mathcal{P}_i$, one simply needs to assume that there are $\alpha > 0$ and $\rho < 1$ such that $\alpha(\mathcal{P}_i) \geq \alpha$ and $\rho(\mathcal{P}_i) \leq \rho$.

3.4.2 Equal marginals: necessary

We quickly discuss the case when $\mathcal{P}$ does not have equal marginals. Recall that $\beta(\mathcal{P}) = \min_{x^{(1)},\dots,x^{(\ell)} \in \Omega} \mathcal{P}(x^{(1)}, \dots, x^{(\ell)})$. If $\beta(\mathcal{P}) > 0$, then, by Theorem 3, $\mathcal{P}$ is set hitting, and therefore also same-set hitting.

In case $\beta(\mathcal{P}) = 0$, we demonstrate an example which shows that $\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right]$ can be exponentially small in n. For concreteness, we set $\ell := 2$ and $\Omega := \{0, 1\}$ and consider $\mathcal{P}$ which picks uniformly among $\{00, 01, 11\}$. We then set

$$S_1 := \{(x_1, \dots, x_n) \mid x_1 = 1 \wedge |\mathrm{wt}(x) - n/3| \leq 0.01n\} \tag{7}$$

$$S_2 := \{(x_1, \dots, x_n) \mid x_1 = 0 \wedge |\mathrm{wt}(x) - 2n/3| \leq 0.01n\} \tag{8}$$

where $\mathrm{wt}(x)$ is the Hamming-weight of x, i.e., the number of ones in x.

For large enough n, a concentration bound implies that $\Pr[\underline{X}^{(1)} \in S_1] > \frac{1}{3} - 0.01$ and $\Pr[\underline{X}^{(2)} \in S_2] > \frac{1}{3} - 0.01$. Hence, if we set f to be the indicator function of $S := S_1 \cup S_2$, the assumption of Theorem 13 holds. However, because of the first coordinate we have $\Pr[\underline{X}^{(1)} \in S \wedge \underline{X}^{(2)} \in S] \leq \Pr[\underline{X}^{(1)} \in S_2 \vee \underline{X}^{(2)} \in S_1]$, and the right hand side is easily seen to be exponentially small.

It is not difficult to extend this example to any distribution with $\beta(\mathcal{P}) = 0$ that does not have equal marginals.

4 Proof Sketch

In this section we briefly outline the proof of Theorem 13. For simplicity, we assume that the probability space is the one from Section 1.1, i.e., (X_i, Y_i) are distributed uniformly in $\{00, 11, 22, 01, 12, 20\}$. Additionally, we assume that we are given a set $S \subseteq \{0, 1, 2\}^n$ with $\mu(S) = |S|/3^n > 0$, so that we want a bound of the form

$$\Pr\left[\underline{X} \in S \wedge \underline{Y} \in S\right] \geq c(\mu) > 0 \,.$$

The proof consists of three steps. Intuitively, in the first step we deal with dictator sets, e.g., $S_{\mathrm{dict}} = \{\underline{x} : x_1 = 0\}$, in the second step with linear sets, e.g., $S_{\mathrm{lin}} = \{\underline{x} : \sum_{i=1}^{n} x_i \pmod 3 = 0\}$ and in the third step with threshold sets, e.g., $S_{\mathrm{thr}} = \{\underline{x} : |\{i : x_i = 0\}| \geq n/3\}$.

4.1 Step 1 – making a set resilient

We call a set resilient if $\Pr[\underline{X} \in S]$ does not change by more than a (small) multiplicative constant factor whenever conditioned on $(X_{i_1} = x_{i_1}, \dots, X_{i_s} = x_{i_s})$ on a constant number s of coordinates.

In particular, S_{dict} is not resilient (because conditioning on $x_1 = 0$ increases the measure of the set to 1), while S_{lin} and S_{thr} are.

If a set is not resilient, using $\mathcal{P}(x, x) = 1/6$ for every $x \in \Omega$, one can find an event $\mathcal{E} :\equiv X_{i_1} = Y_{i_1} = x_{i_1} \wedge \ldots \wedge X_{i_s} = Y_{i_s} = x_{i_s}$ such that for some constant $\epsilon > 0$ we have $\Pr[\mathcal{E}] \geq \epsilon$ and, at the same time, $\Pr[\underline{X} \in S \mid \mathcal{E}] \geq (1+\epsilon)\Pr[\underline{X} \in S]$.

Since each such conditioning increases the measure of the set S by a constant factor, S must become resilient after a constant number of its iterations. Furthermore, each conditioning induces only a constant factor loss in $\Pr[\underline{X} \in S \wedge \underline{Y} \in S]$.

4.2 Step 2 – eliminating high influences

In this step, assuming that S is resilient, we condition on a constant number of coordinates to transform it into two sets S' and T' such that:

- Both of them have low influences on all coordinates.
- Both of them are supersets of S (after the conditioning).

The first property allows us to apply low-influence set hitting from [9] to S' and T'. The second one, together with the resilience of S, ensures that $\mu(S'), \mu(T') \geq (1-\epsilon)\mu(S)$.

In fact, it is more convenient to assume that we are initially given two resilient sets S and T.

Assume w.l.o.g. that $\mathrm{Inf}_i(T) \geq \tau$ for some $i \in [n]$. Given $z \in \{0, 1, 2\}$, let $T_z := \{\underline{x} : \underline{x} \in T \wedge x_i = z\}$. Furthermore, let $T_z^* := T_z \cup T_{z+1 \pmod 3}$.

Since $\mathrm{Inf}_i(T) \geq \tau$, we can show that there exists $z \in \{0, 1, 2\}$ such that, after conditioning on $X_i = Y_i = z$, the sum $\mu(S_z) + \mu(T_z^*)$ is strictly greater than the sum $\mu(S) + \mu(T)$:

$$\Pr[\underline{X} \in S_z \mid X_i = z] + \Pr[\underline{Y} \in T_z^* \mid Y_i = z] \geq \Pr[\underline{X} \in S] + \Pr[\underline{Y} \in T] + c(\tau)\ . \tag{9}$$

We choose to delete the coordinate i and replace S with $S' := S_z$ and T with $T' := T_z^*$. Equation (9) implies that after a constant number of such operations, neither S nor T has any remaining high-influence coordinates.

Crucially, with respect to same-set hitting our set replacement is essentially equivalent to conditioning on $X_i = z$ and $Y_i = z \vee Y_i = z + 1 \pmod 3$. Therefore, each operation induces only a constant factor loss in $\Pr[\underline{X} \in S \wedge \underline{Y} \in T]$.

4.3 Step 3 – applying low-influence theorem from [9]

Once we are left with two low-influence, somewhat-large sets S and T, we obtain $\Pr[\underline{X} \in S \wedge \underline{Y} \in T] \geq c(\mu) > 0$ by a straightforward application of a slightly modified version of Theorem 1.14 from [9]. The theorem gives that $\rho(\mathcal{P}) < 1$ implies that the distribution $\mathcal{P}$ is set hitting for low-influence functions:

▶ **Theorem 15.** *Let $\overline{\underline{X}}$ be a random vector distributed according to $(\underline{\overline{\Omega}}, \underline{\mathcal{P}})$ such that $\mathcal{P}$ has equal marginals, $\rho(\mathcal{P}) \leq \rho < 1$ and $\min_{x\in\Omega} \pi(x) \geq \alpha > 0$.*

Then, for all $\epsilon > 0$, there exists $\tau := \tau(\epsilon, \rho, \alpha, \ell) > 0$ such that if functions $f^{(1)}, \ldots, f^{(\ell)} : \underline{\Omega} \to [0, 1]$ satisfy

$$\max_{i\in[n], j\in[\ell]} \mathrm{Inf}_i(f^{(j)}(\underline{X}^{(j)})) \leq \tau\ , \tag{10}$$

then, for $\mu^{(j)} := \mathrm{E}[f^{(j)}(\underline{X}^{(j)})]$:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq \left(\prod_{j=1}^{\ell} \mu^{(j)}\right)^{\ell/(1-\rho^2)} - \epsilon\ . \tag{11}$$

Furthermore, there exists an absolute constant $C \geq 0$ such that for $\epsilon \in (0, 1/2]$ one can take

$$\tau := \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}}\right)^{C^{\frac{\ell \ln(\ell/\epsilon)\ln(1/\alpha)}{(1-\rho)\epsilon}}} . \tag{12}$$

The proof of Theorem 15 can be found in the Arxiv version of the paper [8].

4.4 The case $\rho = 1$: open question

Theorem 13 requires that $\rho < 1$ in order to give a meaningful bound. It is unclear whether this is an artifact of our proof or if it is necessary. In particular, consider the three step distribution $\mathcal{P}$ which picks a uniform triple from $\{000, 111, 222, 012, 120, 201\}$. One easily checks that $\rho(\mathcal{P}) = 1$ and that all marginals are uniform. We do not know if this distribution is same-set hitting.

However, the method of our proof breaks down. We illustrate the reason in the following lemma.

▶ **Lemma 16.** *For every $n > n_0$ there exist three sets $S^{(1)}$, $S^{(2)}$, and $S^{(3)}$ such that for the distribution $\mathcal{P}$ as described above we have*

- $\forall j : \Pr[\underline{X}^{(j)} \in S^{(j)}] \geq 0.49$.
- $\Pr[\forall j : \underline{X}^{(j)} \in S^{(j)}] = 0$.
- *The characteristic functions $1_{S^{(j)}}$ of the three sets all satisfy*

$$\max_{i \in [n]} \mathrm{Inf}_i(1_{S^{(j)}}(\underline{X}^{(j)})) \to 0 \text{ as } n \to \infty .$$

While the lemma does not give information about whether $\mathcal{P}$ is same-set hitting, it shows that our proof fails (since the analogue of Theorem 15 fails).

Proof. We let

$$S^{(1)} := \{\underline{x}^{(1)} : \underline{x}^{(1)} \text{ has less than } n/3 \text{ twos}\} ,$$
$$S^{(2)} := \{\underline{x}^{(2)} : \underline{x}^{(2)} \text{ has less than } n/3 \text{ ones}\} ,$$
$$S^{(3)} := \{\underline{x}^{(3)} : \underline{x}^{(3)} \text{ has less than } n/3 \text{ zeros}\} .$$

Whenever we pick $\underline{X}^{(1)}, \underline{X}^{(2)}, \underline{X}^{(3)}$, the number of twos in $\underline{X}^{(1)}$ plus the number of ones in $\underline{X}^{(2)}$ plus the number of zeros in $\underline{X}^{(3)}$ always equals n (there is a contribution of one from each coordinate). All three properties are now easy to check. ◀

References

1. Per Austrin and Elchanan Mossel. Noise correlation bounds for uniform low degree functions. *Arkiv för Matematik*, 51(1):29–52, 2013. doi:10.1007/s11512-011-0145-5.
2. Irit Dinur, Ehud Friedgut, and Oded Regev. Independent sets in graph powers are almost contained in juntas. *Geometric and Functional Analysis*, 18(1):77–97, 2008.
3. Harry Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *Journal d'Analyse Mathématique*, 31(1):204–256, 1977. doi:10.1007/BF02813304.
4. Harry Furstenberg and Yitzhak Katznelson. A density version of the Hales-Jewett theorem. *Journal d'Analyse Mathématique*, 57(1):64–119, 1991. doi:10.1007/BF03041066.

5 W. T. Gowers. A new proof of Szemerédi's theorem. *Geometric & Functional Analysis GAFA*, 11(3):465–588, 2001. doi:10.1007/s00039-001-0332-9.

6 W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics*, 166(3):897–946, 2007. Available from: http://www.jstor.org/stable/20160083.

7 Venkatesan Guruswami and Euiwoong Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 822–836, 2015. doi:10.1137/1.9781611973730.56.

8 Jan Hązła, Thomas Holenstein and Elchanan Mossel. Lower Bounds on Same-Set Inner Product in Correlated Spaces, 2015. Arxiv https://arxiv.org/abs/1509.06191.

9 Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010. doi:10.1007/s00039-010-0047-x.

10 Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013. doi:10.1007/s00039-013-0229-4.

11 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. Available from: http://www.cambridge.org/de/academic/subjects/computer-science/algorithmics-complexity-computer-algebra-and-computational-g/analysis-boolean-functions.

12 Karl Petersen. *Ergodic theory*. Cambridge University Press, corrected edition, 1989. Available from: http://opac.inria.fr/record=b1088586.

13 Klaus F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, s1-28(1):104–109, 1953. doi:10.1112/jlms/s1-28.1.104.

14 Vojtěch Rödl and Jozef Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures & Algorithms*, 25(1):1–42, 2004. doi:10.1002/rsa.20017.

15 Vojtěch Rödl and Jozef Skokan. Applications of the regularity lemma for uniform hypergraphs. *Random Structures & Algorithms*, 28(2):180–194, 2006. doi:10.1002/rsa.20108.

16 Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27(1):199–245, 1975.

17 Terence Tao and Van H. Vu. *Additive Combinatorics*. Cambridge University Press, 2006. Available from: http://opac.inria.fr/record=b1122796.