Product Space Models of Correlation: Between Noise Stability and Additive Combinatorics

arXiv:1509.06191v3 [cs.DM] 23 Dec 2018

Product Space Models of Correlation: Between Noise Stability and Additive Combinatorics

Jan Hązła*   Thomas Holenstein   Elchanan Mossel†

Abstract: There is a common theme to some research questions in additive combinatorics and noise stability. Both study the following basic question: Let $\mathcal{P}$ be a probability distribution over a space 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 $\mathcal{P}$.

A central question that is addressed in both areas is:

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

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

Instances of this question include the finite field model versions of Roth's and Szemerédi's theorems as well as Borell's result about the optimality of noise stability of half-spaces.

Our goal in this paper is to interpolate between the noise stability theory and the finite field additive combinatorics theory and address the question above in greater generality than considered before. In particular, we settle the question for $\ell = 2$ and when $\ell > 2$ and

$\mathcal{P}$ has bounded correlation $\rho(\mathcal{P}) < 1$. Under the same conditions we also characterize the *obstructions* for similar lower bounds in the case of ℓ different functions. Part of the novelty in our proof is the combination of analytic arguments from the theories of influences and hyper-contraction with arguments from additive combinatorics.

Key words and phrases: correlated product spaces, invariance principle, noise stability

1 Introduction

1.1 Setup and same-set hitting

In this paper we analyze a general framework which includes many fundamental questions in both the theory of noise stability and in finite field models of additive combinatorics. We begin with formally defining this general setting. Let Ω be a finite set 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 1.1. Let $\mu, \delta \in (0, 1]$. We say that a distribution $\mathcal{P}$ *is (μ, δ)-same-set hitting*, if, for all $n \geq 1$, 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.

It is not difficult to see that the definition of same-set hitting is equivalent to the one where functions f are restricted to be set indicators $f : \Omega^n \to \{0, 1\}$. The value $\mathrm{E}\left[\prod_{j=1}^{\ell} f\left(\underline{X}^{(j)}\right)\right]$ then can be interpreted as $\Pr\left[\bigwedge_{j=1}^{\ell} \underline{X}^{(j)} \in S\right]$ for the respective set $S := \{\underline{x} : f(\underline{x}) = 1\}$ of density at least μ. This special case motivated the name "same-set hitting", and all our theorems and proofs can be read with that case in mind.

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

The question of set hitting was studied extensively in additive combinatorics and in the theory of influences and noise stability. Perhaps the most well-studied case is that of 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:

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

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

Theorem 1.2 ([RS04], [RS06], [Gow07], cf. Theorem 11.27, Proposition 11.28 and Exercise 11.6.3 in [TV06]). *If $|Z|$ is coprime to $(\ell-1)!$, then $\mathcal{P}_{Z,\ell}$ is same-set hitting.*

Taking $\ell = p$ and $Z = \mathbb{F}_p$ we obtain the classical formulation of Szemerédi's theorem for progressions of length p in the finite field model. The special case $\ell = 3$ is also known as the *capset problem*. As is well known, the case $\ell = 3$ follows from the arguments of Roth [Rot53] applied to the finite field setup [Mes95], while the general case follows a long line of work, starting by Szemerédi's regularity lemma [Sze75], its proof by Furstenberg using the ergodic theorem [Fur77] as well as the finite group and multi-dimensional versions, see, e.g., [Rot53, FK91, Gow01, Gre05a].

It is natural to consider a generalization of the question where different functions are applied to different $X^{(j)}$. This question was studied in the theories of Gaussian noise stability and hyper-contraction as we explain next.

1.2 Set hitting

The generalization to multiple sets is defined as follows.

Definition 1.3. Let $\mu, \delta \in (0,1]$. We say that a distribution $\mathcal{P}$ *is* (μ, δ)-*set hitting*, if, whenever functions $f^{(1)}, \ldots, f^{(\ell)} : \Omega^n \to [0,1]$ satisfy $\mathrm{E}[f^{(j)}(\underline{X}^{(j)})] \geq \mu$ for every $j \in [\ell]$, we have

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

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.

Borell [Bor85] established the set hitting property in the Gaussian case where $(X_i, Y_i) \sim \mathcal{N}(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$ are i.i.d and $\rho \in (0,1)$. In fact [Bor85] does much more: it finds the optimal δ in terms of μ and ρ in this case. (Note that in this case Ω is infinite).

In earlier work, [Bor82] Borell also proved some of the first reverse hypercontractive inequalities. These give a different proof that the Gaussian example above is set hitting but also imply the same for the binary analog where $(X_i, Y_i) \in \{-1, 1\}^2$ satisfy $E[X_i] = E[Y_i] = 0$ and $E[X_i Y_i] = \rho$. See [MOR$^+$06] for a discussion of this result and some of its implications.

The full classification of set hitting distributions can be deduced from a paper on reverse hypercontractivity[1] by Mossel, Oleszkiewicz and Sen [MOS13]:

Theorem 1.4 ([MOS13]). *A finite probability space $\mathcal{P}$ is set hitting if and only if:*

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

In many interesting settings, including the finite field models in additive combinatorics, the distribution $\mathcal{P}$ does not have full support. In these settings, as we discuss next, the goal is to understand sufficient conditions on the functions which imply that (1) does hold.

---

[1]That $\mathcal{P}$ is set hitting if (2) holds is a consequence of Lemma 8.3 in [MOS13]. If (2) does not hold, an appropriate combination of dictators establishes a counterexample.

1.3 Obstructions in additive combinatorics

In general much of the interest in additive combinatorics is in understanding what conditions on functions f imply (1). For example, the starting point of the proof of Roth's theorem [Rot53] on arithmetic progressions of length three is that if the functions $f^{(1)}, f^{(2)}, f^{(3)}$ all satisfy that $\|\widehat{f^{(j)} - E[f]}\|_\infty$ is small then (1) holds. That is, the distribution of arithmetic progressions of length three is set hitting for all functions $f^{(j)}$ with all (positive degree) Fourier coefficients small in absolute value. As a matter of fact, in that case $f^{(1)}, f^{(2)}, f^{(3)}$ are known to be pseudorandom in the sense that $\delta(\mu) \approx \mu^3$.

The proof of Roth's theorem then proceeds roughly as follows: If a function f is pseudorandom, we are done. Otherwise, we are guaranteed a large Fourier coefficient. This is then exploited in a density increment argument: It turns out that a large Fourier coefficient implies that f must have increased relative density on an affine subspace of $\mathbb{F}_p^n$ of codimension one. One iterates the density increment until f becomes pseudorandom.

A similar situation arises in a more recent proof for longer arithmetic progression by Gowers: If the functions $f^{(j)}$ have low Gowers uniformity norm, then (1) holds, see e.g. [Gre05a].

In one of our main results (see Section 1.5.3 below) we show that in a pretty general setup (which does not include the additive combinatorics setup), the only obstruction for (1) to hold is for $f^{(j)}$ to have a large low-degree Fourier coefficient.

1.4 Basic example

At this point we would like to introduce the simplest example that is not covered by either the theory of influences or techniques from additive combinatorics. 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 $\{X_i, X_i + 1 \mod 3\}$. 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. Similarly, given sets S and T of density at least μ, we want to find under what conditions does it hold that the probability $\Pr[\underline{X} \in S \wedge \underline{Y} \in S]$ can be lower bounded effectively. We note that the support of the distribution on $\{0,1,2\}^2$ is not full (hence, Theorem 1.4 does not apply) and that the distribution is not of arithmetic nature.

1.5 Our results

1.5.1 Same-set hitting for two steps

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

Theorem 1.5 (cf. Theorem 3.1). *A two-step probability distribution with equal marginals $\mathcal{P}$ is same-set hitting if and only if $\alpha(\mathcal{P}) := \min_{x \in \Omega} \mathcal{P}(x,x) > 0$.*

Of course, if $\beta(\mathcal{P}) > 0$, then Theorem 1.5 follows from Theorem 1.4. Our work is novel in case $\beta(\mathcal{P}) = 0$, i.e., when the distribution is same-set hitting but not set hitting. In particular we establish same-set hitting for the probability space from Section 1.4.

1.5.2 Same-set hitting for more than two steps

In a general case of an ℓ-step distribution with equal marginals, it is still clear that, letting $\alpha(\mathcal{P}) := \min_{x \in \Omega} \mathcal{P}(x, x, \ldots, x)$, the condition $\alpha(\mathcal{P}) > 0$ 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 3.1, we show that multi-step probability spaces induced by Markov chains are same-set hitting (cf. Section 8).

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:

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

For the full definition of $\rho(\mathcal{P})$, see Definition 2.1.

Theorem 1.6 (cf. Theorem 3.2). *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 $X_i^{(1)}, X_i^{(2)}, X_i^{(3)}$ are 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. One might conjecture that $\alpha(\mathcal{P}) > 0$ is the sole sufficient condition for same-set hitting. Unfortunately, the techniques used to prove Theorem 1.2 do not seem to extend easily to spaces with less algebraic structure.

1.5.3 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 1.7 (Informal, cf. Theorem 3.3). *Let $\mathcal{P}$ be a probability distribution with $\rho(\mathcal{P}) < 1$. Then, $\mathcal{P}$ is set-hitting for functions $f^{(1)}, \ldots, f^{(\ell)} : \underline{\Omega} \to [0, 1]$ that have both:*

- *Noticeable expectations, i.e., $\mathrm{E}[f^{(j)}(\underline{X}^{(j)})] \geq \Omega(1)$.*

- *No large Fourier coefficients, i.e., $\max_\sigma |\hat{f}^{(j)}(\sigma)| \leq o(1)$.*

1.6 Other related work

In the case of symmetric two-step spaces (which can be thought of as product graphs) works by Dinur, Friedgut and Regev [DFR08, FR18] establish a removal lemma: They show that if $\Pr[\underline{X} \in S \wedge \underline{Y} \in S]$

is small, then it must be possible to remove a small number of elements from S to obtain S' with $\Pr[\underline{X} \in S' \wedge \underline{Y} \in S'] = 0$. They go on to use this result to characterize all sets with $\Pr[\underline{X} \in S \wedge \underline{Y} \in S] = 0$: It turns out that every such set must be almost contained in a junta. Interestingly, [FR18] obtain a tower-type dependence between μ and δ in the removal lemma, in contrast to ours which is "merely" triply exponential.

The case of $\rho < 1$ has also been studied in the context of extremal combinatorics and hardness of approximation. In particular, Mossel [Mos10] 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 1.6. Additionally, Theorem 1.7 can be seen as a strengthening of [Mos10].

Furthermore, Austrin and Mossel [AM13] establish the result equivalent to Theorem 1.7 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).

Our work is related to problems and results in inapproximability in theoretical computer science. For example, our theorem is related to the proof of hardness for rainbow colorings of hypergraphs by Guruswami and Lee [GL15]. In particular, it is connected to their Theorem 4.3 and partially answers their Questions C.4 and C.6.

There are works in additive combinatorics that treat specific classes of distributions with $\rho = 1$. For example, one can take $\mathcal{P}$ to be uniform over solutions to a fixed full-rank system of r linear equations with ℓ variables over $\mathbb{F}_p$. There is extensive work on removal lemmas (which imply same-set hitting) for different cases in this setting, see, e.g., [Gre05b, KSV09, Sha10, FLS18].

Follow-up work There are two subsequent preprints by some of the authors: [Mos17] strengthens Theorem 3.3 to obtain precise Gaussian bounds for functions with small low-degree Fourier coefficients in case $\rho(\mathcal{P}) < 1$ (one can also use the technique from [Mos17] to deduce an alternative proof of Theorem 3.2 with roughly the same dependence). Another author [Hąz18] shows same-set hitting for symmetric sets for the distribution of arithmetic progressions with restricted differences mentioned in Section 4.4.

1.7   Proof ideas: additive combinatorics and theory of influences

Interestingly, the proof of our results interpolates between additive combinatorics and the theory of influences. Results of [Mos10] imply that if a collection of functions have low influences then they are same-set hitting. In the proof of Theorem 3.2 we apply a variant of a density increment argument to reduce to this case. First, we apply the standard density increment argument to assume without loss of generality that conditioning on a small number of coordinates does not change the measure of the set by much. Then we show, under this assumption, by applying another variant of density increment that we can additionally assume w.l.o.g. that all influences are small.

1.8   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.
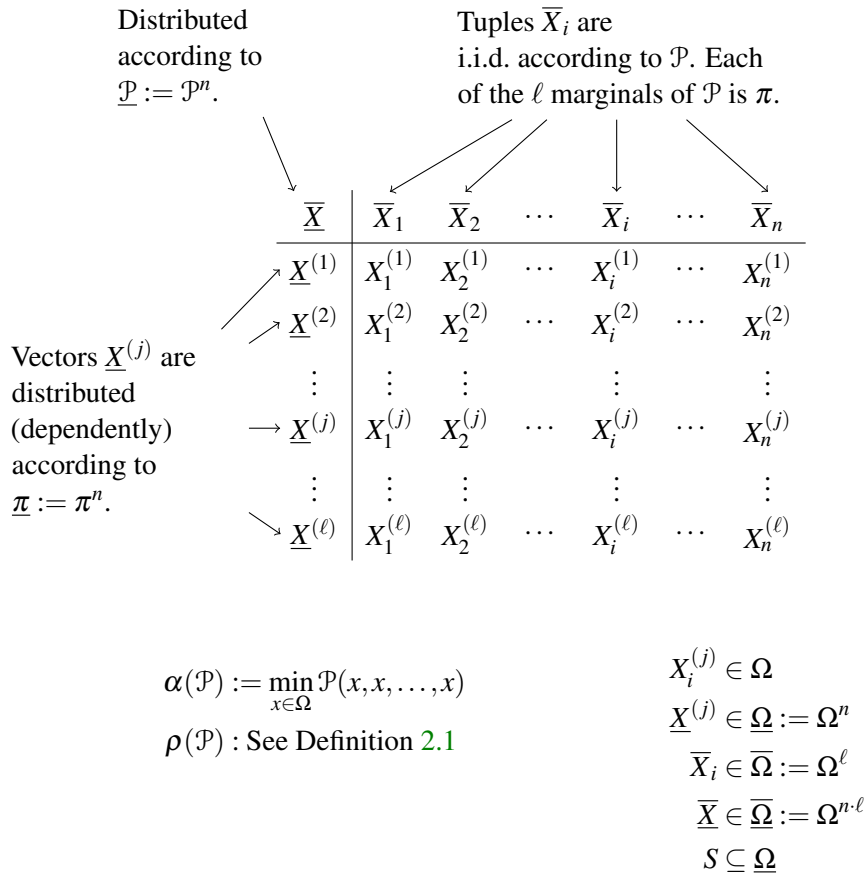
Distributed
according to
$\underline{\mathcal{P}} := \mathcal{P}^n$.

Tuples $\overline{X}_i$ are
i.i.d. according to $\mathcal{P}$. Each
of the ℓ marginals of $\mathcal{P}$ is π.

Vectors $\underline{X}^{(j)}$ are
distributed
(dependently)
according to
$\underline{\pi} := \pi^n$.

$\overline{X}$	$\overline{X}_1$	$\overline{X}_2$	$\cdots$	$\overline{X}_i$	$\cdots$	$\overline{X}_n$
$\underline{X}^{(1)}$	$X_1^{(1)}$	$X_2^{(1)}$	$\cdots$	$X_i^{(1)}$	$\cdots$	$X_n^{(1)}$
$\underline{X}^{(2)}$	$X_1^{(2)}$	$X_2^{(2)}$	$\cdots$	$X_i^{(2)}$	$\cdots$	$X_n^{(2)}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\underline{X}^{(j)}$	$X_1^{(j)}$	$X_2^{(j)}$	$\cdots$	$X_i^{(j)}$	$\cdots$	$X_n^{(j)}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\underline{X}^{(\ell)}$	$X_1^{(\ell)}$	$X_2^{(\ell)}$	$\cdots$	$X_i^{(\ell)}$	$\cdots$	$X_n^{(\ell)}$

$$\alpha(\mathcal{P}) := \min_{x \in \Omega} \mathcal{P}(x, x, \ldots, x)$$
$$\rho(\mathcal{P}) : \text{See Definition 2.1}$$

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

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 $\underline{\mathcal{P}}$.

The full proof of the multi-step theorem follows in Section 5. The proof of the two-step theorem is in Section 6 and the proof for functions with small Fourier coefficients in Section 7. A theorem for Markov chains is introduced in Section 8 and better bounds for symmetric spaces in Section 9. Finally, the modified proof of the low-influence theorem from [Mos10] is presented in the appendix. We note that an extended abstract of our results appeared in [HHM16].

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 $\overline{X}_i = (X_i^{(1)}, \ldots, X_i^{(j)}, \ldots, X_i^{(\ell)})$ 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 $\underline{\overline{X}} = (\overline{X}_1, \ldots, \overline{X}_n) = (\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)})$.

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 $\{X^{(j)} : j \in S\}$.

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

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 [Mos10].

Definition 2.1. Let $\mathcal{P}$ be a single-coordinate distribution and let $S, T \subseteq [\ell]$. We define the *correlation*:

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

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 1.6 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 2.2. 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 3.1. *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 $i \in \{1, \dots, n\}$.*

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

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

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

We remark that Theorem 3.1 does not depend on $\rho(\mathcal{P})$ in any way. This is in contrast to the case $\ell > 2$. It is possible to obtain an inverse polynomial bound $c(\mu) \geq \mu^C$ for symmetric two-step spaces (see Section 9).

To prove Theorem 3.1 we make a convex decomposition argument and then apply the multi-step Theorem 3.2 (see Section 6). For completeness, we provide a proof of Theorem 1.5 assuming Theorem 3.1.

Proof of Theorem 1.5. The "if" part follows from Theorem 3.1. The "only if" can be seen by taking f to be an appropriate dictator. □

3.2 The general case

Theorem 3.2. *Let Ω be a finite set and $\mathcal{P}$ a distribution over Ω^ℓ in which all marginals are equal. Let tuples $\overline{X}_i = (X_i^{(1)}, \dots, X_i^{(\ell)})$ be i.i.d. according to $\mathcal{P}$ for $i \in \{1, \dots, n\}$.*

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

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

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 (see Section 8).

3.3 Hitting of different sets by uniform functions

Finally, we state the generalization of low-influence theorem from [Mos10]. 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 [O'D14]. Note that this theorem requires neither equal marginals nor $\alpha(\mathcal{P}) > 0$. For the proof see Section 7.

Theorem 3.3. *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 3.1, 3.2 and 3.3 we assumed that the tuples $(X_i^{(1)}, \ldots, X_i^{(\ell)})$ 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 $\beta(\mathcal{P}) = \min_{x^{(1)}, \ldots, x^{(\ell)} \in \Omega} \mathcal{P}(x^{(1)}, \ldots, x^{(\ell)})$. If $\beta(\mathcal{P}) > 0$, then, by Theorem 1.4, $\mathcal{P}$ 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

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

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

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 3.2 holds. However, because of the first coordinate we have $\Pr[\underline{X}^{(1)} \in S \wedge \underline{X}^{(2)} \in S] \leq \Pr[\underline{X}^{(1)} \in S_2] + \Pr[\underline{X}^{(2)} \in S_1]$, 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 3.2. For simplicity, we assume that the probability space is the one from Section 1.4, 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

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

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

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 $(X_{i_1} = x_{i_1}, \ldots, X_{i_s} = x_{i_s})$ 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 $\varepsilon > 0$ we have $\Pr[\mathcal{E}] \geq \varepsilon$ and, at the same time, $\Pr[\underline{X} \in S \mid \mathcal{E}] \geq (1 + \varepsilon) \Pr[\underline{X} \in S]$.

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

It is worth noting that this is the only stage of the proof where we assume the same-set property (and utilize the assumption $\alpha(\mathcal{P}) > 0$).

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 conditioning).

The first property allows us to apply low-influence set hitting from [Mos10] to S' and T'. The second one, together with the resilience of S, ensures that $\mu(S'), \mu(T') \geq (1 - \varepsilon)\mu(S)$.

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 $\mathrm{Inf}_1(T) \geq \tau$ for some $i \in [n]$. Given $z \in \{0,1,2\}$, let $T_z := \{(x_1, x_2, \ldots, x_n) : (z, x_2, \ldots, x_n) \in T\}$. Furthermore, let $T_z^* := T_z \cup T_{z+1 \pmod 3}$.

Since $\mathrm{Inf}_1(T) \geq \tau$, we can show that there exists $z \in \{0,1,2\}$ such that, after conditioning on $X_1 = Y_1 = z$, the sum $\mu(S_z) + \mu(T_z^*)$ is strictly greater than the sum $\mu(S) + \mu(T)$:

$$\Pr[\underline{X} \in S_z \mid X_1 = z] + \Pr[\underline{Y} \in T_z^* \mid Y_1 = z] \geq \Pr[\underline{X} \in S] + \Pr[\underline{Y} \in T] + c(\tau) . \tag{9}$$

We choose to disregard the first coordinate 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 $X_1 = z$ and $Y_1 = z \lor Y_1 = z + 1 \pmod 3$. Therefore, each operation induces only a constant factor loss in $\Pr[\underline{X} \in S \land \underline{Y} \in T]$.

4.3 Step 3 — applying low-influence theorem from [Mos10]

Once we are left with two low-influence, somewhat-large sets S and T, we obtain $\Pr[\underline{X} \in S \land \underline{Y} \in T] \geq c(\mu) > 0$ by a straightforward application of a slightly modified version of Theorem 1.14 from [Mos10]. The theorem gives that $\rho(\mathcal{P}) < 1$ implies that the distribution $\mathcal{P}$ is set hitting for low-influence functions:

Theorem 4.1. *Let $\overline{X}$ be a random vector distributed according to $(\overline{\Omega}, \mathcal{P})$ 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 $\varepsilon > 0$, there exists $\tau := \tau(\varepsilon, \rho, \alpha, \ell) > 0$ 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 $\mu^{(j)} := \mathrm{E}[f^{(j)}(\underline{X}^{(j)})]$:

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

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

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

The proof of Theorem 4.1 can be found in Appendix A. The first part of the appendix contains a short explanation of differences between [Mos10] and our version.

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

Theorem 3.2 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 $\{000, 111, 222, 012, 120, 201\}$. In other words, sampling from $\underline{\mathcal{P}}$ picks a random arithmetic progression $x, x + d, x + 2d$ with $x \in \mathbb{F}_3^n$ and $d \in \{0, 1\}^n$. 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 4.2. *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 $\mathbb{1}_{S^{(j)}}$ of the three sets all satisfy*

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

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 4.1 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. $\qquad\square$

5   Proof for General ℓ and $\rho(\mathcal{P}) < 1$

The goal of this section is to prove our second main result, which we restate here for convenience.

Theorem 3.2. *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 $\mathrm{E}[f(\underline{X}^{(j)})] = \mu > 0$:

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

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:

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

5.1 Properties of the correlation

Recall Definition 2.1. We now give an alternative characterization of $\rho(\mathcal{P}, \{j\}, [\ell] \setminus \{j\})$ which will be useful later. For this, we first define certain random process and an associated Markov chain.

Definition 5.1. Let $\mathcal{P}$ be a single-coordinate distribution and let $j \in [\ell]$. We call a collection of random variables $(\overline{X}^{\setminus j} = (X^{(1)}, \dots, X^{(j-1)}, X^{(j+1)}, \dots, X^{(\ell)}), Y, Z)$ a *double sample on step j from* $\mathcal{P}$ if:

- $\overline{X}$ is first sampled according to $\mathcal{P}$, ignoring step j.

- Assuming that $\overline{X}^{\setminus j} = \overline{x}^{\setminus j}$, the random variables Y and Z are then sampled independently of each other according to the j-th step of $\mathcal{P}$ conditioned on $\overline{X}^{\setminus j} = \overline{x}^{\setminus j}$.

Sometimes we will omit $\overline{X}^{\setminus j}$ from the notation and refer as double sample to (Y, Z) alone.

An equivalent interpretation of a double sample is that after sampling $(\overline{X}^{\setminus j}, Y)$ according to $\mathcal{P}$ we "forget" about Y and sample Z again from the same distribution (keeping the same value of $\overline{X}^{\setminus j}$). Therefore, both $(\overline{X}^{\setminus j}, Y)$ and $(\overline{X}^{\setminus j}, Z)$ are distributed according to $\mathcal{P}$.

If we let
$$K(y, z) := \Pr[Z = z | Y = y] = \mathrm{E}\left[\Pr\left[Z = z | Y = y, \overline{X}^{\setminus j}\right]\right],$$

we see that

$$\pi(y)K(y, z) = \Pr[Y = y \wedge Z = z] = \Pr[Y = z \wedge Z = y] = \pi(z)K(z, y), \tag{13}$$

which means that K is the kernel of a Markov chain that is reversible with respect to π (see e.g., [LPW08, Section 1.6]). Thus, K has an orthonormal eigenbasis with eigenvalues $1 = \lambda_1(K) \geq \lambda_2(K) \geq \cdots \geq \lambda_{|\Omega|}(K) \geq -1$, (e.g., [LPW08, Lemma 12.2]). We will say that K is the *Markov kernel induced by the double sample* (Y, Z).

A standard fact from the Markov chain theory expresses $\lambda_2(K)$ in terms of covariance of functions $f \in L^2(\Omega, \pi)$:

Lemma 5.2 (Lemma 13.12 in [LPW08])**.** *Let Y, Z be two consecutive steps of a reversible Markov chain with kernel K such that both Y and Z are distributed according to a stationary distribution of K. Then,*

$$\lambda_2(K) = \max_{\substack{f: \Omega \to \mathbb{R} \\ \mathrm{E}[f(Y)] = 0 \\ \mathrm{Var}[f(Y)] = 1}} \mathrm{E}[f(Y)f(Z)] . \tag{14}$$

Lemma 5.3. *Let $\mathcal{P}$ be a single-coordinate distribution and let $(\overline{X}^{\backslash j}, Y, Z)$ be a double sample from $\mathcal{P}$ that induces a Markov kernel K. Then,*

$$\lambda_2(K) = \rho(\mathcal{P}, \{j\}, [\ell] \setminus \{j\})^2 .$$

Proof. For readability, let us write $\overline{X}$ instead of $\overline{X}^{\backslash j}$.

Consider first two functions f and g as in Definition 2.1 and assume without loss of generality that $\mathrm{E}[f(Y)] = \mathrm{E}[g(\overline{X})] = 0$. Of course, we also assume that $\mathrm{Var}[f(Y)] = \mathrm{Var}[g(\overline{X})] = 1$ as specified by Definition 2.1. We will show that

$$\mathrm{Cov}\left[f(Y), g(\overline{X})\right]^2 \leq \lambda_2(K) , \tag{15}$$

and that there exists a choice of f and g that achieves equality in (15).

Let $h(\overline{x}) := \mathrm{E}[f(Y)|\overline{X} = \overline{x}]$ and observe that

$$\begin{aligned}
\mathrm{E}[f(Y)f(Z)] &= \sum_{\overline{x},y,z} \Pr[\overline{X} = \overline{x}] \Pr[Y = y \mid \overline{X} = \overline{x}] \Pr[Z = z \mid \overline{X} = \overline{x}] f(y)f(z) \\
&= \mathrm{E}[h(\overline{X})^2] .
\end{aligned} \tag{16}$$

Now, by Cauchy-Schwarz, (16) and Lemma 5.2 we see that

$$\begin{aligned}
\mathrm{Cov}[f(Y), g(\overline{X})]^2 &= \mathrm{E}[f(Y)g(\overline{X})]^2 = \mathrm{E}[h(\overline{X})g(\overline{X})]^2 \leq \mathrm{E}[h(\overline{X})^2] \mathrm{E}[g(\overline{X})^2] \\
&= \mathrm{E}[h(\overline{X})^2] = \mathrm{E}[f(Y)f(Z)] \leq \lambda_2(K) .
\end{aligned}$$

The equality is obtained for f that maximizes the right-hand side of (14) and $g := c \cdot h$ for some $c > 0$. □

For later use, we make the following implication of Lemma 5.3.

Corollary 5.4. *Let (Y, Z) be a double sample on step j from a single-coordinate distribution $(\Omega, \mathcal{P})$ with $\rho(\mathcal{P}) = \rho$. Then, for every function $f : \Omega \rightarrow \mathbb{R}$,*

$$\mathrm{E}[(f(Y) - f(Z))^2] \geq 2(1 - \rho^2) \mathrm{Var}[f(Y)]. \tag{17}$$

Proof. Assume w.l.o.g. that $\mathrm{E}[f(Y)] = 0$. By Lemmas 5.2 and 5.3,

$$\mathrm{E}\left[(f(Y) - f(Z))^2\right] = 2\left(\mathrm{Var}[f(Y)] - \mathrm{E}[f(Y)f(Z)]\right) \geq 2(1 - \rho^2) \mathrm{Var}[f(Y)] .$$

□

5.2 Reduction to the resilient case

In this section, we will prove that we can assume that the function f is *resilient* in the following sense: whenever we fix a constant number of inputs to some value, the expected value of f remains roughly the same.

The intuitive reason for this is simple: if there is some way to fix the coordinates which changes the expected value of f, we can fix these coordinates such that the expected value *increases*, which only makes our task easier (and can be done only a constant number of times).

We first make the concept of "fixing" a subset of the coordinates formal.

Definition 5.5. Let $f : \underline{\Omega} \to [0,1]$ be a function. A *restriction* $\mathcal{R}$ is a sequence $\mathcal{R} = (r_1, \dots, r_n)$ where each r_i is either an element $r_i \in \Omega$, or the special symbol $r_i = \star$.

The coordinates with $r_i = \star$ are *unrestricted*, the coordinates where $r_i \in \Omega$ are *restricted*. The *size* of a restriction is the number of restricted coordinates.

A restriction $\mathcal{R}$ operates on a function f as

$$(\mathcal{R}f)(x_1, \dots, x_n) := f(y_1, \dots, y_n) \tag{18}$$

where $y_i = r_i$ if $r_i \neq \star$ and $y_i = x_i$ otherwise.

Next, we define what it means for a function to be resilient: restrictions do not change the expectation too much.

Definition 5.6. Let $\underline{X}$ be a random vector distributed according to a (single-step) distribution $(\underline{\Omega}, \underline{\pi})$. A function $f : \underline{\Omega} \to [0,1]$ is ε-*resilient up to size* k if for every restriction $\mathcal{R}$ of size at most k we have that $(1 - \varepsilon)\mathrm{E}[f(\underline{X})] \leq \mathrm{E}[\mathcal{R}f(\underline{X})] \leq (1 + \varepsilon)\mathrm{E}[f(\underline{X})]$.

The function is upper resilient if the expectation cannot increase too much.

Definition 5.7. Let $\underline{X}$ be a random vector distributed according to a distribution $(\underline{\Omega}, \underline{\pi})$. A function $f : \underline{\Omega} \to [0,1]$ is ε-*upper resilient up to size* k if for every restriction $\mathcal{R}$ of size at most k we have that $\mathrm{E}[\mathcal{R}f(\underline{X})] \leq (1 + \varepsilon)\mathrm{E}[f(\underline{X})]$.

Resilience and upper resilience are equivalent up to a multiplicative factor which depends only on k and the smallest probability in the marginal distribution $\alpha(\pi)$. Intuitively the reason is that if there is some restriction which decreases the 1-norm, then some other restriction on the same coordinates must increase the 1-norm somewhat.

Lemma 5.8. *Suppose that a function f is ε-upper resilient up to size k. Then, f is ε'-resilient up to size k, where $\varepsilon' = \varepsilon/(\alpha(\pi))^k$.*

Proof. Fix a subset $S \subseteq [n]$ of the coordinates of size $|S| \leq k$. We consider a random variable $\mathcal{R}$ whose values are restrictions with restricted coordinates being exactly S. The elements $r_i \in \Omega$ for $i \in S$ are picked according to the distribution π. We let $p(\mathcal{R}')$ be the probability a certain restriction $\mathcal{R}'$ is picked, and get

$$\mathrm{E}[f(\underline{X})] = \sum_{\mathcal{R}'} p(\mathcal{R}') \cdot \mathrm{E}\left[\mathcal{R}'f(\underline{X})\right], \tag{19}$$

where we sum over all restrictions $\mathcal{R}'$ that restrict exactly the coordinates in S.

Let now $\mathcal{R}^*$ be one of the possible choices for $\mathcal{R}$. Then,

$$
\begin{aligned}
p(\mathcal{R}^*) \cdot \mathrm{E}[\mathcal{R}^*f(\underline{X})] &= \mathrm{E}[f(\underline{X})] - \sum_{\mathcal{R}' \neq \mathcal{R}^*} p(\mathcal{R}') \cdot \mathrm{E}[\mathcal{R}'f(\underline{X})] \\
&\geq \mathrm{E}[f(\underline{X})] - (1 + \varepsilon) \sum_{\mathcal{R}' \neq \mathcal{R}^*} p(\mathcal{R}') \cdot \mathrm{E}[f(\underline{X})] \\
&= (1 - (1 + \varepsilon)(1 - p(\mathcal{R}^*))) \cdot \mathrm{E}[f(\underline{X})] \\
&\geq (p(\mathcal{R}^*) - \varepsilon) \cdot \mathrm{E}[f(\underline{X})],
\end{aligned}
$$

and hence:

$$\mathrm{E}[\mathcal{R}^* f(\underline{X})] \geq \left(1 - \frac{\varepsilon}{p(\mathcal{R}^*)}\right) \cdot \mathrm{E}[f(\underline{X})] \,.$$

Since $p(\mathcal{R}^*) \geq \alpha(\pi)^k$ we get the bound for the restriction $\mathcal{R}^*$, which was chosen arbitrarily. $\qquad\square$

Lemma 5.9. *Let $\overline{X}$ be a random vector distributed according to a distribution with equal marginals $(\underline{\Omega}, \underline{\mathcal{P}})$ and $f : \underline{\Omega} \to [0, 1]$ be a function with $\mathrm{E}[f(\underline{X}^{(1)})] = \mu > 0$.*
Let $\varepsilon \in (0, 1], k \in \mathbb{N}$. Then, there exists a restriction $\mathcal{R}$ such that $g := (\mathcal{R}f)$ is ε-resilient up to size k and

$$\mathrm{E}[g(\underline{X}^{(1)})] \geq \mu \,, \tag{20}$$

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g(\underline{X}^{(j)})\right] \,, \tag{21}$$

where $c := \exp\left(-\frac{2\ln 1/\mu}{\alpha^{2k} \cdot \varepsilon}\right)$ with $\alpha := \alpha(\mathcal{P}) > 0$.

In particular, c depends only on $\varepsilon, k, \alpha(\mathcal{P})$ and μ (requiring $\varepsilon, \alpha(\mathcal{P}), \mu > 0$).

Proof. Let $\varepsilon' := \alpha^k \cdot \varepsilon$ and choose a restriction $\mathcal{R}$ such that $\mathrm{E}[\mathcal{R}f(\underline{X}^{(1)})] \geq \mathrm{E}[f(\underline{X}^{(1)})] \cdot (1 + \varepsilon')$. We repeat this, replacing f with $(\mathcal{R}f)$, until there is no such restriction.

Since the expectation of f only increases, we get (20). Finally, once the process stops, the resulting function is ε-resilient due to Lemma 5.8 (note that $\alpha(\pi) \geq \alpha$).

It remains to argue that (21) holds for the resulting function. Note first that the expectation cannot exceed 1, and hence the process will be repeated at most $p := \ln(1/\mu)/\ln(1 + \varepsilon') \leq \frac{2\ln(1/\mu)}{\varepsilon'}$ times. Therefore, the final restriction $\mathcal{R}$ obtained after at most p iterations of the process above is of size at most pk.

Define $g := (\mathcal{R}f)$ and let $\mathcal{E}$ be the event that all strings $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}$ agree with the restriction $\mathcal{R}$ in its restricted coordinates. We will use $\mathbb{1}(\mathcal{E})$ to denote the function which is 1 if event $\mathcal{E}$ happens and 0 otherwise. We see that

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(X^{(j)})\right] \geq \mathrm{E}\left[\prod_{j=1}^{\ell} f(X^{(j)}) \cdot \mathbb{1}(\mathcal{E})\right] = \mathrm{E}\left[\prod_{j=1}^{\ell} g(X^{(j)}) \cdot \mathbb{1}(\mathcal{E})\right]$$

$$\geq \alpha^{pk} \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g(X^{(j)})\right] \,.$$

Finally,

$$\alpha^{pk} \geq \exp\left(-\frac{2k\ln(1/\alpha)\ln(1/\mu)}{\alpha^k \cdot \varepsilon}\right) \geq \exp\left(-\frac{2\ln 1/\mu}{\alpha^{2k} \cdot \varepsilon}\right) \,.$$

$\qquad\square$

5.3 Reduction to the low-influence case

We next show that if f is resilient, we can also assume that it has only low influences. However, this part of the proof actually produces a collection of functions $g^{(1)}, \ldots, g^{(\ell)}$ such that each of them has small influences: it operates differently on each function. In turn, it is more convenient to do this part of the proof also starting from a collection $f^{(1)}, \ldots, f^{(\ell)}$, as long as all of them are sufficiently resilient.

As in the previous section, we use restrictions. Here, however, we are only interested in restrictions of size one. Consequently, we write $\mathcal{R}[i, a]$ to denote the restriction $\mathcal{R} = (r_1, \ldots, r_n)$ with $r_i = a$ and $r_{i'} = \star$ for $i' \neq i$.

Furthermore, we require a new operator.

Definition 5.10. Let $f : \underline{\Omega} \to [0, 1]$, $i \in [n]$, and fix values $y, z \in \Omega$.
We define the operator $\mathcal{M}[i, y, z]$ as

$$(\mathcal{M}[i, y, z] f)(x_1, \ldots, x_n) := \max\big(f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n),$$
$$f(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n)\big) .$$

The operator $\mathcal{M}[i, y, z]$ is useful for two reasons. First, if $\mathrm{Inf}_i(f^{(j)})$ is "large", then $\mathrm{E}\big[\mathcal{M}[i, y, z] f^{(j)}(\underline{X}^{(j)})\big] \geq \mathrm{E}[f^{(j)}(\underline{X}^{(j)})] + c$ for some $y, z \in \Omega$ and $c > 0$. This implies that we can use this operator to increase the expectation of a function unless all of its influences are small. We will prove this property later.

Second, fix a step $j^* \in [\ell]$ and assume that for some values $\bar{x}^{\backslash j^*} = (x^{(1)}, \ldots, x^{(j^*-1)}, x^{(j^*+1)}, \ldots, x^{(\ell)}), y, z \in \Omega$ both conditional probabilities $\Pr[X_i^{(j^*)} = y \mid \overline{X}_i^{\backslash j^*} = \bar{x}^{\backslash j^*}]$ and $\Pr[X_i^{(j^*)} = z \mid \overline{X}_i^{\backslash j^*} = \bar{x}^{\backslash j^*}]$ are "somewhat large" (larger than some constant). We imagine now that $\overline{X}_i^{(\backslash j^*)} = \bar{x}^{\backslash j^*}$ and that we have also picked all values $\underline{X}_{\backslash i}^{(j^*)} = (X_1^{(j^*)}, \ldots, X_{i-1}^{(j^*)}, X_{i+1}^{(j^*)}, \ldots, X_n^{(j^*)})$. We then hope that $X_i^{(j^*)}$ is picked among y and z such that it maximizes $f^{(j^*)}$. Since this happens with constant probability, we conclude the following: Suppose we replace $f^{(j^*)}$ with $\mathcal{M}[i, y, z] f^{(j^*)}$ and then prove that afterwards $\mathrm{E}[\prod f^{(j)}(\underline{X}^{(j)})]$ is large. Then, $\mathrm{E}[\prod f^{(j)}(\underline{X}^{(j)})]$ was large before.

This second point is formalized in the following lemma:

Lemma 5.11. *Let $\overline{X}$ be a random vector distributed according to $(\overline{\Omega}, \mathcal{P})$. Fix $i \in [n]$, $j^* \in [\ell]$ and $\bar{x}^{\backslash j^*} = (x^{(1)}, \ldots, x^{(j^*-1)}, x^{(j^*+1)}, \ldots, x^{(\ell)}), y, z \in \Omega$. Suppose that:*

$$\mathcal{P}(\bar{x}^{\backslash j^*}, y) \geq \beta , \tag{22}$$

$$\mathcal{P}(\bar{x}^{\backslash j^*}, z) \geq \beta . \tag{23}$$

Let $f^{(1)}, \ldots, f^{(\ell)} : \underline{\Omega} \to [0, 1]$, and for $j \in [\ell]$ define:

$$g^{(j)} := \begin{cases} \mathcal{R}[i, x^{(j)}] f^{(j)} & \text{if } j \neq j^*, \\ \mathcal{M}[i, y, z] f^{(j)} & \text{if } j = j^*. \end{cases} \tag{24}$$

Then:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq \beta \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g^{(j)}(\underline{X}^{(j)})\right] . \tag{25}$$

Proof. We first define a random variable A, which is the value among y and z which $X_i^{(j^*)}$ needs to take in order to maximize $f^{(j^*)}$. Formally,

$$A = \begin{cases} y & \text{if } f\left(\underline{X}_{\backslash i}^{(j^*)}, y\right) > f\left(\underline{X}_{\backslash i}^{(j^*)}, z\right), \\ z & \text{otherwise.} \end{cases} \tag{26}$$

Consider now the event $\mathcal{E}$ which occurs if $\overline{X}_i = (\bar{x}^{\backslash j}, A)$. We get

$$\begin{aligned} \mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] &\geq \mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)}) \cdot \mathbb{1}(\mathcal{E})\right] \\ &= \mathrm{E}\left[\prod_{j=1}^{\ell} g^{(j)}(\underline{X}^{(j)}) \cdot \mathbb{1}(\mathcal{E})\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[\prod_{j=1}^{\ell} g^{(j)}(\underline{X}^{(j)}) \cdot \mathbb{1}(\mathcal{E}) \Big| \overline{X}_{\backslash i}\right]\right] \\ &= \mathrm{E}\left[\prod_{j=1}^{\ell} g^{(j)}(\underline{X}^{(j)}) \cdot \mathrm{E}\left[\mathbb{1}(\mathcal{E}) \Big| \overline{X}_{\backslash i}\right]\right] \\ &\geq \beta \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g^{(j)}(\underline{X}^{(j)})\right] . \end{aligned}$$

The equality from the first to the second line follows because if the event $\mathcal{E}$ happens, then the functions $f^{(j)}(\underline{X}^{(j)})$ and $g^{(j)}(\underline{X}^{(j)})$ are equal. From the third to the fourth line we use that conditioned on $\overline{X}_{\backslash i}$ the functions $g^{(j)}(\underline{X}^{(j)})$ are constant. Finally, the last inequality follows because by (22) and (23), for every choice of $\overline{X}_{\backslash i} = (\overline{X}_1, \ldots, \overline{X}_{i-1}, \overline{X}_{i+1}, \ldots, \overline{X}_n)$ event $\mathcal{E}$ has probability at least β. $\qquad\square$

The obvious idea for the next step would be to find values $\bar{x}^{\backslash j}, y, z$ such that

$$\mathrm{E}\left[\mathcal{M}[i, y, z] f^{(j^*)}(\underline{X}^{(j^*)})\right] \geq \mathrm{E}\left[f^{(j^*)}(\underline{X}^{(j^*)})\right] + c$$

and fix them.

Unfortunately, there is a problem with this strategy. To replace the function $f^{(j^*)}$ with $\mathcal{M}[i, y, z] f^{(j^*)}$, Lemma 5.11 also replaces $f^{(j)}$ with $\mathcal{R}[i, x^{(j)}] f^{(j)}$ for $j \neq j^*$ (and this is required for the proof to work). Unfortunately, it is possible that $\mathrm{E}\left[\mathcal{R}[i, x^{(j)}] f^{(j)}(\underline{X}^{(j)})\right] \ll \mathrm{E}\left[f^{(j)}(\underline{X}^{(j)})\right]$. We remark that we *cannot* use that $f^{(j)}$ is resilient here: while $f^{(j)}$ is resilient the first time we condition, the functions $\mathcal{M}[i, y, z] f^{(j)}$ obtained in the subsequent steps are not resilient in general, so later steps will not have the guarantee.

Our solution is to pick the values $(\overline{X}^{\backslash j^*}, Y, Z)$ at random, as a double sample on coordinate j^* (cf. Definition 5.1). Let:

$$G^{(j)} := \begin{cases} \mathcal{R}[i, X^{(j)}] f^{(j)} & \text{if } j \neq j^*, \\ \mathcal{M}[i, Y, Z] f^{(j)} & \text{if } j = j^*. \end{cases}$$

Note that the random variable $X^{(j)}$ is part of the double sample $(\overline{X}^{\setminus j^*}, Y, Z)$ and sampled separately (and independently) from the random vector $\overline{X}$. In particular, it should not be confused with the "input" random variable $X_i^{(\setminus j)}$. We prove that (in expectation over $\overline{X}^{(\setminus j)}, Y, Z$) the sum of expectations $\sum_{j=1}^{\ell} \mathrm{E}[G^{(j)}(\underline{X}^{(j)})]$ is greater by a constant than the sum $\sum_{j=1}^{\ell} \mathrm{E}[f^{(j)}(\underline{X}^{(j)})]$. To argue that the sum of expectations increases, the key part is to show that $\mathrm{E}\left[G^{(j^*)}(\underline{X}^{(j^*)})\right]$ increases by a constant.

Lemma 5.12. *Let $(\overline{X}^{\setminus j^*}, Y, Z)$ be a double sample from a single-coordinate distribution $\mathcal{P}$.*

Let $\underline{X}$ be a random vector, independent of this double sample and distributed according to a single-step distribution $(\underline{\Omega}, \underline{\pi})$ such that π is the j^*-th marginal distribution of $\mathcal{P}$.

Then, for every $i \in [n]$ and every function $f : \underline{\Omega} \to [0,1]$ we have

$$\mathrm{E}\left[\mathcal{M}[i, Y, Z]f(\underline{X})\right] \geq \mathrm{E}[f(\underline{X})] + \tau(1 - \rho^2(\mathcal{P})) , \tag{27}$$

where $\tau = \mathrm{Inf}_i(f(\underline{X}))$.

Recall that the distribution of (Y, Z) depends on j^*. We do not need to consider the full multi-step process in this lemma, but when applying it later we will set $\underline{X} = \underline{X}^{(j^*)}$ and $f = f^{(j^*)}$.

Proof. Fix a vector $\underline{x}_{\setminus i}$ for $\underline{X}_{\setminus i}$, and define the function $h : \Omega \to [0,1]$ as $h(x) := f(\underline{x}_{\setminus i}, x)$. By Corollary 5.4,

$$\mathrm{E}[|h(Y) - h(Z)|] \geq \mathrm{E}[(h(Y) - h(Z))^2] \geq 2(1 - \rho^2)\, \mathrm{Var}[h(Y)] ,$$

and hence, averaging over $\underline{X}_{\setminus i}$,

$$\mathrm{E}\left[\left|f(\underline{X}_{\setminus i}, Y) - f(\underline{X}_{\setminus i}, Z)\right|\right] \geq 2(1 - \rho^2)\, \mathrm{Inf}_i(f(\underline{X}_{\setminus i}, Y)) = 2\tau(1 - \rho^2) . \tag{28}$$

Since Y and Z are symmetric (i.e., they define a reversible Markov chain, cf. remarks after Definition 5.1) and by (28),

$$\mathrm{E}\left[\left(\mathcal{M}[i, Y, Z]f - f\right)(\underline{X})\right] = \mathrm{E}\left[\max(f(\underline{X}_{\setminus i}, Y), f(\underline{X}_{\setminus i}, Z)) - f(\underline{X}_{\setminus i}, Y)\right]$$
$$= \frac{1}{2}\mathrm{E}\left[\left|f(\underline{X}_{\setminus i}, Y) - f(\underline{X}_{\setminus i}, Z)\right|\right] \geq \tau(1 - \rho^2) ,$$

as claimed. $\qquad\square$

Lemma 5.13. *Let a random vector $\overline{X}$ be distributed according to $(\overline{\Omega}, \mathcal{P})$ and functions $f^{(1)}, \ldots, f^{(\ell)} : \underline{\Omega} \to [0,1]$. Let i, j^* and τ be such that $\mathrm{Inf}_i(f^{(j^*)}) \geq \tau \geq 0$ and let $\rho(\mathcal{P}) \leq \rho \leq 1$.*

Pick a double sample $(\overline{X}^{\setminus j^*}, Y, Z)$ from $\mathcal{P}$ and let:

$$G^{(j)} := \begin{cases} \mathcal{R}[i, X^{(j)}]f^{(j)} & \text{if } j \neq j^* \\ \mathcal{M}[i, Y, Z]f^{(j^*)} & \text{if } j = j^*. \end{cases} \tag{29}$$

Then:

$$\mathrm{E}\left[\sum_{j=1}^{\ell} \mathrm{E}[G^{(j)}(\underline{X}^{(j)}) \mid G^{(j)}]\right] \geq \sum_{j=1}^{\ell} \mathrm{E}\left[f^{(j)}(\underline{X}^{(j)})\right] + \tau \cdot (1 - \rho^2) . \tag{30}$$

Note that (29) defines the functions $G^{(j)}$ as random variables which is why we use capital letters.

Proof. If $j \neq j^*$ we have

$$\mathrm{E}\left[\mathrm{E}[G^{(j)}(\underline{X}^{(j)}) \mid G^{(j)}]\right] = \mathrm{E}[f^{(j)}(\underline{X}^{(j)})] \,, \tag{31}$$

since the marginal distribution of $\underline{X}^{(j)}$ is exactly as in the marginal π of $\mathcal{P}$. Hence, it suffices to show that

$$\mathrm{E}\left[\mathrm{E}[G^{(j^*)}(\underline{X}^{(j^*)}) \mid G^{(j^*)}]\right] = \mathrm{E}\left[\mathcal{M}[i,Y,Z]f^{(j^*)}(\underline{X}^{(j^*)})\right]$$
$$\geq \mathrm{E}[f^{(j^*)}(\underline{X}^{(j^*)})] + \tau(1-\rho^2) \,,$$

but this is exactly Lemma 5.12. □

Lemma 5.14. *Let $\overline{X}$ be a random vector distributed according to $(\overline{\underline{\Omega}}, \underline{\mathcal{P}})$ and also let $f^{(1)},\ldots,f^{(\ell)} : \underline{\Omega} \to [0,1]$, $i \in [n]$, $j^* \in [\ell]$, $\mathrm{Inf}_i(f^{(j^*)}) \geq \tau \geq 0$, $\rho(\mathcal{P}) \leq \rho \leq 1$.*
Then, there exist values $\overline{x}^{\setminus j^*} = (x^{(1)},\ldots,x^{(j^*-1)},x^{(j^*+1)},\ldots,x^{(\ell)}),y,z$ such that the functions

$$g^{(j)} := \begin{cases} \mathcal{R}[i,x^{(j)}]f^{(j)} & \text{if } j \neq j^* \\ \mathcal{M}[i,y,z]f^{(j)} & \text{if } j = j^* \end{cases} \tag{32}$$

satisfy

$$\sum_{j=1}^{\ell} \mathrm{E}[g^{(j)}(\underline{X}^{(j)})] \geq \sum_{j=1}^{\ell} \mathrm{E}[f^{(j)}(\underline{X}^{(j)})] + \tau(1-\rho^2)/2 \,, \tag{33}$$

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq \frac{\tau(1-\rho^2)}{2\ell|\Omega|^{\ell+1}} \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g^{(j)}(\underline{X}^{(j)})\right] \,. \tag{34}$$

While (33) is immediate from Lemma 5.13, we have to do a little bit of work to guarantee that it holds simultaneously with (34).

Proof. Choose $(\overline{X}^{\setminus j^*}, Y, Z)$ as a double sample from $\mathcal{P}$ and let $G^{(j)}$ be defined as in (29).

Define $p(\overline{x}^{\setminus j^*}, y, z) := \Pr[\overline{X}^{\setminus j^*} = \overline{x}^{\setminus j^*} \wedge Y = y \wedge Z = z]$, $\beta := \frac{\tau(1-\rho^2)}{2\ell|\Omega|^{\ell+1}}$, an event $\mathcal{E} :\equiv p(\overline{X}^{\setminus j^*}, Y, Z) < \beta$ and a random variable

$$A := \sum_{j=1}^{\ell} \mathrm{E}[G^{(j)}(\underline{X}^{(j)}) \mid G^{(j)}] - \mathrm{E}[f^{(j)}(\underline{X}^{(j)})] \,.$$

By Lemma 5.13, we have $\mathrm{E}[A] \geq \tau(1-\rho^2)$.

Since there are $|\Omega|^{\ell+1}$ possible tuples $(\overline{x}^{\setminus j}, y, z)$, by union bound we have $\Pr[\mathcal{E}] \leq |\Omega|^{\ell+1}\beta = \tau(1-\rho^2)/2\ell$. Bearing in mind the above and that $A \in [-\ell, \ell]$,

$$\mathrm{E}[A \cdot \mathbb{1}(\neg\mathcal{E})] \geq \mathrm{E}[A] - \ell\Pr[\mathcal{E}] \geq \tau(1-\rho^2)/2 \,.$$

As a consequence, we can choose $(\overline{x}^{\setminus j^*}, y, z)$ such that $A \geq \tau(1-\rho^2)/2$ and $\mathcal{E}$ does not happen. (33) is now immediate, while for (34) observe that $\neg\mathcal{E}$ implies $\mathcal{P}(\overline{X}^{\setminus j^*}, Y) \geq \beta$ and $\mathcal{P}(\overline{X}^{\setminus j^*}, Z) \geq \beta$ and apply Lemma 5.11. □

We can now repeat the process from Lemma 5.14 multiple times to get the result of this section.

Corollary 5.15. *Let* $\overline{X}$ *be a random vector distributed according to* $(\overline{\Omega}, \mathcal{P})$ *with* $\rho(\mathcal{P}) \leq \rho < 1$. *Then, for every* $\tau > 0$ *there exist* $k \in \mathbb{N}$ *and* $\beta > 0$ *such that:*

For every $\varepsilon \in [0,1]$ *and functions* $f^{(1)}, \ldots, f^{(\ell)} : \underline{\Omega} \to [0,1]$ *such that each* $f^{(j)}$ *is* ε-*resilient up to size* k, *there exist* $g^{(1)}, \ldots, g^{(\ell)} : \underline{\Omega} \to [0,1]$ *with the following properties:*

1. $\max_{j \in [\ell]} \max_{i \in [n]} \mathrm{Inf}_i(g^{(j)}(\underline{X}^{(j)})) \leq \tau$.

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

3. *For all* $j \in [\ell]$*:* $\mathrm{E}[g^{(j)}(\underline{X}^{(j)})] \geq (1-\varepsilon)\,\mathrm{E}[f^{(j)}(\underline{X}^{(j)})]$.

Furthermore, one can take $k := \lfloor \frac{2\ell}{\tau(1-\rho^2)} \rfloor$ *and* $\beta := \left(\frac{\tau(1-\rho^2)}{2\ell|\Omega|^{\ell+1}} \right)^k$.

In particular, both k and β depend only on τ and $\mathcal{P}$ (requiring $\tau > 0$ and $\rho(\mathcal{P}) < 1$).

Proof. We repeat the process from Lemma 5.14, always replacing the collection of functions $f^{(1)}, \ldots, f^{(\ell)}$ with $g^{(1)}, \ldots, g^{(\ell)}$ until condition 1 is satisfied. Since $\sum_{j=1}^{\ell} \mathrm{E}[f^{(j)}(\underline{X}^{(j)})]$ cannot exceed ℓ and every time it increases by $\tau(1-\rho^2)/2$, we have to do this at most $\frac{2\ell}{\tau(1-\rho^2)}$ times.

The first point is then obvious, and the second point follows from Lemma 5.14.

Finally, the third point follows because the functions $f^{(j)}$ are all ε-resilient up to size k, and each of the functions $g^{(j)}$ can be written as a maximum of restrictions of size at most k of $f^{(j)}$. Since the maximum only increases expectations, the proof follows. $\square$

5.4   Finishing the proof

Proof of Theorem 3.2. Let us assume that $\mu \in (0, 0.99]$, the computations being only easier if this is not the case. To establish (5), whenever we say "constant", in the $O()$ notation or otherwise, we mean "depending only on $\mathcal{P}$ (in particular, on α, ρ, $|\Omega|$ and ℓ), but not on μ".

The proof consecutively applies Lemma 5.9, Corollary 5.15 and Theorem 4.1.

Given $f : \underline{\Omega} \to [0,1]$ with $\mathrm{E}[f(\underline{X}^{(1)})] = \mu$, first apply Lemma 5.9 to f with $\varepsilon := 1/2$ and $k := \exp\left((1/\mu)^D \right)$ for a constant D large enough (where "large enough" will depend on another constant D' to be defined later). This gives us a function $g : \underline{\Omega} \to [0,1]$ such that:

* g is ε-resilient up to size k.

* $\mathrm{E}[g(\underline{X}^{(1)})] \geq \mu$.

*

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g(\underline{X}^{(j)})\right], \tag{35}$$

where:

$$c = 1/\exp\left((1/\alpha)^{2k} \cdot 4\ln 1/\mu\right) \geq 1/\exp\left(\exp\left(O(k)\right) \cdot 4\ln 1/\mu\right)$$

$$\geq 1/\exp\left(\exp\left(\exp\left((1/\mu)^{O(1)}\right)\right) \cdot 4\ln 1/\mu\right)$$

$$\geq 1/\exp\left(\exp\left(\exp\left((1/\mu)^{O(1)}\right)\right)\right).$$

Next, apply Corollary 5.15. Set $g^{(1)} := \ldots := g^{(\ell)} := g$ and $\tau := 1/\exp\left((1/\mu)^{D'}\right)$ for a constant D' large enough. We need to check if k we have chosen satisfies the assumption of Corollary 5.15:

$$\frac{2\ell}{\tau(1-\rho^2)} \leq O\left(\exp\left((1/\mu)^{D'}\right)\right) \leq \exp\left((1/\mu)^{O(1)}\right) \leq k.$$

Therefore, Corollary 5.15 is applicable and yields $h^{(1)}, \ldots, h^{(\ell)} : \underline{\Omega} \to [0,1]$ such that:

- $\max_{j\in[\ell]} \max_{i\in[n]} \mathrm{Inf}_i(h^{(j)}(\underline{X}^{(j)})) \leq \tau.$

- $\forall j \in [\ell] : \mathrm{E}[h^{(j)}(\underline{X}^{(j)})] \geq \mu/2.$

-

$$\mathrm{E}\left[\prod_{j=1}^{\ell} g(\underline{X}^{(j)})\right] \geq \beta \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} h^{(j)}(\underline{X}^{(j)})\right], \tag{36}$$

where:

$$\beta = \left(\frac{\tau(1-\rho^2)}{2\ell|\Omega|^{\ell+1}}\right)^k \geq 1/O\left(\exp\left((1/\mu)^{D'}\right)\right)^k$$

$$\geq 1/\exp\left((1/\mu)^{O(1)} \cdot k\right) \geq 1/\exp\left(\exp\left((1/\mu)^{O(1)}\right)\right).$$

Finally, we need to apply Theorem 4.1. To this end, set:

$$\varepsilon := (\mu/2)^{\ell^2/(1-\rho^2)}/2 \geq \mu^{O(1)}$$

and verify (12):

$$\left(\frac{(1-\rho^2)\varepsilon}{\ell^{5/2}}\right)^{O\left(\frac{\ln(\ell/\varepsilon)\ln(1/\alpha)}{(1-\rho)\varepsilon}\right)} \geq \Omega(\varepsilon)^{(O(1)+\ln 1/\varepsilon)\cdot O(1/\varepsilon)}$$

$$\geq 1/\exp\left((O(1)+\ln 1/\varepsilon)^2 \cdot O(1/\varepsilon)\right)$$

$$\geq 1/\exp\left((1/\varepsilon)^{O(1)}\right) \geq 1/\exp\left((1/\mu)^{O(1)}\right).$$

Hence, from Theorem 4.1:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} h^{(j)}(\underline{X}^{(j)})\right] \geq \varepsilon/2 \geq \mu^{O(1)}. \tag{37}$$

(35), (36) and (37) put together give:

$$
\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c \cdot \beta \cdot \mu^{O(1)} \geq 1/\exp\left(\exp\left(\exp\left(\left((1/\mu)^{O(1)}\right)\right)\right)\right),
$$

as claimed. □

6 Proof for Two Steps

Our goal in this section is to prove Theorem 3.1 assuming Theorem 3.2.

In the following we will sometimes drop the assumption that Ω is necessarily the support of a probability distribution $\mathcal{P}$. One can check that this will not cause problems.

6.1 Correlation of a cycle

Assume we are given a support set Ω of size $|\Omega| = k$. Let $s \geq 2, p \in (0,1)$ and let $(x_0, \ldots, x_{s-1})$ be a sequence of distinct $x_i \in \Omega$.

Definition 6.1. We call a probability distribution $\mathcal{C}$ over Ω an (s,p)-cycle if

$$
\mathcal{C}(x,y) = \begin{cases} p/s & \text{if } x = y = x_i \text{ for } i \in \{0, \ldots, s-1\}, \\ (1-p)/s & \text{if } x = x_i \wedge y = x_{(i+1) \bmod s} \text{ for } i \in \{0, \ldots, s-1\}, \\ 0 & \text{otherwise.} \end{cases}
$$

Lemma 6.2. *Let $\mathcal{C}$ be an (s,p)-cycle. Then*

$$
\rho(\mathcal{C}) \leq 1 - \frac{7p(1-p)}{s^2}.
$$

Proof. Let K be the Markov kernel induced by a double sample on $\mathcal{C}$ (K is the same whether a sample is on the first or the second step, cf. Section 5.1). Observe that

$$
K(y,z) := \begin{cases} p^2 + (1-p)^2 & \text{if } y = z = x_i, \\ p(1-p) & \text{if } y = x_i \text{ and } z = x_{(i\pm1) \bmod s}. \end{cases}
$$

Let $\alpha_k := \frac{2\pi k}{s}$. One can check that the eigenvalues of K are $\lambda_0, \ldots, \lambda_{s-1}$ with $\lambda_k := 1 - 2p(1-p)(1 - \cos\alpha_k)$. This is easiest if one knows the respective (complex) eigenvectors $v_k := (1, \exp(\alpha_k \iota), \ldots, \exp((s-1)\alpha_k \iota))$ (where ι is the imaginary unit).

Using $\cos x \leq 1 - x^2/5$ for $x \in [0, \pi]$ and $\sqrt{1-x} \leq 1 - x/2$ for $x \in [0,1]$ we obtain that if $k > 0$, then

$$
\sqrt{\lambda_k} \leq \sqrt{1 - 2p(1-p)(1 - \cos\alpha_1)} \leq \sqrt{1 - 2p(1-p)\frac{4\pi^2}{5s^2}} \leq 1 - \frac{7p(1-p)}{s^2}.
$$

The bound on $\rho(\mathcal{C})$ now follows from Lemma 5.3. □

6.2   Convex decomposition of $\mathcal{P}$

In this section we show that if a distribution $\mathcal{P}$ can be decomposed into a convex combination of distributions $\mathcal{P} = \sum_{k=1}^{r} \alpha_k \mathcal{P}_k$ and each distribution $\mathcal{P}_k$ is same-set hitting, then also $\mathcal{P}$ is same-set hitting.

Definition 6.3. We say that a probability distribution with equal marginals $\mathcal{P}$ has an (α, ρ)-convex decomposition if there exist $\beta_1, \ldots, \beta_r > 0$ with $\sum_{k=1}^{r} \beta_k = 1$ and distributions with equal marginals $\mathcal{P}_1, \ldots, \mathcal{P}_r$ such that

$$\mathcal{P} = \sum_{k=1}^{r} \beta_k \cdot \mathcal{P}_k \,.$$

and $\alpha(\mathcal{P}_k) \geq \alpha$ and $\rho(\mathcal{P}_k) \leq \rho$ for every $k \in [r]$.

Lemma 6.4. *Let an ℓ-step distribution $\mathcal{P}$ with equal marginals have an (α, ρ)-convex decomposition for some $\alpha > 0$ and $\rho < 1$.*

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

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

Proof. Let us write the relevant decomposition as $\mathcal{P} = \sum_{k=1}^{r} \beta_k \mathcal{P}_k$. The existence of this decomposition implies that there exists a random vector $\underline{Z} = (Z_1, \ldots, Z_n)$ such that:

- The variables $Z_i \in [r]$ are i.i.d. with $\Pr[Z_i = k] = \beta_k$.

- For every $i \in [n]$ and $k \in [r]$, conditioned on $Z_i = k$, the tuple $\overline{X}_i$ is distributed according to $\mathcal{P}_k$.

Let $\underline{z}$ be an arbitrary assignment to $\underline{Z}$ and let $\mu_{\underline{z}} := \mathrm{E}[f(\underline{X}^{(1)}) \mid \underline{Z} = \underline{z}]$. If $\mu_{\underline{z}} \geq \mu/2$, by Theorem 3.2[2]

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)}) \mid \underline{Z} = \underline{z}\right] \geq c(\alpha, \rho, \ell, \mu) > 0 \,. \tag{38}$$

Since $\mathrm{E}[\mu_{\underline{Z}}] = \mathrm{E}[\mathrm{E}[f(\underline{X}^{(1)}) \mid \underline{Z}]] = \mathrm{E}[f(\underline{X}^{(1)})] = \mu$, by Markov

$$\Pr[\mu_{\underline{Z}} \geq \mu/2] \geq \mu/2 \,. \tag{39}$$

(38) and (39) together give

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq \mu/2 \cdot c(\alpha, \rho, \ell, \mu) > 0 \,.$$

$\square$

---

[2]Technically, Theorem 3.2 requires the distributions to be the same for each coordinate, which is not the case in our setting. However, this is not a problem, cf. Section 3.4.1.

6.3 Decomposition of $\mathcal{P}$ into cycles

Definition 6.5. Let us consider weighted directed graphs with non-negative weights over a vertex set Ω. We will identify such a digraph G with its weight matrix.

We say that such a weighted digraph is *regular*, if for every vertex the total weight of the incoming edges is equal to the total weight of the outgoing edges.

We call a weighted digraph a *weighted cycle*, if it is a directed cycle over a subset of Ω with all edges of the same weight $w > 0$. We call w the *weight* of the cycle and number of its edges s the *size* of the cycle.

We say that a weighted digraph G can be *decomposed into r weighted cycles* if there exist weighted cycles $C_1, \ldots, C_r$ such that $G = \sum_{k=1}^r C_k$.

Lemma 6.6. *Every regular weighted digraph G over a set Ω of size k can be decomposed into at most k^2 weighted cycles.*

Proof. Since the digraph is regular, it must have a cycle. Remove it from the graph (taking as weight w the minimum weight of the edge on this cycle).

Since the resulting graph is still regular, proceed by induction until the graph is empty.

At each step at least one edge is completely removed from the graph, therefore there will be at most k^2 steps. $\qquad\square$

To see that a two-step distribution $\mathcal{P}$ can be decomposed into cycles, it will be useful to take $\mathcal{P}' := \mathcal{P} - \alpha \cdot \mathrm{Id}$ and look at it as a weighted directed graph $(\Omega, \mathcal{P}')$, where $\mathcal{P}'$ is interpreted as a weight function $\mathcal{P}' : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$.

Lemma 6.7. *Let $\mathcal{P}$ be a two-step distribution with equal marginals over an alphabet Ω with size t.*

Then, $\mathcal{P}$ has a convex decomposition $\mathcal{P} = \sum_{k=1}^r \beta_k \mathcal{P}_k$ such that each $\mathcal{P}_k$ either has support of size 1 or is an (s,p)-cycle with $2 \leq s \leq t$ and $p \in [\alpha(\mathcal{P})^3, 1/2]$.

Consequently, $\mathcal{P}$ has an (α, ρ)-convex decomposition with $\alpha := \alpha(\mathcal{P})^4$ and $\rho := 1 - 3\alpha(\mathcal{P})^5$.

Proof. Throughout this proof we will treat $\mathcal{P}$ as a weight matrix of a digraph. Since $\mathcal{P}$ has equal marginals, this weighted digraph is regular. Use Lemma 6.6 to decompose $\mathcal{P} - \alpha(\mathcal{P}) \cdot \mathrm{Id}$ into weighted cycles, which allows us to write

$$\mathcal{P} = \alpha(\mathcal{P}) \cdot \mathrm{Id} + \sum_{k=1}^r C_k \,,$$

where C_k is a weighted cycle with weight w_k and size s_k and $r \leq t^2$. Take $\beta_k := \min(w_k, \alpha(\mathcal{P})/t^2)$ and let Id_k be the identity matrix restricted to the support of C_k. Now we can write $\mathcal{P}$ as

$$\mathcal{P} = \left(\alpha(\mathcal{P}) \cdot \mathrm{Id} - \sum_{k=1}^r \beta_k \, \mathrm{Id}_k \right) + \left(\sum_{k=1}^r s_k(w_k + \beta_k) \cdot \frac{\beta_k \, \mathrm{Id}_k + C_k}{s_k(w_k + \beta_k)} \right) \,.$$

Firstly, $(\alpha(\mathcal{P}) \cdot \mathrm{Id} - \sum_{k=1}^r \beta_k \, \mathrm{Id}_k)$ can be decomposed into distributions with support size 1.

As for the other term, note that $\mathcal{C}_k := \frac{\beta_k \, \mathrm{Id}_k + C_k}{s_k(w_k + \beta_k)}$ is a probability distribution that either has support of size 1 (iff C_k has support of size 1) or is an (s,p)-cycle with $2 \leq s \leq t$ and $p = \beta_k/(\beta_k + w_k)$.

If $\beta_k = w_k$, then $p = 1/2$. If $\beta_k < w_k$, then $1/2 \geq p = \beta_k/(\beta_k + w_k) \geq \beta_k = \alpha(\mathcal{P})/t^2 \geq \alpha(\mathcal{P})^3$. Therefore, $p \in [\alpha(\mathcal{P})^3, 1/2]$, as stated.

Consequently, $\alpha(\mathcal{C}_k) = p/s_k \geq \alpha(\mathcal{P})^4$ and, by Lemma 6.2, $\rho(\mathcal{C}_k) \geq 1 - 3\alpha(\mathcal{P})^5$ and, since every (s, p)-cycle has equal marginals, we obtained an (α, ρ)-convex decomposition of $\mathcal{P}$. □

6.4 Putting things together

Proof of Theorem 3.1. From Lemmas 6.7 and 6.4. □

Remark 6.8. One can see that see that, as in Theorem 3.2, we obtain a triply exponential explicit bound, i.e, there exists $D(\alpha(\mathcal{P})) > 0$ such that if $\mu \in (0, 0.99]$, then

$$\mathrm{E}[f(\underline{X})f(\underline{Y})] \geq 1/\exp\left(\exp\left(\exp\left((1/\mu)^D\right)\right)\right) .$$

7 Local Variance

In this section we state and prove a generalization of the low-influence theorem from [Mos10]. 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 [O'D14].

[Mos10] shows that $\rho(\mathcal{P}) < 1$ implies that $\mathcal{P}$ is set hitting for low-influence functions. We extend this result to a weaker notion of influence. In particular, we show that $\mathcal{P}$ is set hitting for functions with $\Omega(1)$ measure and $o(1)$ largest Fourier coefficient. The main result of this section is Theorem 3.3.

We remark that Theorem 3.3 does *not* require equal marginals. The rest of this section contains the proof of Theorem 3.3. First, from Corollary 5.15 and Theorem 4.1 it is easy to establish[3] the following:

Theorem 7.1. *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 $\varepsilon \in [0, 1)$.*

Then, for all $\mu^{(1)}, \ldots, \mu^{(\ell)} \in (0, 1]$ there exists $k(\mathcal{P}, \varepsilon, \mu^{(1)}, \ldots, \mu^{(\ell)}) \in \mathbb{N}$ 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 if $f^{(1)}, \ldots, f^{(\ell)}$ are all ε-resilient up to size k, then

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq c(\mathcal{P}, \varepsilon, \mu^{(1)}, \ldots, \mu^{(\ell)}) > 0 . \tag{40}$$

Definition 7.2. Let π be a single-step distribution and let $f : \underline{\Omega} \to \mathbb{R}$ be a function. Let $S \subseteq [n]$ with $|S| = k$. We define $f^{\subseteq S} : \underline{\Omega} \to \mathbb{R}$ as

$$f^{\subseteq S}(\underline{x}) := \mathrm{E}[f(\underline{x}_S, \underline{X}_{\overline{S}})] , \tag{41}$$

where $\overline{S} := [n] \setminus S$, $\underline{x}_S$ is the vector $\underline{x}$ restricted to coordinates in S, and $\underline{X}_{\overline{S}}$ is a random vector of $n - k$ elements with each coordinate distributed i.i.d. in π.

A proof of the following claim can be found, e.g., in [O'D14]:

---

[3]One needs to check that the assumption about equal marginals is not necessary, but that turns out to be the case (the bound in Theorem 4.1 then depends on $\min_{j \in [\ell], x \in \mathrm{supp}(X^{(j)})} \pi^{(j)}(x)$).

Claim 7.3. *Let π be a single-step distribution and let $f : \underline{\Omega} \to \mathbb{R}$, $S \subseteq [n]$. If a random vector $\underline{X}$ is distributed according to $\underline{\pi}$ and $\phi_0, \ldots, \phi_{m-1}$ form a Fourier basis for π and $f = \sum_{\sigma \in \mathbb{N}_{<m}^n} \hat{f}(\sigma) \phi_\sigma$, then $f^{\subseteq S} = \sum_{\sigma : \mathrm{supp}(\sigma) \subseteq S} \hat{f}(\sigma) \phi_\sigma$. In particular,*

$$\mathrm{Var}\left[f^{\subseteq S}(\underline{X}) \right] = \sum_{\substack{\sigma : \mathrm{supp}(\sigma) \subseteq S, \\ \sigma \neq 0^n}} \left| \hat{f}(\sigma) \right|^2 .$$

Lemma 7.4. *Let a random vector $\underline{X}$ be distributed according to a single-step distribution $\underline{\pi}$ with $\min_{x \in \Omega} \pi(x) \geq \alpha$ and let $\varepsilon \in [0,1]$, $k \in \mathbb{N}$.*

Then, for every $f : \underline{\Omega} \to \mathbb{R}_{\geq 0}$ with $\mathrm{E}[f(\underline{X})] = \mu$, if for every $S \subseteq [n]$ with $|S| = k$ it holds that

$$\mathrm{Var}\left[f^{\subseteq S}(\underline{X}) \right] \leq \alpha^k (\varepsilon \mu)^2 ,$$

then f is ε-resilient up to size k.

Proof. We prove the contraposition.

If f is not ε-resilient up to size k, by definition of $f^{\subseteq S}$ it implies that there exist $S \subseteq [n]$ with $|S| = k$ and $\underline{x}$ such that

$$\left| f^{\subseteq S}(\underline{x}) - \mathrm{E}[f^{\subseteq S}(\underline{X})] \right| > \varepsilon \, \mathrm{E}[f^{\subseteq S}(\underline{X})] = \varepsilon \mu .$$

But this gives

$$\mathrm{Var}\left[f^{\subseteq S}(\underline{X}) \right] \geq \alpha^k \left(f^{\subseteq S}(\underline{x}) - \mathrm{E}[f^{\subseteq S}(\underline{X})] \right)^2 > \alpha^k (\varepsilon \mu)^2 ,$$

as required. $\qquad \square$

Using Lemma 7.4 we can weaken the assumption in Theorem 7.1 such that it only requires that all Fourier coefficients of degree at most k are small:

Proof of Theorem 3.3. From Theorem 7.1, there exists $k := k(\mathcal{P}, \mu^{(1)}, \ldots, \mu^{(\ell)})$ such that if $f^{(1)}, \ldots, f^{(\ell)}$ are all $1/2$-resilient up to size k, then (6) holds. Therefore, it is sufficient to show that the functions $f^{(j)}$ are indeed $1/2$-resilient up to size k if the parameter γ is chosen small enough.

By Claim 7.3, if $\max_{\sigma : 0 < |\sigma| \leq k} |\hat{f}^{(j)}(\sigma)| \leq \gamma$, then for any $S \subseteq [n]$ with $|S| = k$ we have $\mathrm{Var}\left[(f^{(j)})^{\subseteq S}(\underline{X}^{(j)}) \right] \leq |\Omega|^k \gamma^2$. With that in mind it is easy to choose γ such that Lemma 7.4 can be applied to each $f^{(j)}$. $\qquad \square$

8 Multiple Steps of a Markov Chain

Next, we consider the case where the distribution $\mathcal{P}$ is such that the random variables $X^{(1)}, X^{(2)}, \ldots, X^{(\ell)}$ form a Markov chain.

Definition 8.1. *Let $\mathcal{P}$ be a an ℓ-step distribution with equal marginals and let $\overline{X} = (X^{(1)}, \ldots, X^{(\ell)})$ be a random variable distributed according to $\mathcal{P}$. We say that $\mathcal{P}$ is generated by Markov chains[4] if for every $j \in \{2, \ldots, \ell\}$ and $x^{(1)}, \ldots, x^{(j)} \in \Omega$ we have*

$$\Pr[X^{(j)} = x^{(j)} | X^{(1)} = x^{(1)} \wedge \cdots \wedge X^{(j-1)} = x^{(j-1)}]$$
$$= \Pr[X^{(j)} = x^{(j)} | X^{(j-1)} = x^{(j-1)}] .$$

---

[4]Note that our definition allows for different Markov chains in different steps.

Observe that since we still require $\mathcal{P}$ to have equal marginals, the marginal π is then simply a stationary distribution of the chain.

In this case, we give a reduction to Theorem 3.1 to prove a bound that does not depend on $\rho(\mathcal{P})$:

Theorem 8.2. *Let Ω be a finite set and $\mathcal{P}$ a probability distribution over Ω^ℓ with equal marginals generated by Markov chains. 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 $f : \Omega^n \to [0,1]$ with $E[f(\underline{X}^{(1)})] = \mu > 0$:

$$E\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)}) \right] \geq c(\alpha(\mathcal{P}), \ell, \mu) , \tag{42}$$

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

Proof. Let $\mathcal{P}$ be a distribution generated by Markov chains with $\alpha := \alpha(\mathcal{P}) > 0$ and let $f : \underline{\Omega} \to [0,1]$ with $E[f(\underline{X}^{(1)})] = \mu > 0$.

The proof is by induction on ℓ. For $\ell = 2$, apply Theorem 3.1 directly. For $\ell > 2$, define the function $g : \underline{\Omega} \to [0,1]$ as

$$g(\underline{x}) := E\left[f(\underline{X}^{(\ell-1)}) f(\underline{X}^{(\ell)}) \mid \underline{X}^{(\ell-1)} = \underline{x} \right] = f(\underline{x}) \cdot E\left[f(\underline{X}^{(\ell)}) \mid \underline{X}^{(\ell-1)} = \underline{x} \right] .$$

Applying Theorem 3.1 for the distribution of the last two steps,

$$E[g(\underline{X}^{(1)})] = E[g(\underline{X}^{(\ell-1)})] = E[f(\underline{X}^{(\ell-1)}) f(\underline{X}^{(\ell)})] \geq c(\alpha, \mu) > 0 . \tag{43}$$

Now we have

$$E\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)}) \right] = E\left[\left(\prod_{j=1}^{\ell-2} f(\underline{X}^{(j)}) \right) g(\underline{X}^{(\ell-1)}) \right] \tag{44}$$

$$\geq E\left[\prod_{j=1}^{\ell-1} g(\underline{X}^{(j)}) \right] \tag{45}$$

$$\geq c(\alpha, \ell-1, c(\alpha, \mu)) = c(\alpha, \ell, \mu) > 0, \tag{46}$$

where (44) holds since $\mathcal{P}$ is generated by Markov chains, (45) is due to $f \geq g$ pointwise and (46) is an application of the induction and (43). $\qquad \square$

Remark 8.3. Unfortunately, this proof worsens the explicit bound. One can check that for a Markov-generated distribution with ℓ steps the dependence on μ is a tower of exponentials of height $3(\ell - 1)$.

9 Polynomial Same-Set Hitting

The property of set hitting establishes a lower bound on $E\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)}) \right]$ that is independent of n. However, it might be the case that this bound is very small, perhaps far from the best possible one. In particular, our bound from Theorem 3.2 is triply exponentially small, and the bound from Theorem 1.2 is not even primitive recursive.

Definition 9.1. A distribution $\mathcal{P}$ is *polynomially set hitting* (resp. polynomially same-set hitting) if there exists $C \geq 0$ such that $\mathcal{P}$ is (μ, μ^C)-set hitting (resp. same-set hitting) for every $\mu \in (0,1]$.

As a matter of fact, [MOS13] (cf. Theorem 1.4) establishes that all distributions that are set hitting are also polynomially set hitting. We suspect that this is also the case for two-step same-set hitting, but this remains an open problem.

However, it is possible to harness reverse hypercontractivity to show that all *symmetric* two-step distributions are polynomially same-set hitting:

Theorem 9.2. *Let a two-step probability distribution with equal marginals $\mathcal{P}$ be symmetric, i.e., $\mathcal{P}(x,y) = \mathcal{P}(y,x)$ for all $x, y \in \Omega$. If $\alpha(\mathcal{P}) > 0$, then $\mathcal{P}$ is polynomially same-set hitting.*

We omit the proof of Theorem 9.2, noting that the idea is similar as in Section 6: one performs an obvious convex decomposition of $\mathcal{P}$ into cycles of length two and applies the result of [MOS13] to each term of this decomposition.

A   Appendix: Proof of Theorem 4.1

Our proof of Theorem 4.1 follows in this appendix. It is only a slight adaptation of the argument from [Mos10], but we include it in full for the sake of completeness.

We first restate the theorem and discuss the differences between our proof and the one in [Mos10]:

Theorem 4.1. *Let $\overline{X}$ be a random vector distributed according to $(\overline{\Omega}, \underline{\mathcal{P}})$ 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 $\varepsilon > 0$, there exists $\tau := \tau(\varepsilon, \rho, \alpha, \ell) > 0$ 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 $\mu^{(j)} := \mathrm{E}[f^{(j)}(\underline{X}^{(j)})]$:

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

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

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

Theorem 4.1 is very similar to a subcase of Theorem 1.14 from [Mos10]. We make a stronger claim with one respect: in [Mos10] the influence threshold τ depends among others on:

$$\alpha^* := \min_{(x^{(1)}, \ldots, x^{(\ell)}) \in \mathrm{supp}(\mathcal{P})} \mathcal{P}(x^{(1)}, \ldots, x^{(\ell)}) \,, \tag{47}$$

while our bound depends only on the smallest marginal probability:

$$\alpha = \min_{x \in \Omega} \pi(x) \, . \tag{48}$$

The main differences to the proof in [Mos10] are:

- [Mos10] proves the base case $\ell = 2$ and then obtains the result for general ℓ by an inductive argument (cf., Theorem 6.3 and Proposition 6.4 in [Mos10]). Since the induction is applied to functions $f^{(1)}$ and $g := \prod_{j=2}^{\ell} f^{(j)}$, where g is viewed as a function on a single-step space, the information on the smallest marginal is lost in the case of g. To avoid this, our proof proceeds directly for general ℓ. However, the structure and the main ideas are really the same as in [Mos10].

- In Section A.4, in hypercontractivity bounds for Gaussian and discrete spaces (Theorem A.42 and Lemma A.43) we are slightly more careful to obtain bounds which depend on α rather than α^* (as defined in (48) and (47)). This better bound is then propagated in the proof of the invariance principle.

- Another change is not related to the dependency on the smallest marginal. In Section A.8, in the Gaussian reverse hypercontractivity bound (Theorem A.76) instead of using the result of Borell ([Bor85], Theorem 5.1 in [Mos10]) for a bound expressed in terms of the cdf of bivariate Gaussians, we utilize the results of [CDP15] and [Led14] for a more convenient bound of the form $\left(\prod_{j=1}^{\ell} \mu^{(j)} \right)^{c(\rho, \ell)}$.

The proof can be generalized in several directions, but for the sake of clarity we present the simplest version sufficient for our purposes.

A.1 Preliminaries — the general framework

We start with explaining the notation of random variables and L^2 spaces that we will use throughout the proof.

Definition A.1. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. We define the real inner product space $L^2(\Omega, \mathcal{P})$ as the set of all square-integrable functions $f : \Omega \to \mathbb{R}$, i.e., the functions that satisfy

$$\int_{\Omega} f^2 \, \mathrm{d}\mathcal{P} < +\infty \, , \tag{49}$$

with inner product defined as

$$\langle f, g \rangle := \int_{\Omega} fg \, \mathrm{d}\mathcal{P} \, . \tag{50}$$

Remark A.2. As we will see shortly, if X is a random variable sampled from Ω according to $\mathcal{P}$, the equations (49) and (50) can be written as

$$\mathrm{E}[f^2(X)] < +\infty \, ,$$
$$\langle f, g \rangle = \mathrm{E}[f(X)g(X)] \, .$$

Remark A.3. We omitted the event space $\mathcal{F}$ in the definition of $L^2(\Omega, \mathcal{P})$. This is because $\mathcal{F}$ is always implicit in the choice of the measure $\mathcal{P}$.

In particular, when $\mathcal{P}$ is discrete, of course we choose $\mathcal{F}$ to be the powerset of Ω. When $\mathcal{P}$ is continuous over $\mathbb{R}^n$, we use the "standard" real event space, i.e., the completion of the Borel algebra.

While this will not be our usual way of thinking, at this point it makes sense to introduce the formal definition of a random variable: a function from a probability space to some set.

Definition A.4. Let $(\Sigma, \mathcal{F}, \mathcal{P})$ be a probability space. We say that X is a random variable over a set Σ' if it is a measurable function $X : \Sigma \to \Sigma'$.

As usual, we will assume throughout the proof that all random variables are induced by some underlying probability space $(\Sigma, \mathcal{F}, \mathcal{P})$.

Using this, a random variable induces some distribution, which we can study.

Definition A.5. We say that a random variable X over a set Ω is *distributed according to a probability space* $(\Omega, \mathcal{P})$ if for every event $A \in \mathcal{F}$:

$$\Pr[X \in A] = \mathcal{P}(A) .$$

Definition A.6. Let X be a random variable distributed over Ω. By $L^2(X)$ we denote the inner product space of random variables that correspond to square-integrable functions $f : \Omega \to \mathbb{R}$:

$$L^2(X) := \{Z \mid Z = f \circ X \text{ for some } f : \Omega \to \mathbb{R} \text{ with } \mathrm{E}[f(X)^2] < +\infty\} ,$$

with the inner product given as

$$\langle Z_1, Z_2 \rangle := \mathrm{E}[Z_1 \cdot Z_2] .$$

Remark A.7. We consider the formal setting again, i.e., suppose $(\Sigma, \mathcal{F}, \mathcal{P})$ is the underlying probability space, and $X : \Sigma \to \Omega$ a random variable. Then, $L^2(X)$ is a subspace of $L^2(\Sigma, \mathcal{P})$. Intuitively, it contains all real valued functions which "depend only on X".

Example A.8. Fix $(\Omega, \mathcal{P})$ to be the uniform distribution on $\Omega := \{0, 1, 2\}$ and let X be distributed according to $(\Omega, \mathcal{P})$. Then $L^2(X)$ has dimension three and one of its orthonormal bases is

$$Z_0 :\equiv 1$$
$$Z_1 := \begin{cases} \sqrt{6}/2 & \text{if } X = 0, \\ -\sqrt{6}/2 & \text{if } X = 1, \\ 0 & \text{if } X = 2. \end{cases}$$
$$Z_2 := \begin{cases} \sqrt{2}/2 & \text{if } X \in \{0, 1\}, \\ -\sqrt{2} & \text{if } X = 2. \end{cases}$$

After this point, we will have no need to refer explicitly to the underlying probability space $(\Sigma, \mathcal{F}, \mathcal{P})$ anymore. Nevertheless, it will be useful to remember that random variables are functions of this underlying space.

It immediately follows from the definitions that:

Lemma A.9. *Let X be a random variable distributed according to $(\Omega, \mathcal{P})$. Then $L^2(X)$ is isomorphic to $L^2(\Omega, \mathcal{P})$.*

A.2 Preliminaries — orthonormal ensembles and multilinear polynomials

In this section we introduce orthonormal ensembles and multilinear polynomials over them.

Definition A.10. We call a finite family $(\mathfrak{X}_0, \ldots, \mathfrak{X}_p)$ of random variables *orthonormal* if they satisfy $\mathrm{E}[\mathfrak{X}_k^2] = 1$ for every k and $\mathrm{E}[\mathfrak{X}_j \mathfrak{X}_k] = 0$ for every $j \neq k$.

Definition A.11. We call a finite family of orthonormal random variables $\mathfrak{X} = (\mathfrak{X}_{\star,0} = 1, \mathfrak{X}_{\star,1}, \ldots, \mathfrak{X}_{\star,p})$ an *orthonormal ensemble*. We call p the *size* of the ensemble.

An *ensemble sequence* is a sequence of independent families of random variables $\underline{\mathfrak{X}} = (\mathfrak{X}_1, \ldots, \mathfrak{X}_n)$ such that each $\mathfrak{X}_i$ is an orthonormal ensemble $\mathfrak{X}_i = (\mathfrak{X}_{i,0} = 1, \mathfrak{X}_{i,1}, \ldots, \mathfrak{X}_{i,p})$ of the same size p. We call n the *size* of the sequence.

The notation $\mathfrak{X}_{\star,k}$ is a little awkward, but we do not need to use it often. The reason for it is that we want to to make sure that one cannot confuse one of the random variables $\mathfrak{X}_{\star,k}$ within an orthonormal ensemble with the orthonormal ensemble $\mathfrak{X}_i$ itself. Whenever a random variable $\mathfrak{X}_{i,k}$ is part of an ensemble $\mathfrak{X}_i$, there is no reason to use the $\star$-symbol. Instead we use the index of the ensemble.

Note that in an orthonormal ensemble for $k > 0$ we have $\mathrm{E}[\mathfrak{X}_{\star,k}] = \mathrm{E}[\mathfrak{X}_{\star,k} \mathfrak{X}_{\star,0}] = 0$.

Definition A.12. We call two ensemble sequences $\underline{\mathfrak{X}} = (\mathfrak{X}_1, \ldots, \mathfrak{X}_n)$ and $\underline{\mathcal{Y}} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_m)$ *compatible* if $n = m$ and the sizes of the individual ensembles $\mathfrak{X}_i$ and $\mathcal{Y}_i$ are the same.

Definition A.13. Let $\underline{\mathfrak{X}} = (\mathfrak{X}_1, \ldots, \mathfrak{X}_n)$ be an ensemble sequence such that each ensemble $\mathfrak{X}_i$ is of size p.

A *monomial compatible with* $\underline{\mathfrak{X}}$ is a term

$$x_\sigma := \prod_{i=1}^{n} x_{i,\sigma_i},$$

where $\sigma = (\sigma_1, \ldots, \sigma_n)$ with $\sigma_i \in \{0, \ldots, p\}$.

A (formal) *multilinear polynomial compatible with* $\underline{\mathfrak{X}}$ is a sum of compatible monomials, i.e., a polynomial P of the form

$$P(\underline{x}) = \sum_{\sigma \in \{0,\ldots,p\}^n} \alpha(\sigma) x_\sigma = \sum_{\sigma \in \{0,\ldots,p\}^n} \alpha(\sigma) \prod_{i=1}^{n} x_{i,\sigma_i},$$

where the sum goes over all tuples $\sigma = (\sigma_1, \ldots, \sigma_n)$ as above, and $\alpha(\sigma) \in \mathbb{R}$.

For a tuple σ we define its *support* as $\mathrm{supp}(\sigma) := \{i \in [n] : \sigma_i \neq 0\}$ and its *degree* as the size of its support: $|\sigma| := |\mathrm{supp}(\sigma)|$. Also, we will write the tuple $(0, \ldots, 0)$ as 0^n.

Let a multilinear polynomial P compatible with $\underline{\mathfrak{X}}$ be given. Then, $P(\underline{\mathfrak{X}})$ is what one expects: the random variable obtained by evaluating the polynomial on the given input. Analogously, if σ is a tuple as above we write $\mathfrak{X}_\sigma$ for the random variable corresponding to the evaluation of the monomial x_σ.

Lemma A.14. *Let $\underline{\mathfrak{X}}$ be an ensemble sequence and σ, τ two tuples whose monomials x_σ, x_τ are compatible with $\underline{\mathfrak{X}}$. Then,*

$$\mathrm{E}[\mathfrak{X}_\sigma \mathfrak{X}_\tau] = \begin{cases} 1 & \text{if } \sigma = \tau \\ 0 & \text{otherwise} \end{cases} \tag{51}$$

and

$$\mathrm{E}[\mathfrak{X}_\sigma] = \begin{cases} 1 & \textit{if } \sigma = 0^n \\ 0 & \textit{otherwise.} \end{cases} \tag{52}$$

Proof. By independence of the coordinates we have $\mathrm{E}[\mathfrak{X}_\sigma \mathfrak{X}_\tau] = \prod_{i=1}^n \mathrm{E}[\mathfrak{X}_{i,\sigma_i} \cdot \mathfrak{X}_{i,\tau_i}]$ and now we can use the orthonomality of each ensemble $\mathfrak{X}_i$. For the second part, we apply the first on $\tau = 0^n$. □

Definition A.15. Given a multilinear polynomial $P(\underline{x}) = \sum_\sigma \alpha(\sigma) x_\sigma$ we define its following properties:

$$\deg(P) := \begin{cases} \max_{\sigma:\alpha_\sigma \neq 0} |\sigma| & \text{if } P \text{ is non-zero} \\ -\infty & \text{if } P \text{ is the zero polynomial} \end{cases} \tag{53}$$

$$\mathrm{E}[P] := \alpha(0^n) \tag{54}$$

$$\mathrm{E}[P^2] := \sum_\sigma \alpha(\sigma)^2 \tag{55}$$

$$\mathrm{Var}[P] := \mathrm{E}[P^2] - \mathrm{E}^2[P] \tag{56}$$

$$\mathrm{Inf}_i(P) := \sum_{\sigma:\sigma_i \neq 0} \alpha(\sigma)^2 \tag{57}$$

$$\mathrm{Inf}(P) := \sum_{i=1}^n \mathrm{Inf}_i(P) \tag{58}$$

The next lemma states that the formal expressions defined above are consistent with the corresponding probabilistic interpretations for every ensemble sequence.

Lemma A.16. *For an ensemble sequence $\underline{\mathfrak{X}}$ and a multilinear polynomial P compatible with it we have*

$$\mathrm{E}[P] = \mathrm{E}[P(\underline{\mathfrak{X}})] \tag{59}$$

$$\mathrm{E}[P^2] = \mathrm{E}[(P(\underline{\mathfrak{X}}))^2] \tag{60}$$

$$\mathrm{Var}[P] = \mathrm{Var}[P(\underline{\mathfrak{X}})]. \tag{61}$$

Furthermore, if all random variables in $\underline{\mathfrak{X}}$ are discrete, then

$$\mathrm{Inf}_i(P) = \mathrm{E}\left[\mathrm{Var}\left[P(\underline{\mathfrak{X}}) \mid \mathfrak{X}_1, \ldots, \mathfrak{X}_{i-1}, \mathfrak{X}_{i+1}, \ldots, \mathfrak{X}_n\right]\right]. \tag{62}$$

Proof. Linearity of expectation and (52) yield $\mathrm{E}[P(\underline{\mathfrak{X}})] = \sum_\sigma \alpha(\sigma) \mathrm{E}[\mathfrak{X}_\sigma] = \alpha(0^n)$, which is (59). Next, (51) gives $\mathrm{E}[P^2(\underline{\mathfrak{X}})] = \sum_{\sigma,\tau} \alpha(\sigma)\alpha(\tau)\mathfrak{X}_\sigma \mathfrak{X}_\tau = \sum_\sigma \alpha(\sigma)^2$, i.e. (60), and hence (61) by the definition of the variance.

As for (62), fix an assignment $\underline{x}_{\backslash i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ to the ensemble sequence $\underline{\mathfrak{X}}_{\backslash i} = (\mathfrak{X}_1, \ldots, \mathfrak{X}_{i-1}, \mathfrak{X}_{i+1}, \ldots, \mathfrak{X}_n)$.[5] We suppose that this tuple has a non-zero probability of occurence. Since $\mathfrak{X}_i$ is an orthornormal ensemble,

$$\mathrm{Var}[P(\underline{\mathfrak{X}}) \mid \underline{\mathfrak{X}}_{\backslash i} = \underline{x}_{\backslash i}] = \sum_{k=1}^p \left(\sum_{\sigma:\sigma_i=k} \alpha(\sigma) \cdot \prod_{j \neq i} x_{j,\sigma_j} \right)^2$$

---

[5]Note that each entry in this tuple is itself a tuple: $x_i = (x_{i,0} = 1, x_{i,1}, \ldots, x_{i,p})$, where p is the size of the ensemble.

From Lemma A.14, for a fixed $k \in \{1, \dots, p\}$,

$$\mathrm{E}\left[\left(\sum_{\sigma:\sigma_i=k} \alpha(\sigma) \cdot \prod_{j \neq i} \mathcal{X}_{j,\sigma_j}\right)^2\right] = \sum_{\sigma:\sigma_i=k} \alpha(\sigma)^2 \,.$$

Together this gives

$$\mathrm{E}\left[\mathrm{Var}\left[P(\underline{\mathcal{X}}) \mid \underline{\mathcal{X}}_{\setminus i}\right]\right] = \sum_{\sigma:\sigma_i \neq 0} \alpha(\sigma)^2 \,,$$

as claimed. $\qquad\square$

Definition A.17. For a multilinear polynomial $P(\underline{x}) = \sum_\sigma \alpha(\sigma)x_\sigma$ and $S \subseteq [n]$ we let P_S be P restricted to tuples σ with $\mathrm{supp}(\sigma) = S$, i.e., $P_S := \sum_{\sigma:\mathrm{supp}(\sigma)=S} \alpha(\sigma)x_\sigma$.

Then, let $P^{>d} := \sum_{S:|S|>d} P_S$ be P restricted to tuples with the degree greater than d. We also define $P^{=d}$, $P^{\leq d}$ etc. in the analogous way.

Lemma A.18. *Let P and Q be multilinear polynomials compatible with an ensemble sequence $\underline{\mathcal{X}}$. Then,*

$$\mathrm{E}[P(\underline{\mathcal{X}})Q(\underline{\mathcal{X}})] = \sum_{S \subseteq [n]} \mathrm{E}[P_S(\underline{\mathcal{X}})Q_S(\underline{\mathcal{X}})] \,.$$

Proof. It is enough to show that for $S \neq T$

$$\mathrm{E}[P_S(\underline{\mathcal{X}})Q_T(\underline{\mathcal{X}})] = 0 \,.$$

Let $P(\underline{\mathcal{X}}) = \sum_\sigma \alpha(\sigma) \cdot \mathcal{X}_\sigma$ and $Q(\underline{\mathcal{X}}) = \sum_\sigma \beta(\sigma) \cdot \mathcal{X}_\sigma$. Assume w.l.o.g. that there exists $i^* \in S \setminus T$. Then,

$$\mathrm{E}[P_S(\underline{\mathcal{X}})Q_T(\underline{\mathcal{X}})] =$$

$$= \sum_{\substack{\sigma:\mathrm{supp}(\sigma)=S \\ \sigma':\mathrm{supp}(\sigma')=T}} \alpha(\sigma)\beta(\sigma')\,\mathrm{E}\left[\mathcal{X}_{i^*,\sigma_{i^*}}\right]\mathrm{E}\left[\prod_{i \neq i^*} \mathcal{X}_{i,\sigma_i}\mathcal{X}_{i,\sigma'_i}\right] = 0 \,.$$

$\qquad\square$

Corollary A.19. *Let P be a multilinear polynomial. Then, $\mathrm{E}[P^2] = \sum_{S \subseteq [n]} \mathrm{E}[P_S^2]$.*

Proof. Taking any ensemble sequence $\underline{\mathcal{X}}$ compatible with P,

$$\mathrm{E}[P^2] = \mathrm{E}[P(\underline{\mathcal{X}})^2] = \sum_{S \subseteq [n]} \mathrm{E}[P_S(\underline{\mathcal{X}})^2] = \sum_{S \subseteq [n]} \mathrm{E}[P_S^2] \,. \qquad\square$$

Claim A.20. *Let P be a multilinear polynomial. Then, $\mathrm{Var}[P] = \sum_{S \subseteq [n]} \mathrm{Var}[P_S]$.*

Proof. Observing that $\mathrm{Var}[P_\emptyset] = 0$, $\mathrm{E}[P_\emptyset^2] = \alpha(0^n)^2$ and $\mathrm{Var}[P_S] = \mathrm{E}[P_S^2]$ for $S \neq \emptyset$, by Corollary A.19

$$\mathrm{Var}[P] = \mathrm{E}[P^2] - \alpha(0^n)^2 = \sum_{S \subseteq [n], S \neq \emptyset} \mathrm{E}[P_S^2] = \sum_{S \subseteq [n]} \mathrm{Var}[P_S] \,. \qquad\square$$

Lemma A.21. *Let P be a multilinear polynomial with $\deg(P) \leq d$. Then,*

$$\mathrm{Inf}(P) \leq d \cdot \mathrm{Var}[P] .$$

Proof.

$$\mathrm{Inf}(P) = \sum_\sigma |\sigma| \cdot \alpha(\sigma)^2 \leq d \cdot \sum_{\sigma \neq 0^n} \alpha(\sigma)^2 = d \cdot \mathrm{Var}[P] . \qquad \square$$

Definition A.22. Let $\rho \in \mathbb{R}$. We define the operator T_ρ as follows: let $P(\underline{x}) = \sum_\sigma \alpha(\sigma) x_\sigma$ be a multilinear polynomial. Then,

$$(T_\rho P)(\underline{x}) := \sum_\sigma \rho^{|\sigma|} \alpha(\sigma) x_\sigma .$$

We will mostly use the operator T_ρ with $\rho \in [0,1]$.

Definition A.23. We call an orthonormal ensemble $\mathcal{G}_\star$ of size p *Gaussian* if random variables $\mathcal{G}_{\star,1}, \ldots, \mathcal{G}_{\star,p}$ are independent $\mathcal{N}(0,1)$ Gaussians.

We say that an ensemble sequence $\underline{\mathcal{G}} = (\mathcal{G}_1, \ldots, \mathcal{G}_n)$ is Gaussian if for each $i \in [n]$ the ensemble $\mathcal{G}_i$ is Gaussian.

We remark than as in all ensemble sequences, in a Gaussian ensemble sequence we have $\mathcal{G}_{i,0} \equiv 1$ for all i.

Definition A.24. For tuples of multilinear polynomials $\overline{P} = (P^{(1)}, \ldots, P^{(\ell)})$ such that each polynomial $P^{(j)}$ is compatible with an ensemble sequence $\underline{\mathcal{X}}$ we write $\overline{P}(\underline{\mathcal{X}})$ for the tuple $(P^{(1)}(\underline{\mathcal{X}}), \ldots, P^{(\ell)}(\underline{\mathcal{X}}))$.

Similarly, given multilinear polynomials $\overline{P} = (P^{(1)}, \ldots, P^{(\ell)})$ and a collection of ensemble sequences $\overline{\underline{\mathcal{X}}} = (\underline{\mathcal{X}}^{(1)}, \ldots \underline{\mathcal{X}}^{(\ell)})$ such that $P^{(j)}$ is compatible with $\underline{\mathcal{X}}^{(j)}$ we write $\overline{P}(\overline{\underline{\mathcal{X}}})$ for $(P^{(1)}(\underline{\mathcal{X}}^{(1)}), \ldots, P^{(\ell)}(\underline{\mathcal{X}}^{(\ell)}))$.

A.3 Preliminaries — ensemble collections

In this section we recall the setting of Theorem 4.1 and introduce some other concepts we will need throughout the proof.

From now on we will always implicitly assume that all multi-step distributions $\mathcal{P}$ have equal marginals (denoted as π). This assumption is not necessary, but sufficient for our main purpose, while making the notation easier.

Definition A.25. Let X be a random variable distributed according to a single-step, single-coordinate distribution (Ω, π). We say that an orthonormal ensemble $\mathcal{X}_\star$ is *constructed from X* if the elements of $\mathcal{X}_\star$ form an orthonormal basis of $L^2(X)$.

Similarly, let $\underline{X}$ be a random vector distributed according to $(\underline{\Omega}, \underline{\pi})$. We say that an ensemble sequence $\underline{\mathcal{X}} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ is constructed from $\underline{X}$ if for each $i \in [n]$ the ensemble $\mathcal{X}_i$ is constructed from X_i.

The definition of ensemble sequences requires that $\mathcal{X}_{i,0} \equiv 1$ for every i; of course we can find a basis of $L^2(X_i)$ which satisfies this requirement, so that ensemble sequences constructed from $\underline{X}$ indeed exist.

Lemma A.26. *Let $\underline{\mathcal{X}}$ be an ensemble sequence constructed from a random vector $\underline{X}$ distributed according to $(\underline{\Omega}, \underline{\pi})$. Assume that the size of each ensemble $\mathcal{X}_i$ is p. Then the set of monomials*

$$\underline{\mathcal{B}} := \{\mathcal{X}_\sigma \mid \sigma = (\sigma_1, \ldots, \sigma_n), \sigma_i \in \{0, \ldots, p\}\}$$

is an orthonormal basis of $L^2(\underline{X})$.

Proof. Observe that the dimension of $L^2(X_i)$ is $p + 1$, (note that it is the support size of the single-coordinate distribution (Ω, π)). Hence, the dimension of $L^2(\underline{X})$ is $(p+1)^n$, which equals the size of $\underline{\mathcal{B}}$. Therefore, it is enough to check that $\underline{\mathcal{B}}$ is orthonormal, which is done in Lemma A.14. $\square$

Definition A.27. Let $\underline{\mathcal{X}}$ be an ensemble sequence constructed from a random vector $\underline{X}$ distributed according to $(\underline{\Omega}, \underline{\pi})$.

For a function $f : \underline{\Omega} \to \mathbb{R}$ and a multilinear polynomial P compatible with $\underline{\mathcal{X}}$ we say that $f(\underline{X})$ is *equivalent* to P if it always holds that

$$f(\underline{X}) = P(\underline{\mathcal{X}}) .$$

Recall the operator T_ρ from Definition A.22. We show that it has a natural counterpart in $L^2(\underline{\Omega}, \underline{\pi})$.

Definition A.28. Let $\rho \in [0, 1]$ and let $(\underline{\Omega}, \underline{\pi})$ be a single-step probability space (with (Ω, π) a corresponding single-coordinate probability space).

We define a linear operator $T_\rho : L^2(\underline{\Omega}, \underline{\pi}) \to L^2(\underline{\Omega}, \underline{\pi})$ as

$$T_\rho f(\underline{x}) := \mathrm{E}\left[f\left(\underline{Y}^{\rho, \underline{x}}\right)\right] ,$$

where $\underline{Y}^{\rho, \underline{x}} = (Y_1^{\rho, \underline{x}}, \ldots, Y_n^{\rho, \underline{x}})$ is a random vector with independent coordinates distributed such that $Y_i^{\rho, \underline{x}} = x_i$ with probability ρ and $Y_i^{\rho, \underline{x}}$ is (independently) distributed according to (Ω, π) with probability $(1 - \rho)$.

The next lemma states that taking operator T_ρ preserves the equivalence of functions and polynomials:

Lemma A.29. *Let $\underline{\mathcal{X}}$ be an ensemble sequence constructed from a random vector $\underline{X}$ distributed according to $(\underline{\Omega}, \underline{\pi})$.*

Let $\rho \in [0, 1]$, $f : \underline{\Omega} \to \mathbb{R}$ and P be a multilinear polynomial equivalent to f. Then, $T_\rho P$ and $T_\rho f$ are equivalent, i.e.,

$$T_\rho f(\underline{X}) = T_\rho P(\underline{\mathcal{X}}) .$$

Proof. Fix an input $\underline{x} \in \underline{\Omega}$ in the support of $\underline{\mathcal{P}}$. Let $\underline{\mathcal{Y}}^{\rho, \underline{x}} = (\mathcal{Y}_1^{\rho, \underline{x}}, \ldots, \mathcal{Y}_n^{\rho, \underline{x}})$ be the random sequence where for each coordinate $i \in [n]$, independently

$$\mathcal{Y}_i^{\rho, \underline{x}} := \begin{cases} \mathcal{X}_i(x_i) & \text{with probability } \rho, \\ \text{a random ensemble distributed as } \mathcal{X}_i & \text{with probability } 1 - \rho. \end{cases}$$

Note that $\underline{\mathcal{Y}}^{\rho,\underline{x}}$ is not an ensemble sequence, but this will not cause problems.

Writing $P(\underline{x}) = \sum_\sigma \alpha(\sigma) \cdot x_\sigma$ we can calculate

$$T_\rho f(\underline{x}) = \mathrm{E}[f(\underline{Y}^{\rho,\underline{x}})] = \mathrm{E}[P(\underline{\mathcal{Y}}^{\rho,\underline{x}})] = \sum_\sigma \alpha(\sigma) \mathrm{E}[\mathcal{Y}_\sigma^{\rho,\underline{x}}]$$

$$= \sum_\sigma \rho^{|\sigma|} \alpha(\sigma) \cdot \mathcal{X}_\sigma(\underline{x}) = T_\rho P(\underline{x}) \ .$$

Since $\underline{x}$ was arbitrary, the claim is proved. $\qquad\square$

Recall Definition A.23. In the proof we will construct a tuple of ensemble sequences $\overline{\underline{\mathcal{X}}} = (\underline{\mathcal{X}}^{(1)}, \ldots, \underline{\mathcal{X}}^{(\ell)})$ from a random vector $\overline{X}$ and consider relations between those sequences and compatible Gaussian ensemble sequences. To this end, we need to introduce the Gaussian equivalent of marginal ensemble sequences $\underline{\mathcal{X}}^{(j)}$.

Definition A.30. Let $\mathcal{G}_\star = (\mathcal{G}_{\star,0}, \ldots, \mathcal{G}_{\star,p})$ be a Gaussian orthonormal ensemble of size p. We define an inner product space $V(\mathcal{G})$ as

$$V(\mathcal{G}) := \left\{ \sum_{k=0}^p \alpha_k \cdot \mathcal{G}_k \mid \alpha_0, \ldots, \alpha_k \in \mathbb{R} \right\}$$

with the inner product of $A, B \in V(\mathcal{G})$ given by $\langle A, B \rangle := \mathrm{E}[A \cdot B]$.

Similarly, given a Gaussian ensemble sequence $\underline{\mathcal{G}}$ such that each of its ensembles is of size p we let

$$V(\underline{\mathcal{G}}) := \left\{ \sum_\sigma \alpha(\sigma) \cdot \mathcal{G}_\sigma \mid \sigma = (\sigma_1, \ldots, \sigma_n) \in \{0, \ldots, p\}, \alpha(\sigma) \in \mathbb{R} \right\} \ ,$$

with the inner product $\langle A, B \rangle := \mathrm{E}[A \cdot B]$.

Lemma A.31. *Let a random tuple $\overline{X} = (X^{(1)}, \ldots, X^{(\ell)})$ be distributed according to a single-coordinate distribution $(\overline{\Omega}, \mathcal{P})$. Let $\overline{\underline{\mathcal{X}}}_\star = (\underline{\mathcal{X}}_\star^{(1)}, \ldots, \underline{\mathcal{X}}_\star^{(\ell)})$ be such that $\mathcal{X}_\star^{(j)}$ is an orthonormal ensemble constructed from $X^{(j)}$.*

Then, there exist Gaussian orthonormal ensembles $\overline{\mathcal{G}}_\star = (\mathcal{G}_\star^{(1)}, \ldots, \mathcal{G}_\star^{(\ell)})$ compatible with $\overline{\underline{\mathcal{X}}}_\star$ such that for all $j_1, j_2 \in [\ell]$, and all $k_1, k_2 \geq 0$ we have

$$\mathrm{Cov}\left[\mathcal{X}_{\star,k_1}^{(j_1)}, \mathcal{X}_{\star,k_2}^{(j_2)}\right] = \mathrm{Cov}\left[\mathcal{G}_{\star,k_1}^{(j_1)}, \mathcal{G}_{\star,k_2}^{(j_2)}\right] \ . \tag{63}$$

Proof. Consider $(\overline{\Omega}, \mathcal{P})$ as a single-step probability space, and let $\overline{X}$ be the corresponding random variable. Let now $\mathcal{Z}_\star$ be an orthonormal ensemble constructed from $\overline{X}$. Recall that this means that the elements of $\mathcal{Z}_\star$ form an orthonormal basis of $L^2(\overline{X})$.

Let $\mathcal{H}_\star$ be a Gaussian ensemble sequence compatible with $\mathcal{Z}_\star$. Define the map $\Psi : L^2(\overline{X}) \to V(\mathcal{H}_\star)$ by linearly extending $\Psi(\mathcal{Z}_{\star,k}) := \mathcal{H}_{\star,k}$. In this way Ψ becomes an isomorphism between $L^2(\overline{X})$ and $V(\mathcal{H}_\star)$ (and as such it preserves inner products).

Since $L^2(X^{(j)})$ is a subspace of $L^2(\overline{X})$, we can define $\mathcal{G}_{\star,k}^{(j)}$ as $\mathcal{G}_{\star,k}^{(j)} := \Psi(\mathcal{X}_{\star,k}^{(j)})$. Since Ψ preserves inner products we get (63).

We still need to argue that for each $j \in [\ell]$ the orthonormal ensemble $\mathcal{G}_\star^{(j)}$ is Gaussian. The fact that $\mathcal{G}_\star^{(j)}$ is an ensemble sequence follows from (63) for $j_1 = j_2 = j$ (note that $\Psi(1) = 1$).

The variables $\mathcal{G}_{\star,k}^{(j)}$ are clearly jointly Gaussian, since they can be written as sums of independent Gaussians. By (63), their covariance matrix is identity. This finishes the proof, since joint Gaussians with the identity covariance matrix must be independent. $\square$

Since the proof of Lemma A.31 is somewhat abstract, we illustrate the construction of $\overline{\mathcal{G}}_\star$ with an example.

Example A.32. Consider $(X^{(1)}, X^{(2)})$ distributed according to $\mathcal{P}$ over $\Omega = \{0,1\}$ with $\mathcal{P}(0,0) = \mathcal{P}(1,1) = 1/8$ and $\mathcal{P}(0,1) = \mathcal{P}(1,0) = 3/8$. We can take the following for the ensemble $\mathcal{Z}_\star$:

$(X^{(1)}, X^{(2)}) :=$	(0,0)	(0, 1)	(1, 0)	(1, 1)
$\mathcal{Z}_{\star,0}$	1	1	1	1
$\mathcal{Z}_{\star,1}$	2	0	0	-2
$\mathcal{Z}_{\star,2}$	0	$2\sqrt{3}/3$	$-2\sqrt{3}/3$	0
$\mathcal{Z}_{\star,3}$	$\sqrt{3}$	$-\sqrt{3}/3$	$-\sqrt{3}/3$	$\sqrt{3}$

For the marginal ensemble $\mathcal{X}_\star^{(1)}$ we can take

$X^{(1)} :=$	0	1
$\mathcal{X}_{\star,0}^{(1)}$	1	1
$\mathcal{X}_{\star,1}^{(1)}$	1	-1

Now one can check that $\mathcal{X}_{\star,0}^{(1)} = \mathcal{Z}_{\star,0}$ and $\mathcal{X}_{\star,1}^{(1)} = 1/2 \cdot \mathcal{Z}_{\star,1} + \sqrt{3}/2 \cdot \mathcal{Z}_{\star,2}$. Defining the ensemble $\mathcal{X}_\star^{(2)}$ in the same way we get $\mathcal{X}_{\star,0}^{(2)} = \mathcal{Z}_{\star,0}$ and $\mathcal{X}_{\star,1}^{(2)} = 1/2 \cdot \mathcal{Z}_{\star,1} - \sqrt{3}/2 \cdot \mathcal{Z}_{\star,2}$.

Let $\mathcal{H}_\star = (\mathcal{H}_{\star,0} \equiv 1, \mathcal{H}_{\star,1}, \mathcal{H}_{\star,2}, \mathcal{H}_{\star,3})$ be a Gaussian ensemble sequence compatible with $\mathcal{Z}$. One easily checks that our construction gives

$$\mathcal{G}_{\star,0}^{(1)} = \mathcal{G}_{\star,0}^{(2)} = \mathcal{H}_{\star,0}$$

$$\mathcal{G}_{\star,1}^{(1)} = 1/2 \cdot \mathcal{H}_{\star,1} + \sqrt{3}/2 \cdot \mathcal{H}_{\star,2}$$

$$\mathcal{G}_{\star,1}^{(2)} = 1/2 \cdot \mathcal{H}_{\star,1} - \sqrt{3}/2 \cdot \mathcal{H}_{\star,2} \ .$$

Since the covariances between independent coordinates are always zero, Lemma A.31 applied to each coordinate separately gives:

Corollary A.33. *Let a random vector $\overline{X} = (X^{(1)}, \ldots, X^{(\ell)})$ be distributed according to a distribution $(\overline{\Omega}, \mathcal{P})$. Let $\overline{\mathcal{X}} = (\mathcal{X}^{(1)}, \ldots, \mathcal{X}^{(\ell)})$ be such that $\mathcal{X}^{(j)}$ is an ensemble sequence constructed from $\underline{X}^{(j)}$.*

Then, there exist Gaussian ensemble sequences $\overline{\mathcal{G}} = (\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(\ell)})$ compatible with $\overline{\mathcal{X}}$ such that for all $i_1, i_2 \in [n]$, $j_1, j_2 \in [\ell]$, and all $k_1, k_2 \geq 0$ we have

$$\mathrm{Cov}\left[\mathcal{X}_{i_1,k_1}^{(j_1)}, \mathcal{X}_{i_2,k_2}^{(j_2)}\right] = \mathrm{Cov}\left[\mathcal{G}_{i_1,k_1}^{(j_1)}, \mathcal{G}_{i_2,k_2}^{(j_2)}\right] \ . \tag{64}$$

Definition A.34. An *ensemble collection for* $(\overline{\Omega}, \mathcal{P})$ is a tuple

$$\left(\overline{X}, \underline{\overline{X}} = (\underline{X}^{(1)}, \dots, \underline{X}^{(\ell)}), \overline{\underline{G}} = (\underline{G}^{(1)}, \dots, \underline{G}^{(\ell)}) \right)$$

where

- $\overline{X}$ is a random vector distributed according to $(\overline{\Omega}, \mathcal{P})$,

- $\underline{X}^{(1)}, \dots, \underline{X}^{(\ell)}$ are ensemble sequences constructed from $\underline{X}^{(1)}, \dots, \underline{X}^{(\ell)}$, respectively,

- and $\underline{G}^{(1)}, \dots, \underline{G}^{(\ell)}$ are obtained from Corollary A.33.

A.4 Hypercontractivity

In this section we develop a version of hypercontractivity for products of multilinear polynomials. Our goal is to prove Lemma A.43.

Recall the operator T_ρ from Definition A.22.

Definition A.35. Let $\underline{X}$ be an ensemble sequence and let $1 \le p \le q < \infty$ and $\rho \in [0,1]$. We say that the sequence $\underline{X}$ is (p,q,ρ)-*hypercontractive* if for every multilinear polynomial P compatible with $\underline{X}$ we have

$$\mathrm{E}\left[\left| T_\rho P(\underline{X}) \right|^q \right]^{1/q} \le \mathrm{E}\left[|P(\underline{X})|^p \right]^{1/p}$$

Definition A.36. Let $\mathcal{X}$ be an orthonormal ensemble and let $1 \le p \le q < \infty$ and $\rho \in [0,1]$. We say that the ensemble $\mathcal{X}$ is (p,q,ρ)-*hypercontractive* if the one-element ensemble sequence $\underline{X} := (\mathcal{X})$ is (p,q,ρ)-hypercontractive.

We start with stating without proofs the hypercontractivity of orthonormal ensembles that we use in the invariance principle:

Theorem A.37 ([Bon70, Nel73, Gro75, Bec75])**.** *Let $\mathcal{G}$ be a Gaussian orthonormal ensemble and $\rho \in [0, \sqrt{2}/2]$. Then, $\mathcal{G}$ is $(2,3,\rho)$-hypercontractive.*

Theorem A.38 (Special case of Theorem 3.1 in [Wol07])**.** *Let $\mathcal{X}$ be an orthonormal ensemble constructed from a random variable X distributed according to a (single-coordinate, single-step) probability space (Ω, π) with $\min_{x \in \Omega} \pi(x) \ge \alpha \ge 0$.*
Then, $\mathcal{X}$ is $(2,3,\alpha^{1/6}/2)$-hypercontractive.

Subsequently, we observe that an ensemble sequence constructed from hypercontractive ensembles is itself hypercontractive:

Theorem A.39. *Let $1 \le p \le q < \infty$, $\rho \in [0,1]$ and let $\underline{X} := (\mathcal{X}_1, \dots, \mathcal{X}_n)$ be an ensemble sequence such that for every $i \in [n]$, the ensemble $\mathcal{X}_i$ is (p,q,ρ)-hypercontractive. Then, the sequence $\underline{X}$ is also (p,q,ρ)-hypercontractive.*

Yet again, we omit the proof of Theorem A.39. We remark that it is well-known as the *tensorization argument*. The argument can be found, e.g., in the proof of Proposition 3.11 in [MOO10].

Definition A.40. Let $\underline{X}$ be a random vector distributed according to a (single-step, tensorized) probability space $(\underline{\Omega}, \underline{\pi})$. We say that an ensemble sequence $\underline{\mathfrak{X}} = (\mathfrak{X}_1, \ldots, \mathfrak{X}_n)$ is $\underline{X}$-*Gaussian-mixed* if for each $i \in [n]$:

- Either $\mathfrak{X}_i$ is constructed from the random variable X_i,

- or $\mathfrak{X}_i$ is a Gaussian ensemble.

Theorems A.37, A.38 and A.39 immediately imply:

Corollary A.41. *Let $\underline{X}$ be a random vector distributed according to a probability space $(\underline{\Omega}, \underline{\pi})$ with $\min_{x \in \Omega} \pi(x) \geq \alpha \geq 0$ and let $\underline{\mathfrak{X}}$ be an $\underline{X}$-Gaussian-mixed ensemble sequence.*
Then, $\underline{\mathfrak{X}}$ is $(2, 3, \alpha^{1/6}/2)$-hypercontractive.

Theorem A.42. *Let $\underline{X}$ be a random vector distributed according to a probability space $(\underline{\Omega}, \underline{\pi})$ with $\min_{x \in \Omega} \pi(x) \geq \alpha > 0$ and let $\underline{\mathfrak{X}}$ be an $\underline{X}$-Gaussian-mixed ensemble sequence. Let P be a multilinear polynomial compatible with $\underline{\mathfrak{X}}$ of degree at most d. Then,*

$$\mathrm{E}\left[|P(\underline{\mathfrak{X}})|^3\right]^{1/3} \leq \left(\frac{2}{\alpha^{1/6}}\right)^d \sqrt{\mathrm{E}[P^2]}.$$

Proof. Let $\rho := \alpha^{1/6}/2$ and write $P(\underline{\mathfrak{X}}) = \sum_\sigma \beta(\sigma) \mathfrak{X}_\sigma$. By Corollary A.41, definitions of T_ρ and $E[P^2]$, and the degree bound on P,

$$\mathrm{E}\left[|P(\underline{\mathfrak{X}})|^3\right]^{1/3} = \mathrm{E}\left[\left|T_\rho T_{1/\rho} P(\underline{\mathfrak{X}})\right|^3\right]^{1/3} \leq \sqrt{\mathrm{E}\left[(T_{1/\rho}P)^2\right]}$$
$$= \sqrt{\sum_\sigma \rho^{-2|\sigma|} \beta(\sigma)^2} \leq \sqrt{\sum_\sigma \rho^{-2d} \beta(\sigma)^2} = \rho^{-d} \sqrt{\mathrm{E}[P^2]}.$$

$\square$

Lemma A.43. *Let $\overline{X}$ be a random vector distributed according to a (multi-step) probability space with equal marginals $(\overline{\underline{\Omega}}, \mathcal{P})$ with $\min_{x \in \Omega} \pi(x) \geq \alpha > 0$.*
Let $\underline{S}^{(1)}, \ldots, \underline{S}^{(\ell)}$ be ensemble sequences such that $\underline{S}^{(j)}$ is $\underline{X}^{(j)}$-Gaussian-mixed. Let $P^{(1)}, \ldots, P^{(\ell)}$ be multilinear polynomials such that $P^{(j)}$ is compatible with $\underline{S}^{(j)}$ and also $\deg(P^{(j)}) \leq d$.
Then, for every triple $j_1, j_2, j_3 \in [\ell]$:

$$\mathrm{E}\left[\left|\prod_{k=1}^3 P^{(j_k)}(\underline{S}^{(j_k)})\right|\right] \leq \left(\frac{8}{\sqrt{\alpha}}\right)^d \cdot \sqrt{\prod_{k=1}^3 \mathrm{E}\left[(P^{(j_k)})^2\right]}.$$

Proof. Let $\rho := \alpha^{1/6}/2$. By Hölder's inequality and Theorem A.42,

$$\mathrm{E}\left[\left|\prod_{k=1}^3 P^{(j_k)}(\underline{S}^{(j_k)})\right|\right] \leq \prod_{k=1}^3 \mathrm{E}\left[\left|P^{(j_k)}(\underline{S}^{(j_k)})\right|^3\right]^{1/3}$$
$$\leq \rho^{-3d} \cdot \sqrt{\prod_{k=1}^3 \mathrm{E}\left[(P^{(j_k)})^2\right]}.$$

$\square$

A.5 Invariance principle

In this section we prove a basic version of invariance principle for multiple polynomials.

We say that a function is *B*-smooth if all of its third-order partial derivatives are uniformly bounded by *B*:

Definition A.44. For $B \geq 0$ we say that a function $\Psi : \mathbb{R}^\ell \to \mathbb{R}$ is *B-smooth* if $\Psi \in \mathcal{C}^3$ and for every $j_1, j_2, j_3 \in [\ell]$ and every $\bar{x} = (x^{(1)}, \ldots, x^{(\ell)}) \in \mathbb{R}^\ell$ we have

$$\left| \frac{\partial^3}{\partial x^{(j_1)} \partial x^{(j_2)} \partial x^{(j_3)}} \Psi(\bar{x}) \right| \leq B .$$

Theorem A.45 (Invariance Principle). *Let $(\overline{X}, \overline{\mathfrak{X}}, \overline{\mathcal{G}})$ be an ensemble collection for a probability space $(\overline{\Omega}, \mathcal{P})$ with $\min_{x \in \Omega} \pi(x) \geq \alpha > 0$.*

Let $\overline{P} = (P^{(1)}, \ldots, P^{(\ell)})$ be such that $P^{(j)}$ is a multilinear polynomial compatible with the ensemble sequence $\underline{\mathfrak{X}}^{(j)}$.

Let $d \in \mathbb{N}$ and $\tau \in [0,1]$ and assume that $\deg(P^{(j)}) \leq d$ and $\mathrm{Var}[P^{(j)}] \leq 1$ for each $j \in [\ell]$, and that $\sum_{j=1}^\ell \mathrm{Inf}_i(P^{(j)}) \leq \tau$ for each $i \in [n]$.

Finally, let $\Psi : \mathbb{R}^\ell \to \mathbb{R}$ be a B-smooth function. Then,

$$\left| \mathrm{E} \left[\Psi(\overline{P}(\underline{\mathfrak{X}})) - \Psi(\overline{P}(\overline{\mathcal{G}})) \right] \right| \leq \frac{\ell^{5/2} dB}{3} \left(\frac{8}{\sqrt{\alpha}} \right)^d \sqrt{\tau} .$$

Remark A.46. A typical setting of parameters for which Theorem A.45 might be successfully applied is constant ℓ, d, B, and α, while $\tau = o(1)$ (as $n \to \infty$).

The rest of this section is concerned with proving Theorem A.45.

For $i \in \{0, \ldots, n\}$ and $j \in [\ell]$ let the ensemble sequence $\underline{\mathfrak{U}}_{(i)}^{(j)}$ be defined as $\underline{\mathfrak{U}}_{(i)}^{(j)} := (\mathcal{G}_1^{(j)}, \ldots, \mathcal{G}_i^{(j)}, \mathfrak{X}_{i+1}^{(j)}, \ldots, \mathfrak{X}_n^{(j)})$.

Claim A.47.

$$\left| \mathrm{E} \left[\Psi(\overline{P}(\underline{\mathfrak{X}})) - \Psi(\overline{P}(\overline{\mathcal{G}})) \right] \right| \leq \sum_{i=1}^n \left| \mathrm{E} \left[\Psi(\overline{P}(\underline{\mathfrak{U}}_{(i-1)})) - \Psi(\overline{P}(\underline{\mathfrak{U}}_{(i)})) \right] \right| .$$

Proof. By the triangle inequality. □

Due to Claim A.47, we will estimate

$$\left| \mathrm{E} \left[\Psi(\overline{P}(\underline{\mathfrak{U}}_{(i-1)})) - \Psi(\overline{P}(\underline{\mathfrak{U}}_{(i)})) \right] \right|$$

for every $i \in [n]$. Fix $i \in [n]$ and write $\underline{\mathfrak{T}}^{(j)} := \underline{\mathfrak{U}}_{(i-1)}^{(j)}$ and $\underline{\mathfrak{U}}^{(j)} := \underline{\mathfrak{U}}_{(i)}^{(j)}$ for readability. For $j \in [\ell]$ we can write

$$P^{(j)}(\underline{\mathfrak{T}}^{(j)}) = A^{(j)} + \sum_{k>0} \mathfrak{X}_{i,k}^{(j)} \cdot B_k^{(j)} = A^{(j)} + P_i^{(j)}(\underline{\mathfrak{T}}^{(j)}) , \tag{65}$$

where $A^{(j)}$ and $B_k^{(j)}$ do not depend on the coordinate i and, if $P^{(j)}(\underline{\mathfrak{T}}^{(j)}) = \sum_\sigma \alpha(\sigma)\mathfrak{T}_\sigma^{(j)}$, then $P_i^{(j)}(\underline{\mathfrak{T}}^{(j)}) = \sum_{\sigma:i\in\text{supp}(\sigma)} \alpha(\sigma)\mathfrak{T}_\sigma^{(j)}$. At the same time, since $A^{(j)}$ and $B_k^{(j)}$ do not depend on the i-th coordinate,

$$P^{(j)}(\underline{\mathfrak{U}}^{(j)}) = A^{(j)} + \sum_{k>0} \mathfrak{G}_{i,k}^{(j)} \cdot B_k^{(j)} = A^{(j)} + P_i^{(j)}(\underline{\mathfrak{U}}^{(j)}) \,.$$

We note for later use that the construction gives us

$$\deg(P_i^{(j)}) \leq d \tag{66}$$

$$\mathrm{E}\left[\left(P_i^{(j)}\right)^2\right] = \mathrm{Inf}_i\left(P^{(j)}\right) \,. \tag{67}$$

The rest of the proof proceeds as follows: we calculate the multivariate second order Taylor expansion (i.e., with the third-degree rest) of the expression, getting

$$\Psi(\overline{P}(\underline{\mathfrak{T}})) - \Psi(\overline{P}(\underline{\mathfrak{U}})) =$$

$$= \Psi\left(A^{(1)} + \sum_{k>0} \mathfrak{X}_{i,k}^{(1)} B_k^{(1)}, \ldots, A^{(\ell)} + \sum_{k>0} \mathfrak{X}_{i,k}^{(\ell)} B_k^{(\ell)}\right)$$

$$- \Psi\left(A^{(1)} + \sum_{k>0} \mathfrak{G}_{i,k}^{(1)} B_k^{(1)}, \ldots, A^{(\ell)} + \sum_{k>0} \mathfrak{G}_{i,k}^{(\ell)} B_k^{(\ell)}\right)$$

around the point $\overline{A} := (A^{(1)}, \ldots, A^{(\ell)})$. We will see that:

- All the terms up to the second degree cancel in expectation due to the properties of ensemble sequences.

- The remainder, which is of the third degree, can be bounded using that Ψ is B-smooth, properties of $P_i^{(j)}$, and hypercontractivity, in particular Lemma A.43.

We proceed with a detailed description. The first result we will need is multivariate Taylor's theorem for B-smooth functions:

Theorem A.48. *Let $\Psi : \mathbb{R}^\ell \to \mathbb{R}$ be a B-smooth function and let $\overline{x} = (x^{(1)}, \ldots, x^{(\ell)}), \overline{\varepsilon} = (\varepsilon^{(1)}, \ldots, \varepsilon^{(\ell)}) \in \mathbb{R}^\ell$. Then,*

$$\left| \Psi\left(x^{(1)} + \varepsilon^{(1)}, \ldots, x^{(\ell)} + \varepsilon^{(\ell)}\right) - \right.$$

$$\left. \left(\Psi(\overline{x}) + \sum_{j\in[\ell]} \varepsilon^{(j)} \frac{\partial}{\partial x^{(j)}}\Psi(\overline{x}) + \frac{1}{2} \sum_{j_1,j_2\in[\ell]} \varepsilon^{(j_1)}\varepsilon^{(j_2)} \frac{\partial^2}{\partial x^{(j_1)}\partial x^{(j_2)}}\Psi(\overline{x})\right) \right|$$

$$\leq \frac{B}{6} \sum_{j_1,j_2,j_3\in[\ell]} \left|\varepsilon^{(j_1)}\varepsilon^{(j_2)}\varepsilon^{(j_3)}\right| \,.$$

We omit the proof of Theorem A.48.

Lemma A.49. *Fix $i \in [n]$ and write $\underline{\mathfrak{T}}^{(j)} := \underline{\mathfrak{U}}^{(j)}_{(i-1)}$ and $\underline{\mathfrak{U}}^{(j)} := \underline{\mathfrak{U}}^{(j)}_{(i)}$. Then,*

$$
\begin{aligned}
\mathrm{E}\left[\Psi(\overline{P}(\underline{\mathfrak{T}}))\right] &= \\
&= \mathrm{E}\left[\Psi(\overline{A}) + \frac{1}{2} \sum_{j_1, j_2 \in [\ell]} \left(\sum_{k_1, k_2 > 0} \mathcal{X}^{(j_1)}_{i,k_1} \mathcal{X}^{(j_2)}_{i,k_2} B^{(j_1)}_{k_1} B^{(j_2)}_{k_2} \frac{\partial^2}{\partial A^{(j_1)} \partial A^{(j_2)}} \Psi(\overline{A})\right) + R_{\underline{\mathfrak{T}}}\right],
\end{aligned} \tag{68}
$$

and

$$
\begin{aligned}
\mathrm{E}\left[\Psi(\overline{P}(\underline{\mathfrak{U}}))\right] &= \\
&= \mathrm{E}\left[\Psi(\overline{A}) + \frac{1}{2} \sum_{j_1, j_2 \in [\ell]} \left(\sum_{k_1, k_2 > 0} \mathcal{G}^{(j_1)}_{i,k_1} \mathcal{G}^{(j_2)}_{i,k_2} B^{(j_1)}_{k_1} B^{(j_2)}_{k_2} \frac{\partial^2}{\partial A^{(j_1)} \partial A^{(j_2)}} \Psi(\overline{A})\right) + R_{\underline{\mathfrak{U}}}\right],
\end{aligned} \tag{69}
$$

where random variables $R_{\underline{\mathfrak{T}}}$ and $R_{\underline{\mathfrak{U}}}$ are such that

$$
\mathrm{E}\left[\left|R_{\underline{\mathfrak{T}}}\right|\right], \mathrm{E}\left[\left|R_{\underline{\mathfrak{U}}}\right|\right] \le \frac{\ell^{3/2} B}{6} \left(\frac{8}{\sqrt{\alpha}}\right)^d \left(\sum_{j=1}^{\ell} \mathrm{Inf}_i(P^{(j)})\right)^{3/2}. \tag{70}
$$

Proof. We show only (68) and the bound on $\mathrm{E}[|R_{\underline{\mathfrak{T}}}|]$, the proofs for the ensemble sequence $\underline{\mathfrak{U}}$ being analogous.

As a preliminary remark, note that since all the random ensembles we are dealing with are hypercontractive, and since Ψ is B-smooth, all the terms in the expressions above have finite expectations.

Keeping in mind both decompositions from (65), by Theorem A.48

$$
\begin{aligned}
\Psi(\overline{P}(\underline{\mathfrak{T}})) &= \Psi(\overline{A}) + \sum_{j \in [\ell]} \left(\sum_{k > 0} \mathcal{X}^{(j)}_{i,k} B^{(j)}_k \frac{\partial}{\partial A^{(j)}} \Psi(\overline{A})\right) + \\
&\quad + \frac{1}{2} \sum_{j_1, j_2 \in [\ell]} \left(\sum_{k_1, k_2 > 0} \mathcal{X}^{(j_1)}_{i,k_1} \mathcal{X}^{(j_2)}_{i,k_2} B^{(j_1)}_{k_1} B^{(j_2)}_{k_2} \frac{\partial^2}{\partial A^{(j_1)} \partial A^{(j_2)}} \Psi(\overline{A})\right) + R_{\underline{\mathfrak{T}}},
\end{aligned} \tag{71}
$$

where

$$
\mathrm{E}[|R_{\underline{\mathfrak{T}}}|] \le \frac{B}{6} \sum_{j_1, j_2, j_3 \in [\ell]} \mathrm{E}\left[\left|\prod_{k=1}^{3} P^{(j_k)}_i(\underline{\mathfrak{T}}^{(j_k)})\right|\right]. \tag{72}
$$

Since $\mathrm{E}[\mathcal{X}^{(j)}_{i,k}] = 0$, and all other terms are independent of coordinate i, we have

$$
\mathrm{E}\left[\sum_{j \in [\ell]} \sum_{k > 0} \mathcal{X}^{(j)}_{i,k} B^{(j)}_k \frac{\partial}{\partial A^{(j)}} \Psi(\overline{A})\right] = 0,
$$

which together with (71) yields (68).

As for the bound on $\mathrm{E}[|R_{\underline{\mathcal{T}}}|]$, since $T^{(j)}$ is $\underline{X}^{(j)}$-Gaussian-mixed ensemble sequence, due to (72), Lemma A.43 (note that the degree is bounded due to (66)), and (67),

$$
\begin{aligned}
\mathrm{E}[|R_{\underline{\mathcal{T}}}|] &\leq \frac{B}{6}\left(\frac{8}{\sqrt{\alpha}}\right)^d \sum_{j_1,j_2,j_3\in[\ell]} \sqrt{\prod_{k=1}^3 \mathrm{E}\left[\left(P_i^{(j_k)}\right)^2\right]} \\
&= \frac{B}{6}\left(\frac{8}{\sqrt{\alpha}}\right)^d \sum_{j_1,j_2,j_3\in[\ell]} \sqrt{\prod_{k=1}^3 \mathrm{Inf}_i(P^{(j_k)})} \\
&\leq \frac{\ell^{3/2}B}{6}\left(\frac{8}{\sqrt{\alpha}}\right)^d \left(\sum_{j=1}^\ell \mathrm{Inf}_i(P^{(j)})\right)^{3/2},
\end{aligned}
$$

where the last inequality uses $\sum_{j_1,j_2,j_3} v(j_1,j_2,j_3) \leq \sqrt{\ell^3}\sqrt{\sum v^2(j_1,j_2,j_3)}$ for the vector v with entries $v(j_1,j_2,j_3) = \sqrt{\prod_{k=1}^3 \mathrm{Inf}_i(P^{(j_k)})}$. $\qquad\square$

Lemma A.50. *Fix $i \in [n]$ and write $\underline{\mathcal{T}}^{(j)} := \underline{\mathcal{U}}_{(i-1)}^{(j)}$ and $\underline{\mathcal{U}}^{(j)} := \underline{\mathcal{U}}_{(i)}^{(j)}$. Then,*

$$
\left|\mathrm{E}\left[\Psi(\overline{P}(\overline{\mathcal{T}})) - \Psi(\overline{P}(\overline{\mathcal{U}}))\right]\right| \leq \frac{\ell^{3/2}B}{3}\left(\frac{8}{\sqrt{\alpha}}\right)^d \left(\sum_{j=1}^\ell \mathrm{Inf}_i(P^{(j)})\right)^{3/2}.
$$

Proof. First, we need to show that the second-order terms in (68) and (69) cancel out. Since by Lemma A.31 for every $j_1,j_2 \in [\ell]$ and $k_1,k_2 > 0$:

$$
\mathrm{E}\left[\mathcal{X}_{i,k_1}^{(j_1)}\mathcal{X}_{i,k_2}^{(j_2)}\right] = \mathrm{Cov}\left[\mathcal{X}_{i,k_1}^{(j_1)}, \mathcal{X}_{i,k_2}^{(j_2)}\right] = \mathrm{Cov}\left[\mathcal{G}_{i,k_1}^{(j_1)}, \mathcal{G}_{i,k_2}^{(j_2)}\right] = \mathrm{E}\left[\mathcal{G}_{i,k_1}^{(j_1)}\mathcal{G}_{i,k_2}^{(j_2)}\right],
$$

and since all the other terms are independent of coordinate i, we have

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j_1,j_2\in[\ell]}\sum_{k_1,k_2>0} \mathcal{X}_{i,k_1}^{(j_1)}\mathcal{X}_{i,k_2}^{(j_2)} B_{k_1}^{(j_1)} B_{k_2}^{(j_2)} \frac{\partial^2}{\partial A^{(j_1)}\partial A^{(j_2)}}\Psi(\overline{A})\right] \\
&= \mathrm{E}\left[\sum_{j_1,j_2\in[\ell]}\sum_{k_1,k_2>0} \mathcal{G}_{i,k_1}^{(j_1)}\mathcal{G}_{i,k_2}^{(j_2)} B_{k_1}^{(j_1)} B_{k_2}^{(j_2)} \frac{\partial^2}{\partial A^{(j_1)}\partial A^{(j_2)}}\Psi(\overline{A})\right].
\end{aligned}
$$

Therefore, by (68), (69) and (70),

$$
\begin{aligned}
\left|\mathrm{E}\left[\Psi(\overline{P}(\overline{\mathcal{T}})) - \Psi(\overline{P}(\overline{\mathcal{U}}))\right]\right| &\leq \mathrm{E}[|R_{\underline{\mathcal{T}}}|] + \mathrm{E}[|R_{\underline{\mathcal{U}}}|] \\
&\leq \frac{\ell^{3/2}B}{3}\left(\frac{8}{\sqrt{\alpha}}\right)^d \left(\sum_{j=1}^\ell \mathrm{Inf}_i(P^{(j)})\right)^{3/2},
\end{aligned}
$$

as claimed. $\qquad\square$

Proof of Theorem A.45. Recall that $\sum_{j=1}^{\ell} \mathrm{Inf}_i(P^{(j)}) \leq \tau$ and that $\mathrm{Var}[P^{(j)}] \leq 1$. By Claim A.47, Lemma A.50 and Claim A.21,

$$
\begin{aligned}
\left| \mathrm{E}\left[\Psi(\overline{P}(\underline{\overline{\mathbb{X}}})) - \Psi(\overline{P}(\underline{\overline{\mathbb{G}}})) \right] \right| &\leq \sum_{i=1}^{n} \left| \mathrm{E}\left[\Psi(\overline{P}(\underline{\overline{\mathbb{U}}}_{(i-1)})) - \Psi(\overline{P}(\underline{\overline{\mathbb{U}}}_{(i)})) \right] \right| \\
&\leq \frac{\ell^{3/2} B}{3} \left(\frac{8}{\sqrt{\alpha}} \right)^d \sum_{i=1}^{n} \left(\sum_{j=1}^{\ell} \mathrm{Inf}_i(P^{(j)}) \right)^{3/2} \\
&\leq \frac{\ell^{3/2} B}{3} \left(\frac{8}{\sqrt{\alpha}} \right)^d \sqrt{\tau} \sum_{i=1}^{n} \sum_{j=1}^{\ell} \mathrm{Inf}_i(P^{(j)}) \\
&= \frac{\ell^{3/2} B}{3} \left(\frac{8}{\sqrt{\alpha}} \right)^d \sqrt{\tau} \sum_{j=1}^{\ell} \mathrm{Inf}(P^{(j)}) \leq \frac{\ell^{5/2} dB}{3} \left(\frac{8}{\sqrt{\alpha}} \right)^d \sqrt{\tau} .
\end{aligned}
$$

$\square$

A.6 A tailored application of invariance principle

Definition A.51. Define $\phi : \mathbb{R} \to \mathbb{R}$ as

$$
\phi(x) := \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } x \in (0,1), \\ 1 & \text{if } x \geq 1, \end{cases}
$$

and $\chi : \mathbb{R}^\ell \to \mathbb{R}$ as $\chi(\overline{x}) := \prod_{j=1}^{\ell} \phi(x^{(j)})$.

Definition A.52. Let P be a multilinear polynomial and $\gamma \in [0,1]$. We say that P is *γ-decaying* if for each $d \in \mathbb{N}$ we have

$$
\mathrm{E}\left[\left(P^{\geq d} \right)^2 \right] \leq (1 - \gamma)^d .
$$

We also say that a tuple of multilinear polynomials $\overline{P} = (P^{(1)}, \ldots, P^{(\ell)})$ is γ-decaying if $P^{(j)}$ is γ-decaying for each $j \in [\ell]$.

Note that if a multilinear polynomial P is γ-decaying, then, in particular, $\mathrm{Var}[P] \leq \mathrm{E}[P^2] \leq 1$.

Our goal in this section is to prove a version of invariance principle for γ-decaying multilinear polynomials and the function χ:

Theorem A.53. *Let $(\overline{X}, \underline{\overline{\mathbb{X}}}, \overline{\mathbb{G}})$ be an ensemble collection for a probability space $(\overline{\Omega}, \mathbb{P})$ with $\min_{x \in \Omega} \pi(x) \geq \alpha$, $\alpha \in (0, 1/2]$.*

Let $\overline{P} = (P^{(1)}, \ldots, P^{(\ell)})$ be such that $P^{(j)}$ is a multilinear polynomial compatible with the ensemble sequence $\underline{\mathbb{X}}^{(j)}$.

Let $\gamma \in [0,1]$, $\tau \in (0,1]$ and assume that $\overline{P}$ is γ-decaying and that $\sum_{j=1}^{\ell} \mathrm{Inf}_i(P^{(j)}) \leq \tau$ for each $i \in [n]$. There exists an absolute constant $C \geq 0$ such that

$$
\left| \mathrm{E}\left[\chi(\overline{P}(\underline{\overline{\mathbb{X}}})) - \chi(\overline{P}(\overline{\mathbb{G}})) \right] \right| \leq C \ell^{5/2} \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} .
$$

Two obstacles to proving Theorem A.53 by direct application of Theorem A.45 are:

1. The function χ is not $\mathcal{C}^3$.

2. A γ-decaying multilinear polynomial does not have bounded degree.

We will deal with those problems in turn.

A.6.1 Approximating χ with a $\mathcal{C}^3$ function

To apply Theorem A.45, we are going to approximate ϕ and χ with $\mathcal{C}^3$ (in fact, $\mathcal{C}^\infty$) functions.

For that we need to introduce the notion of convolution and a basic calculus theorem, whose proof we omit (see, e.g., Chapter 9 in [Rud87]):

Definition A.54. Let $f : \mathbb{R} \to \mathbb{R}$ and $S \subseteq \mathbb{R}$. We say that S is a *support* of f if $x \notin S$ implies $f(x) = 0$. We say that f has *compact support* if there exists a bounded interval I that is a support of f.

Definition A.55. The convolution $f * g$ of two continuous functions $f, g : \mathbb{R} \to \mathbb{R}$, at least one of which has compact support, is $(f * g)(x) := \int_{-\infty}^{\infty} f(x - t)g(t)\,dt$.

Theorem A.56. *Let functions $f, g : \mathbb{R} \to \mathbb{R}$ be such that f is continuous on $\mathbb{R}$, $g \in \mathcal{C}^\infty$ and g has compact support. Then, $(f * g) \in \mathcal{C}^\infty$. Furthermore, for every $k \in \mathbb{N}$ and $x \in \mathbb{R}$:*

$$\frac{\partial^k}{\partial x^k}(f * g)(x) = \left(f * \frac{\partial^k g}{\partial x^k} \right)(x).$$

We also need a special density function with support $[-1, 1]$:

Theorem A.57. *There exists a function $\psi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that all of the following hold:*

- $\psi \in \mathcal{C}^\infty$.

- ψ *has support* $[-1, 1]$.

- $\forall x : \psi(x) = \psi(-x)$.

- $\int_{-\infty}^{\infty} \psi(x)\,dx = \int_{-1}^{1} \psi(x)\,dx = 1$.

Proof. Consider

$$\Psi(x) := \begin{cases} \exp(-\frac{1}{(x+1)^2}) \cdot \exp(-\frac{1}{(x-1)^2}) & \text{if } x \in (-1, 1) \\ 0 & \text{otherwise} \end{cases} \tag{73}$$

and set $\psi(x) := \Psi(x)/c$ where $c := \int_{-1}^{1} \Psi(x)\,dx$. $\qquad\square$

For any $\lambda > 0$ we can rescale ψ to an analogous distribution with support $[-\lambda, \lambda]$:

Definition A.58. Let $\lambda > 0$ and define $\psi_\lambda : \mathbb{R} \to \mathbb{R}_{\geq 0}$ as $\psi_\lambda(x) := \frac{1}{\lambda} \psi\left(\frac{x}{\lambda}\right)$.

It is easy to see that ψ_λ has properties analogous to ψ:

Claim A.59. *Let $\lambda > 0$. ψ_λ has the following properties:*

- $\psi_\lambda \in \mathcal{C}^\infty$.

- ψ_λ *has support $[-\lambda, \lambda]$.*

- $\forall x : \psi_\lambda(x) = \psi_\lambda(-x)$.

- $\int_{-\infty}^{\infty} \psi_\lambda(x)\,\mathrm{d}x = \int_{-\lambda}^{\lambda} \psi_\lambda(x)\,\mathrm{d}x = 1$.

We see that convoluting ϕ with ψ_λ for a small λ results in a smooth function that is still very close to ϕ:

Definition A.60. Let $\lambda \in (0, 1/2)$ and define $\phi_\lambda : \mathbb{R} \to \mathbb{R}$ as $\phi_\lambda := \phi * \psi_\lambda$.

To start with, we state some easy to verify properties of ϕ_λ:

Claim A.61. *Let $\lambda \in (0, 1/2)$. The function ϕ_λ has the following properties:*

- $\phi_\lambda(x) = \int_{-\lambda}^{\lambda} \psi_\lambda(y)\phi(x+y)\,\mathrm{d}y$.

- $x \leq -\lambda \vee x \in [\lambda, 1-\lambda] \vee x \geq 1 + \lambda \implies \phi_\lambda(x) = \phi(x)$.

- $x \in [-\lambda, \lambda] \implies \phi_\lambda(x) \in [0, \lambda]$.

- $x \in [1-\lambda, 1+\lambda] \implies \phi_\lambda(x) \in [1-\lambda, 1]$.

- $x \leq y \implies \phi_\lambda(x) \leq \phi_\lambda(y)$.

Lemma A.62. *Let $\lambda \in (0, 1/2)$:*

1) $\forall x : |\phi_\lambda(x) - \phi(x)| \leq \lambda$.

2) $\phi_\lambda \in \mathcal{C}_\infty$. *Furthermore, for each $k \in \mathbb{N}$ there exists a constant $B_k \geq 0$ such that $\forall x : \left| \frac{\partial^k}{\partial x^k} \phi_\lambda(x) \right| \leq \frac{B_k}{\lambda^k}$.*

Proof. 1) From Claim A.61.

2) Since $\phi_\lambda = \phi * \psi_\lambda$, due to Theorem A.56 we have $\phi_\lambda \in \mathcal{C}^\infty$.

For $x \notin [-\lambda, 1+\lambda]$ the function ϕ_λ is constant with $\left| \frac{\partial^k}{\partial x^k} \phi_\lambda(x) \right| \leq 1$.

For $x \in [-\lambda, 1+\lambda]$, first note that for every $k \in \mathbb{N}$, since ψ has support $[-1, 1]$, also all of its derivatives have support $[-1, 1]$ and therefore $\left| \frac{\partial^k}{\partial x^k} \psi(x) \right| \leq B_k$. Together with Theorem A.56 this gives (substituting $z := y/\lambda$)

$$
\begin{aligned}
\left| \frac{\partial^k}{\partial x^k} \phi_\lambda(x) \right| &= \left| \frac{\partial^k}{\partial x^k} (\phi * \psi_\lambda)(x) \right| = \left| \int_{-\infty}^{+\infty} \phi(x-y) \frac{\partial^k}{\partial y^k} \psi_\lambda(y)\,\mathrm{d}y \right| \\
&= \left| \int_{-\lambda}^{\lambda} \phi(x-y) \frac{\partial^k}{\partial y^k} \psi_\lambda(y)\,\mathrm{d}y \right| \\
&= \frac{1}{\lambda^{k+1}} \left| \int_{-\lambda}^{\lambda} \phi(x-y) \frac{\partial^k}{\partial z^k} \psi(z)\,\mathrm{d}y \right| \leq \frac{2B_k}{\lambda^k} \ ,
\end{aligned}
$$

as claimed.

$\square$

Now we are ready for the approximation of χ:

Definition A.63. Let $\lambda \in (0, 1/2)$. Define function $\chi_\lambda : \mathbb{R}^\ell \to \mathbb{R}$ as

$$\chi_\lambda(\bar{x}) := \prod_{j=1}^{\ell} \phi_\lambda(x^{(j)}) .$$

From Lemma A.62 we easily get:

Corollary A.64. *Let $\lambda \in (0, 1/2)$. The function χ_λ has the following properties:*

1) $\forall \bar{x} \in \mathbb{R}^\ell : |\chi(\bar{x}) - \chi_\lambda(\bar{x})| \leq \ell\lambda$.

2) *There exists a universal constant $B \geq 0$ such that χ_λ is $\frac{B}{\lambda^3}$-smooth.*

After developing the approximation we are ready to prove the invariance principle for the function χ:

Theorem A.65. *Let $(\overline{X}, \overline{\mathfrak{X}}, \overline{\mathfrak{G}})$ be an ensemble collection for a probability space $(\overline{\Omega}, \mathbb{P})$ with $\min_{x \in \Omega} \pi(x) \geq \alpha > 0$.*

Let $\overline{P} = (P^{(1)}, \ldots, P^{(\ell)})$ be such that $P^{(j)}$ is a multilinear polynomial compatible with the ensemble sequence $\underline{\mathfrak{X}}^{(j)}$.

Let $d \in \mathbb{N}$ and $\tau \in [0, 1]$ and assume that $\deg(P^{(j)}) \leq d$ and $\mathrm{Var}[P^{(j)}] \leq 1$ for each $j \in [\ell]$, and that $\sum_{j=1}^{\ell} \mathrm{Inf}_i(P^{(j)}) \leq \tau$ for each $i \in [n]$.

There exists a universal constant $C \geq 0$ such that

$$\left| \mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{X}})) - \chi(\overline{P}(\underline{\mathfrak{G}})) \right] \right| \leq C \cdot \frac{\ell^{5/2} \tau^{1/8}}{\alpha^{4d}} .$$

Proof. Let $\lambda := \tau^{1/8}/3$. By the triangle inequality we get

$$\begin{aligned}
\left| \mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{X}})) - \chi(\overline{P}(\underline{\mathfrak{G}})) \right] \right| &\leq \left| \mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{X}})) - \chi_\lambda(\overline{P}(\underline{\mathfrak{X}})) \right] \right| \\
&\quad + \left| \mathrm{E}\left[\chi_\lambda(\overline{P}(\underline{\mathfrak{X}})) - \chi_\lambda(\overline{P}(\underline{\mathfrak{G}})) \right] \right| \\
&\quad + \left| \mathrm{E}\left[\chi_\lambda(\overline{P}(\underline{\mathfrak{G}})) - \chi(\overline{P}(\underline{\mathfrak{G}})) \right] \right| .
\end{aligned} \tag{74}$$

From Corollary A.64.1 and the definition of λ we get both

$$\left| \mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{X}})) - \chi_\lambda(\overline{P}(\underline{\mathfrak{X}})) \right] \right| \leq \ell\lambda \leq O\left(\frac{\ell^{5/2} \tau^{1/8}}{\alpha^{4d}} \right) \tag{75}$$

$$\left| \mathrm{E}\left[\chi_\lambda(\overline{P}(\underline{\mathfrak{G}})) - \chi(\overline{P}(\underline{\mathfrak{G}})) \right] \right| \leq \ell\lambda \leq O\left(\frac{\ell^{5/2} \tau^{1/8}}{\alpha^{4d}} \right) . \tag{76}$$

By Theorem A.45 and Corollary A.64.2 we get

$$\left| \mathrm{E}\left[\chi_\lambda(\overline{P}(\underline{\mathfrak{X}})) - \chi_\lambda(\overline{P}(\underline{\mathfrak{G}})) \right] \right| \leq O\left(\frac{\ell^{5/2} d8^d \tau^{1/2}}{\lambda^3 \alpha^{d/2}} \right) . \tag{77}$$

We can assume w.l.o.g. that $\alpha \leq 1/2$ (otherwise the theorem is trivial). Using the definition of λ, $d8^d \leq 9^{d+1}$ and $9 \leq \left(\frac{1}{\alpha}\right)^{3.5}$ we see that

$$\frac{\ell^{5/2}d8^d\tau^{1/2}}{\lambda^3\alpha^{d/2}} \leq O\left(\frac{\ell^{5/2}d8^d\tau^{1/8}}{\alpha^{d/2}}\right) \leq O\left(\frac{\ell^{5/2}\tau^{1/8}}{\alpha^{4d}}\right). \tag{78}$$

Inserting (75), (76), and the combination of (78) and (77) into (74) gives the result. □

A.6.2 Invariance principle for γ-decaying polynomials

Let $\overline{P} = (P^{(1)}, \ldots, P^{(\ell)})$ be a tuple of mutlilinear polynomials and let $\overline{P}^{<d} := \left(\left(P^{(1)}\right)^{<d}, \ldots, \left(P^{(\ell)}\right)^{<d}\right)$.
We will deal with a γ-decaying $\overline{P}$ by estimating $|\mathrm{E}[\chi(\overline{P}^{<d}(\underline{\mathfrak{X}})) - \chi(\overline{P}(\underline{\mathfrak{X}}))]|$ for appropriately chosen d.

First, we need a bound on the change of χ:

Lemma A.66. *For all $\overline{x} = (x^{(1)}, \ldots, x^{(\ell)}), \overline{\varepsilon} = (\varepsilon^{(1)}, \ldots, \varepsilon^{(\ell)}) \in \mathbb{R}^\ell$:*

$$\left|\chi(x^{(1)} + \varepsilon^{(1)}, \ldots, x^{(\ell)} + \varepsilon^{(\ell)}) - \chi(x^{(1)}, \ldots, x^{(\ell)})\right| \leq \sum_{j=1}^{\ell} |\varepsilon^{(j)}|.$$

Proof. Letting $\overline{y}_{(j)} := (x^{(1)}, \ldots, x^{(j)}, x^{(j+1)} + \varepsilon^{(j+1)}, \ldots x^{(\ell)} + \varepsilon^{(\ell)})$,

$$\left|\chi(x^{(1)} + \varepsilon^{(1)}, \ldots, x^{(\ell)} + \varepsilon^{(\ell)}) - \chi(x^{(1)}, \ldots, x^{(\ell)})\right|$$
$$\leq \sum_{j=1}^{\ell} \left|\chi(\overline{y}_{(j-1)}) - \chi(\overline{y}_{(j)})\right| \leq \sum_{j=1}^{\ell} \left|\varepsilon^{(j)}\right|.$$

□

Proof of Theorem A.53. Let $d := \lfloor\frac{\ln 1/\tau}{64\ln 1/\alpha}\rfloor$. By the triangle inequality,

$$\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{X}})) - \chi(\overline{P}(\underline{\mathfrak{G}}))\right]\right| \leq \left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{X}})) - \chi(\overline{P}^{<d}(\underline{\mathfrak{X}}))\right]\right|$$
$$+ \left|\mathrm{E}\left[\chi(\overline{P}^{<d}(\underline{\mathfrak{X}})) - \chi(\overline{P}^{<d}(\underline{\mathfrak{G}}))\right]\right|$$
$$+ \left|\mathrm{E}\left[\chi(\overline{P}^{<d}(\underline{\mathfrak{G}})) - \chi(\overline{P}(\underline{\mathfrak{G}}))\right]\right|. \tag{79}$$

We proceed to demonstrate that all three terms on the right hand side of (79) are $O\left(\ell^4\tau^{\Omega\left(\frac{\gamma}{\ln 1/\alpha}\right)}\right)$, which will finish the proof.

Lemma A.67.

$$\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{X}})) - \chi(\overline{P}^{<d}(\underline{\mathfrak{X}}))\right]\right| \leq \ell(1-\gamma)^{d/2} \leq O\left(\ell\tau^{\Omega\left(\frac{\gamma}{\ln 1/\alpha}\right)}\right) \tag{80}$$

and, similarly,

$$\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{G}})) - \chi(\overline{P}^{<d}(\underline{\mathfrak{G}}))\right]\right| \leq \ell(1-\gamma)^{d/2} \leq O\left(\ell\tau^{\Omega\left(\frac{\gamma}{\ln 1/\alpha}\right)}\right) \tag{81}$$

Proof. We prove only (80), the argument for (81) being the same. Using Lemma A.66, Cauchy-Schwarz, the fact that $\overline{P}$ is γ-decaying and the definition of d,

$$\left| \mathrm{E}\left[\chi(\overline{P}(\underline{\mathfrak{X}})) - \chi(\overline{P}^{<d}(\underline{\mathfrak{X}})) \right] \right| \leq \sum_{j=1}^{\ell} \mathrm{E}\left[\left| \left(P^{(j)} \right)^{\geq d} (\underline{\mathfrak{X}}^{(j)}) \right| \right]$$

$$\leq \sum_{j=1}^{\ell} \sqrt{\mathrm{E}\left[\left(\left(P^{(j)} \right)^{\geq d} \right)^2 \right]} \leq \ell (1-\gamma)^{d/2} \leq 2\ell \tau^{\frac{\gamma}{128 \ln 1/\alpha}} \ .$$

$\square$

Lemma A.68.

$$\left| \mathrm{E}\left[\chi(\overline{P}^{<d}(\underline{\mathfrak{X}})) - \chi(\overline{P}^{<d}(\underline{\mathfrak{G}})) \right] \right| \leq O\left(\ell^{5/2} \tau^{\Omega\left(\frac{\gamma}{\ln 1/\alpha} \right)} \right) \ .$$

Proof. From Theorem A.65,

$$\left| \mathrm{E}\left[\chi(\overline{P}^{<d}(\underline{\mathfrak{X}})) - \chi(\overline{P}^{<d}(\underline{\mathfrak{G}})) \right] \right| \leq O\left(\frac{\ell^{5/2} \tau^{1/8}}{\alpha^{4d}} \right) \ .$$

From the definition of d (recall that $\alpha \leq 1/2$),

$$\frac{\ell^{5/2} \tau^{1/8}}{\alpha^{4d}} \leq \ell^{5/2} \tau^{1/16} \leq \ell^{5/2} \tau^{\Omega\left(\frac{\gamma}{\ln 1/\alpha} \right)} \ ,$$

as claimed. $\square$

This finishes the proof of Theorem A.53.

A.7 Reduction to the γ-decaying case

To apply Theorem A.53 we need to show that "smoothing out" of multilinear polynomials $P^{(1)}, \ldots, P^{(\ell)}$ does not change the expectation of their product too much.

Recall Definitions A.22 and A.28 for the operator T_ρ. Our goal in this section is to prove:

Theorem A.69. *Let $\overline{X}$ be a random vector distributed according to $(\overline{\Omega}, \mathcal{P})$ with $\rho(\overline{\Omega}, \mathcal{P}) \leq \rho \leq 1$. Let $\underline{\mathfrak{Z}}$ be an ensemble sequence constructed from $\overline{X}$ and $\underline{\mathfrak{X}}^{(1)}, \ldots, \underline{\mathfrak{X}}^{(\ell)}$ be ensemble sequences constructed from $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}$, respectively.*

Let $\varepsilon \in (0, 1/2]$ and $\gamma \in \left[0, \frac{(1-\rho)\varepsilon}{\ell \ln \ell/\varepsilon} \right]$.

Then, for all multilinear polynomials $P^{(1)}, \ldots, P^{(\ell)}$ such that $P^{(j)}(\underline{\mathfrak{X}}^{(j)}) \in [0, 1]$:

$$\left| \mathrm{E}\left[\prod_{j=1}^{\ell} P^{(j)}(\underline{\mathfrak{X}}^{(j)}) - \prod_{j=1}^{\ell} T_{1-\gamma} P^{(j)}(\underline{\mathfrak{X}}^{(j)}) \right] \right| \leq \varepsilon \ .$$

Let us start with an intuition: Due to Lemma A.18, it is enough to bound

$$
\mathrm{E}\left[\prod_{j=1}^{\ell} P_S^{(j)} - \prod_{j=1}^{\ell} T_{1-\gamma} P_S^{(j)}\right]
$$

for every $S \subseteq [n]$. If $|S|$ is small, we use the fact that $P_S^{(j)} - T_{1-\gamma} P_S^{(j)}$ shrinks by a factor of $1 - (1 - \gamma)^{|S|}$ for every j. If $|S|$ is large, we exploit that both

$$
\mathrm{E}\left[\prod_{j=1}^{\ell} P_S^{(j)}\right], \mathrm{E}\left[\prod_{j=1}^{\ell} T_{1-\gamma} P_S^{(j)}\right]
$$

are small (roughly $\rho^{|S|}$ times smaller compared to their variances).

To give a formal argument, we use yet another ensemble sequence: let $j \in [\ell]$. We define $\underline{\mathcal{Y}}^{(j)}$ to be an ensemble sequence constructed from $\underline{X}^{[\ell] \setminus \{j\}}$. Furthermore, let

$$
A^{(j)} := \prod_{j' < j} T_{1-\gamma} P(\underline{\mathcal{X}}^{(j')}) \prod_{j' > j} P(\underline{\mathcal{X}}^{(j')}) .
$$

Note that since $A^{(j)} \in L^2(\underline{X}^{[\ell] \setminus \{j\}})$, there exists a multilinear polynomial $Q^{(j)}$ compatible with $\underline{\mathcal{Y}}^{(j)}$ such that

$$
A^{(j)} = Q^{(j)}(\underline{\mathcal{Y}}^{(j)}) .
$$

Lemma A.70.

$$
\prod_{j=1}^{\ell} P^{(j)}(\underline{\mathcal{X}}^{(j)}) - \prod_{j=1}^{\ell} T_{1-\gamma} P^{(j)}(\underline{\mathcal{X}}^{(j)}) = \sum_{j=1}^{\ell} (\mathrm{Id} - T_{1-\gamma}) P^{(j)}(\underline{\mathcal{X}}^{(j)}) \cdot Q^{(j)}(\underline{\mathcal{Y}}^{(j)}) .
$$

Proof. By definition of $Q^{(j)}$. $\qquad\square$

Lemma A.71. *For every $j \in [\ell]$ and $S \subseteq [n]$, $S \neq \emptyset$:*

$$
\left|\mathrm{E}\left[P_S^{(j)}(\underline{\mathcal{X}}^{(j)}) \cdot Q_S^{(j)}(\underline{\mathcal{Y}}^{(j)})\right]\right| \le \rho^{|S|} \sqrt{\mathrm{Var}[P_S^{(j)}] \, \mathrm{Var}[Q_S^{(j)}]} .
$$

Proof. For ease of notation let us write $P := P^{(j)}$, $Q := Q^{(j)}$, $\underline{\mathcal{X}} := \underline{\mathcal{X}}^{(j)}$ and $\underline{\mathcal{Y}} := \underline{\mathcal{Y}}^{(j)}$.

Let $P(\underline{\mathcal{X}}) = \sum_\sigma \alpha(\sigma) \mathcal{X}_\sigma$ and $Q(\underline{\mathcal{Y}}) = \sum_\sigma \beta(\sigma) \mathcal{Y}_\sigma$.

We know that $\mathcal{X}_{i,k} \in L^2(X_i^{(j)})$ and $\mathcal{Y}_{i,k} \in L^2(X_i^{([\ell] \setminus \{j\})})$ for every $i \in [n]$, $k, k' \ge 0$. Furthermore, if $k, k' > 0$, then $\mathrm{E}[\mathcal{X}_{i,k}] = \mathrm{E}[\mathcal{Y}_{i,k'}] = 0$ and $\mathrm{Var}[\mathcal{X}_{i,k}] = \mathrm{Var}[\mathcal{Y}_{i,k'}] = 1$. By definition of ρ, this implies

$$
\left|\mathrm{E}\left[\mathcal{X}_{i,k} \cdot \mathcal{Y}_{i,k'}\right]\right| = \left|\mathrm{Cov}\left[\mathcal{X}_{i,k}, \mathcal{Y}_{i,k'}\right]\right| \le \rho. \tag{82}
$$

Expanding the expectation and using (82) and Cauchy-Schwarz,

$$
|\mathrm{E}\left[P_S(\underline{\mathfrak{X}})Q_S(\underline{\mathfrak{Y}})\right]| = \left|\mathrm{E}\left[\left(\sum_{\sigma:\mathrm{supp}(\sigma)=S}\alpha(\sigma)\mathfrak{X}_\sigma\right)\left(\sum_{\sigma':\mathrm{supp}(\sigma')=S}\beta(\sigma')\mathfrak{Y}_{\sigma'}\right)\right]\right|
$$

$$
\leq \sum_{\substack{\sigma,\sigma':\\ \mathrm{supp}(\sigma)=\mathrm{supp}(\sigma')=S}}\left|\alpha(\sigma)\beta(\sigma')\prod_{i\in S}\mathrm{E}\left[\mathfrak{X}_{i,\sigma_i}\mathfrak{Y}_{i,\sigma'_i}\right]\right|
$$

$$
\leq \rho^{|S|}\sum_{\substack{\sigma,\sigma':\\ \mathrm{supp}(\sigma)=\mathrm{supp}(\sigma')=S}}|\alpha(\sigma)\beta(\sigma')|
$$

$$
\leq \rho^{|S|}\sqrt{\mathrm{Var}[P_S]\,\mathrm{Var}[Q_S]}\,,
$$

$\square$

Lemma A.72. *Let $k \in \mathbb{N}$. Then,* $\min(1-(1-\gamma)^k,\rho^k) \leq \varepsilon/\ell$.

Proof. If $\rho \in \{0,1\}$ we are done, therefore assume that $\rho \in (0,1)$. If $k \geq \log_\rho \varepsilon/\ell$, then $\rho^k \leq \varepsilon/\ell$. If $0 \leq k < \log_\rho \varepsilon/\ell$, then by Bernoulli's inequality,

$$
1-(1-\gamma)^k \leq \gamma k \leq \frac{1-\rho}{\ln(1/\rho)}\cdot\frac{\varepsilon}{\ell} \leq \frac{\varepsilon}{\ell}\,.
$$

$\square$

Lemma A.73. *For every $j \in [\ell]$ and $S \subseteq [n]$, $S \neq \emptyset$:*

$$
\left|\mathrm{E}\left[(\mathrm{Id}-T_{1-\gamma})P_S^{(j)}(\underline{\mathfrak{X}}^{(j)})\cdot Q_S^{(j)}(\underline{\mathfrak{Y}}^{(j)})\right]\right| \leq \frac{\varepsilon}{\ell}\cdot\sqrt{\mathrm{Var}[P_S^{(j)}]\,\mathrm{Var}[Q_S^{(j)}]}\,.
$$

Proof. As in the proof of Lemma A.71, we will write $P := P^{(j)}$, $Q := Q^{(j)}$, $\underline{\mathfrak{X}} := \underline{\mathfrak{X}}^{(j)}$ and $\underline{\mathfrak{Y}} := \underline{\mathfrak{Y}}^{(j)}$. By definition of $T_{1-\gamma}$,

$$
(\mathrm{Id}-T_{1-\gamma})P_S(\underline{\mathfrak{X}}) = (1-(1-\gamma)^{|S|})P_S(\underline{\mathfrak{X}})\,. \tag{83}
$$

From (83), Lemma A.71 and Lemma A.72,

$$
\left|\mathrm{E}\left[(\mathrm{Id}-T_{1-\gamma})P_S(\underline{\mathfrak{X}})\cdot Q_S(\underline{\mathfrak{Y}})\right]\right| \leq \min\left(1-(1-\gamma)^{|S|},\rho^{|S|}\right)\sqrt{\mathrm{Var}[P_S]\,\mathrm{Var}[Q_S]}
$$

$$
\leq \frac{\varepsilon}{\ell}\sqrt{\mathrm{Var}[P_S]\,\mathrm{Var}[Q_S]}\,.
$$

$\square$

Lemma A.74. *Fix $j \in [\ell]$. Then,*

$$
\left|\mathrm{E}\left[(\mathrm{Id}-T_{1-\gamma})P^{(j)}(\underline{\mathfrak{X}}^{(j)})\cdot Q^{(j)}(\underline{\mathfrak{Y}}^{(j)})\right]\right| \leq \varepsilon/\ell\,.
$$

Proof. For ease of notation write $P := P^{(j)}$, $Q := Q^{(j)}$, $\underline{\mathcal{X}} := \underline{\mathcal{X}}^{(j)}$ and $\underline{\mathcal{Y}} := \underline{\mathcal{Y}}^{(j)}$.

Observe that since $P(\underline{\mathcal{X}}), Q(\underline{\mathcal{Y}}) \in [0,1]$, also $\text{Var}[P], \text{Var}[Q] \le 1$.

From Lemma A.18, Lemma A.73 and Cauchy-Schwarz,

$$\left| \text{E}\left[(\text{Id} - T_{1-\gamma}) P(\underline{\mathcal{X}}) \cdot Q(\underline{\mathcal{Y}}) \right] \right| \le \sum_{S \subseteq [n]} \left| \text{E}\left[(\text{Id} - T_{1-\gamma}) P_S(\underline{\mathcal{X}}) \cdot Q_S(\underline{\mathcal{Y}}) \right] \right|$$

$$\le \frac{\varepsilon}{\ell} \sum_{S \ne \emptyset} \sqrt{\text{Var}[P_S] \text{Var}[Q_S]}$$

$$\le \frac{\varepsilon}{\ell} \sqrt{\text{Var}[P] \text{Var}[Q]} \le \varepsilon/\ell \,.$$

$\square$

Proof of Theorem A.69. By Lemma A.70 and Lemma A.74,

$$\left| \text{E}\left[\prod_{j=1}^{\ell} P^{(j)}(\underline{\mathcal{X}}^{(j)}) - \prod_{j=1}^{\ell} T_{1-\gamma} P^{(j)}(\underline{\mathcal{X}}^{(j)}) \right] \right| \le \sum_{j=1}^{\ell} \left| \text{E}\left[(\text{Id} - T_{1-\gamma}) P^{(j)}(\underline{\mathcal{X}}^{(j)}) \cdot Q^{(j)}(\underline{\mathcal{Y}}^{(j)}) \right] \right|$$

$$\le \varepsilon \,.$$

$\square$

A.8 Gaussian reverse hypercontractivity

Definition A.75. Let $L^2(\mathbb{R}^n, \gamma^n)$ be the inner product space of functions with standard $\mathcal{N}(0,1)$ Gaussian measure.

Our goal in this section is to prove the following bound:

Theorem A.76. *Let $(\overline{X}, \underline{\overline{X}}, \overline{\mathcal{G}})$ be an ensemble collection for a probability space $(\overline{\underline{\Omega}}, \underline{\mathcal{P}})$ with $\rho(\mathcal{P}) \le \rho < 1$ and such that each orthonormal ensemble in $\overline{\mathcal{G}}$ has size p.*

Then, for all $f^{(1)}, \dots, f^{(\ell)} \in L^2(\mathbb{R}^{pn}, \gamma^{pn})$ such that $f^{(1)}, \dots, f^{(\ell)} : \mathbb{R}^{pn} \to [0,1]$ and $\text{E}\left[f^{(j)}(\underline{\mathcal{G}}^{(j)}) \right] = \mu^{(j)}$:

$$\text{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{\mathcal{G}}^{(j)}) \right] \ge \left(\prod_{j=1}^{\ell} \mu^{(j)} \right)^{\ell/(1-\rho^2)} \,.$$

Remark A.77. Since the random variables $\mathcal{G}_{i,0}^{(j)}$ are constant, it suffices to consider consider $f^{(j)}$ as functions of pn rather than $(p+1)n$ inputs.

In order to prove Theorem A.76, we will use a multidimensional version of Gaussian reverse hypercontractivity stated as Theorem 1 in [CDP15] (cf. also Corollary 4 in [Led14]).

Theorem A.78 ([CDP15]). *Let $p > 0$ and let $\overline{G} = (\underline{G}^{(1)}, \dots, \underline{G}^{(\ell)})$ be a jointly Gaussian collection of ℓ random vectors such that:*

- *For each $j \in [\ell]$, $\underline{G}^{(j)} = (G_1^{(j)}, \ldots, G_n^{(j)})$ is a random vector distributed as n independent $\mathcal{N}(0,1)$ Gaussians.*

- *For every collection of real numbers $\{\alpha_i^{(j)}\} \in \mathbb{R}$:*

$$\mathrm{Var}\left[\sum_{i,j} \alpha_i^{(j)} \cdot G_i^{(j)}\right] \geq p \cdot \sum_{i,j} \left(\alpha_i^{(j)}\right)^2 . \tag{84}$$

Then, for all functions $f^{(1)}, \ldots, f^{(\ell)} \in L^2(\mathbb{R}^n, \gamma^n)$ such that $f^{(1)}, \ldots, f^{(\ell)} : \mathbb{R}^n \to [0,1]$ and $\mathrm{E}\left[f^{(j)}(\underline{G}^{(j)})\right] = \mu^{(j)}$:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{G}^{(j)})\right] \geq \left(\prod_{j=1}^{\ell} \mu^{(j)}\right)^{1/p} .$$

Remark A.79. An equivalent formulation of the condition in (84) is that the matrix $(T - p\,\mathrm{Id})$ is positive semidefinite, where T is the covariance matrix of $\underline{G}$.

To reduce Theorem A.76 to Theorem A.78 we first look at a single-coordinate variance bound for ensembles from $\overline{\mathcal{X}}$. Next, we will extend this bound to multiple coordinates and ensembles from $\overline{\mathcal{G}}$.

Lemma A.80. *Let $(\overline{X}, \overline{\mathcal{X}}, \overline{\mathcal{G}})$ be an ensemble collection for a probability space $(\overline{\Omega}, \mathcal{P})$ with $\rho(\mathcal{P}) \leq \rho < 1$ and such that each orthonormal ensemble in $\overline{\mathcal{X}}$ has size p.*

Fix $i \in [n]$ and for ease of notation let us write $\mathcal{X}^{(j)} = (\mathcal{X}_0^{(j)}, \ldots, \mathcal{X}_p^{(j)})$ for the random ensemble $\mathcal{X}_i^{(j)} = (\mathcal{X}_{i,0}^{(j)}, \ldots, \mathcal{X}_{i,p}^{(j)})$.

Then, for every collection of real numbers $\{\alpha_k^{(j)}\} \in \mathbb{R}$:

$$\mathrm{Var}\left[\sum_{j \geq 1, k > 0} \alpha_k^{(j)} \cdot \mathcal{X}_k^{(j)}\right] \geq \frac{1 - \rho^2}{\ell} \cdot \sum_{j \geq 1, k > 0} \left(\alpha_k^{(j)}\right)^2 .$$

Proof. For any $j \in [\ell]$ we define $A_j := \sum_{k>0} \alpha_k^{(j)} \cdot \mathcal{X}_k^{(j)}$ and $B_j := \sum_{j' \in [\ell] \setminus \{j\}} \sum_{k>0} \alpha_k^{(j')} \cdot \mathcal{X}_k^{(j')}$.
We compute

$$\begin{aligned}
\mathrm{Var}[B_j] \cdot \mathrm{Var}[A_j + B_j] &= \mathrm{Var}[A_j] \cdot \mathrm{Var}[B_j] + (\mathrm{Var}[B_j])^2 + 2\,\mathrm{Var}[B_j]\,\mathrm{Cov}[A_j, B_j] \\
&= \mathrm{Var}[A_j] \cdot \mathrm{Var}[B_j] + (\mathrm{Var}[B_j] + \mathrm{Cov}[A_j, B_j])^2 - \mathrm{Cov}[A_j, B_j]^2 \\
&\geq \mathrm{Var}[A_j] \cdot \mathrm{Var}[B_j] - \mathrm{Cov}[A_j, B_j]^2 \\
&\geq \mathrm{Var}[A_j]\,\mathrm{Var}[B_j](1 - \rho^2) ,
\end{aligned}$$

where in the last inequality we used that the definition of ρ implies

$$\left|\mathrm{Cov}[A_j, B_j]\right| \leq \rho \sqrt{\mathrm{Var}[A_j]\,\mathrm{Var}[B_j]}$$

since $A_j \in L^2(X_i^{(j)})$ and $B_i \in L^2(X_i^{([\ell] \setminus \{j\})})$.

Therefore,

$$
\begin{aligned}
\mathrm{Var}\left[\sum_{j \geq 1, k > 0} \alpha_k^{(j)} \cdot \mathcal{X}_k^{(j)}\right] &= \frac{1}{\ell} \sum_{j=1}^{\ell} \mathrm{Var}[A_j + B_j] \geq \frac{1 - \rho^2}{\ell} \sum_{j=1}^{\ell} \mathrm{Var}[A_j] \\
&= \frac{1 - \rho^2}{\ell} \sum_{j=1}^{\ell} \sum_{k > 0} \left(\alpha_j^{(k)}\right)^2.
\end{aligned}
$$
$\square$

Lemma A.81. *Let* $(\overline{X}, \overline{\mathcal{X}}, \overline{\mathcal{G}})$ *be an ensemble collection for a probability space* $(\overline{\Omega}, \mathcal{P})$ *with* $\rho(\mathcal{P}) \leq \rho < 1$. *Then, for every collection of real numbers* $\{\alpha_{i,k}^{(j)}\} \in \mathbb{R}$:

$$
\mathrm{Var}\left[\sum_{i,j \geq 1, k > 0} \alpha_{i,k}^{(j)} \cdot \mathcal{X}_{i,k}^{(j)}\right] \geq \frac{1 - \rho^2}{\ell} \cdot \sum_{i,j \geq 1, k > 0} \left(\alpha_{i,j}^{(k)}\right)^2.
$$

Proof. Since ensembles $\overline{\mathcal{X}}_i$ are independent, by Lemma A.80,

$$
\begin{aligned}
\mathrm{Var}\left[\sum_{i,j \geq 1, k > 0} \alpha_{i,k}^{(j)} \cdot \mathcal{X}_{i,k}^{(j)}\right] &= \sum_{i=1}^{n} \mathrm{Var}\left[\sum_{j \geq 1, k > 0} \alpha_{i,k}^{(j)} \cdot \mathcal{X}_{i,k}^{(j)}\right] \\
&\geq \frac{1 - \rho^2}{\ell} \cdot \sum_{i,j \geq 1, k > 0} \left(\alpha_{i,k}^{(j)}\right)^2.
\end{aligned}
$$

$\square$

Lemma A.82. *Let* $(\overline{X}, \overline{\mathcal{X}}, \overline{\mathcal{G}})$ *be an ensemble collection for a probability space* $(\overline{\Omega}, \mathcal{P})$ *with* $\rho(\mathcal{P}) \leq \rho < 1$. *Then, for every collection of real numbers* $\{\alpha_{i,k}^{(j)}\} \in \mathbb{R}$:

$$
\mathrm{Var}\left[\sum_{i,j \geq 1, k > 0} \alpha_{i,k}^{(j)} \cdot \mathcal{G}_{i,k}^{(j)}\right] \geq \frac{1 - \rho^2}{\ell} \cdot \sum_{i,j \geq 1, k > 0} \left(\alpha_{i,j}^{(k)}\right)^2.
$$

Proof. By Corollary A.31 and Lemma A.81. $\square$

Proof of Theorem A.76. By application of Theorem A.78 to $\overline{G} = (\underline{G}^{(1)}, \ldots, \underline{G}^{(\ell)})$, where $\underline{G}^{(j)} = (\mathcal{G}_{i,1}^{(j)}, \ldots, \mathcal{G}_{i,p}^{(j)}, \ldots, \mathcal{G}_{n,1}^{(j)}, \ldots, \mathcal{G}_{n,p}^{(j)})$.

Since $\underline{\mathcal{G}}^{(j)}$ is a Gaussian ensemble sequence, $\underline{G}^{(j)}$ is distributed as pn independent $\mathcal{N}(0, 1)$ Gaussians. Condition (84) for $p := \frac{1 - \rho^2}{\ell}$ is fulfilled due to Lemma A.82. $\square$

A.9 The main theorem

We recall the low-influence theorem that we want to prove:

Theorem 4.1. *Let $\overline{X}$ be a random vector distributed according to $(\overline{\Omega}, \mathcal{P})$ 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 $\varepsilon > 0$, there exists $\tau := \tau(\varepsilon, \rho, \alpha, \ell) > 0$ 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 $\mu^{(j)} := \mathrm{E}[f^{(j)}(\underline{X}^{(j)})]$:

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

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

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

We need to define some new objects in order to proceed with the proof. Let $(\overline{X}, \overline{\mathcal{X}}, \overline{\mathcal{G}})$ be an ensemble collection for $(\overline{\Omega}, \mathcal{P})$.

For $j \in [\ell]$, let $P^{(j)}$ be a multilinear polynomial compatible with $\underline{\mathcal{X}}^{(j)}$ and equivalent to $f^{(j)}(\underline{X}^{(j)})$. For some small $\gamma > 0$ to be fixed later let $Q^{(j)} := T_{1-\gamma} P^{(j)}$. Finally, letting p be the size of each of the ensembles $\mathcal{X}_i^{(j)}$ and $\mathcal{G}_i^{(j)}$, define a function $R^{(j)} : \mathbb{R}^{pn} \to \mathbb{R}$ as

$$R^{(j)}(\underline{x}) := \begin{cases} 0 & \text{if } Q^{(j)}(\underline{x}) < 0, \\ Q^{(j)}(\underline{x}) & \text{if } Q^{(j)}(\underline{x}) \in [0,1], \\ 1 & \text{if } Q^{(j)}(\underline{x}) > 1. \end{cases}$$

Note that it might be impossible to write $R^{(j)}$ as a multilinear polynomial, but it will not cause problems in the proof. Finally, let $\mu'^{(j)} := \mathrm{E}\left[R^{(j)}(\underline{\mathcal{G}}^{(j)})\right]$.

The proof proceeds by decomposing the expression we are bounding into several parts:

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

$$= \mathrm{E}\left[\prod_{j=1}^{\ell} P^{(j)}(\underline{\mathcal{X}}^{(j)}) - \prod_{j=1}^{\ell} Q^{(j)}(\underline{\mathcal{X}}^{(j)})\right] + \tag{85}$$

$$+ \mathrm{E}\left[\prod_{j=1}^{\ell} Q^{(j)}(\underline{\mathcal{X}}^{(j)}) - \prod_{j=1}^{\ell} R^{(j)}(\underline{\mathcal{G}}^{(j)})\right] + \tag{86}$$

$$+ \mathrm{E}\left[\prod_{j=1}^{\ell} R^{(j)}(\underline{\mathcal{G}}^{(j)})\right]. \tag{87}$$

We use the theorems proved so far to bound each of the terms (85), (86) and (87) in turn. First, we apply Theorem A.69 to show that (85) has small absolute value. Then, we use the invariance principle (Theorem A.53) to argue that (86) has small absolute value. Finally, using Gaussian reverse hypercontractivity (Theorem A.76) we show that (87) is bounded from below by (roughly) $\left(\prod_{j=1}^{\ell} \mu^{(j)}\right)^{\ell/(1-\rho^2)}$.

We proceed with a detailed argument in the following lemmas. In the following assume w.l.o.g that $\varepsilon \leq 1/2$ and $\alpha \leq 1/2$.

Lemma A.83. *Set* $\gamma := \frac{(1-\rho)\varepsilon}{2\ell \ln 2\ell/\varepsilon}$. *Then,*

$$\left| \mathrm{E}\left[\prod_{j=1}^{\ell} P^{(j)}(\underline{\mathfrak{X}}^{(j)}) - \prod_{j=1}^{\ell} Q^{(j)}(\underline{\mathfrak{X}}^{(j)}) \right] \right| \leq \varepsilon/2 .$$

Proof. By Theorem A.69. □

Lemma A.84. *There exists an absolute constant* $C > 0$ *such that*

$$\left| \mathrm{E}\left[\prod_{j=1}^{\ell} Q^{(j)}(\underline{\mathfrak{X}}^{(j)}) - \prod_{j=1}^{\ell} R^{(j)}(\underline{\mathfrak{G}}^{(j)}) \right] \right| \leq C\ell^{5/2} \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} .$$

Proof. Note that for every $j \in [\ell]$ the polynomial $Q^{(j)}$ is γ-decaying and that it has bounded influence for every $i \in [n]$:

$$\mathrm{Inf}_i(Q^{(j)}) \leq \mathrm{Inf}_i(P^{(j)}) = \mathrm{Inf}_i(f^{(j)}(\underline{X}^{(j)}) \leq \tau .$$

By definition of χ (Definition A.51) and Theorem A.53,

$$\left| \mathrm{E}\left[\prod_{j=1}^{\ell} Q^{(j)}(\underline{\mathfrak{X}}^{(j)}) - \prod_{j=1}^{\ell} R^{(j)}(\underline{\mathfrak{G}}^{(j)}) \right] \right| = \left| \mathrm{E}\left[\chi\left(\overline{Q}(\underline{\mathfrak{X}})\right) - \chi\left(\overline{Q}(\underline{\mathfrak{G}})\right) \right] \right|$$

$$\leq C\ell^{5/2} \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} .$$

□

Lemma A.85.

$$\mathrm{E}\left[\prod_{j=1}^{\ell} R^{(j)}(\underline{\mathfrak{G}}^{(j)}) \right] \geq \left(\prod_{j=1}^{\ell} \mu'^{(j)} \right)^{\ell/(1-\rho^2)} .$$

Proof. By Theorem A.76. □

Lastly, we need to show that the difference between $\prod_{j=1}^{\ell} \mu'^{(j)}$ and $\prod_{j=1}^{\ell} \mu^{(j)}$ is small.

Claim A.86. *Let* $a \geq 0, \varepsilon \geq 0, a + \varepsilon \leq 1, \beta \geq 1$. *Then,* $(a+\varepsilon)^\beta - a^\beta \leq \beta\varepsilon$.

Proof. The function $h_{\beta,\varepsilon}(a) := (a+\varepsilon)^\beta - a^\beta$ is non-decreasing (since $\frac{\mathrm{d}}{\mathrm{d}a} h_{\beta,\varepsilon} = \beta((a+\varepsilon)^{\beta-1} - a^{\beta-1}) \geq 0$). Hence,

$$(a+\varepsilon)^\beta - a^\beta \leq 1 - (1-\varepsilon)^\beta \leq \beta\varepsilon\,,$$

where in the last step we applied Bernoulli's inequality. □

Lemma A.87. *There exists an absolute constant $C > 0$ such that*

$$\left| \left(\prod_{j=1}^\ell \mu^{(j)} \right)^{\ell/(1-\rho^2)} - \left(\prod_{j=1}^\ell \mu'^{(j)} \right)^{\ell/(1-\rho^2)} \right| \leq \frac{C\ell^2}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C\ln 1/\alpha}}\,.$$

Proof. By Claim A.86,

$$\left| \left(\prod_{j=1}^\ell \mu^{(j)} \right)^{\ell/(1-\rho^2)} - \left(\prod_{j=1}^\ell \mu'^{(j)} \right)^{\ell/(1-\rho^2)} \right| \leq \frac{\ell}{1-\rho^2} \cdot \left| \prod_{j=1}^\ell \mu^{(j)} - \prod_{j=1}^\ell \mu'^{(j)} \right|\,. \tag{88}$$

Since $\mu^{(j)}, \mu'^{(j)} \in [0,1]$,

$$\left| \prod_{j=1}^\ell \mu^{(j)} - \prod_{j=1}^\ell \mu'^{(j)} \right| \leq \sum_{j=1}^\ell \left| \mu^{(j)} - \mu'^{(j)} \right|\,. \tag{89}$$

For a fixed $j \in [\ell]$, from the definition of χ and Theorem A.53 applied with $\ell = 1$,

$$\left| \mu^{(j)} - \mu'^{(j)} \right| = \left| \mathrm{E}\left[\chi\left(Q^{(j)}(\underline{\mathfrak{X}}^{(j)}) \right) - \chi\left(Q^{(j)}(\underline{\mathfrak{G}}^{(j)}) \right) \right] \right| \leq C \cdot \tau^{\frac{\gamma}{C\ln 1/\alpha}}\,. \tag{90}$$

Inequalities (88), (89) and (90) together give the claim. □

Proof of Theorem 4.1. Following the decomposition of $\prod_{j=1}^\ell f^{(j)}(\underline{X}^{(j)})$ into subexpressions (85), (86) and (87), from Lemma A.83, Lemma A.84, Lemma A.85 and Lemma A.87,

$$\mathrm{E}\left[\prod_{j=1}^\ell f^{(j)}(\underline{X}^{(j)}) \right] \geq \left(\prod_{j=1}^\ell \mu^{(j)} \right)^{\ell/(1-\rho^2)} - \varepsilon/2 - C\ell^{5/2} \cdot \tau^{\frac{\gamma}{C\ln 1/\alpha}} - \frac{C\ell^2}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C\ln 1/\alpha}}$$

$$\geq \left(\prod_{j=1}^\ell \mu^{(j)} \right)^{\ell/(1-\rho^2)} - \varepsilon/2 - \frac{2C\ell^{5/2}}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C\ln 1/\alpha}}\,.$$

By choosing $\tau(\varepsilon, \rho, \alpha, \ell, \gamma)$ small enough we get

$$\frac{2C\ell^{5/2}}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C\ln 1/\alpha}} \leq \varepsilon/2\,, \tag{91}$$

which is the main part of the theorem (recall that $\gamma = \frac{(1-\rho)\varepsilon}{2\ell\ln(2\ell/\varepsilon)}$).

To see that we can choose τ as in (12), note that for $D > 0$ big enough we have

$$\tau := \left(\frac{(1-\rho^2)\varepsilon}{\ell^{5/2}}\right)^{D\frac{\ell\ln(\ell/\varepsilon)\ln(1/\alpha)}{(1-\rho)\varepsilon}} \leq \left(\frac{(1-\rho^2)\varepsilon}{\ell^{5/2}}\right)^{D'\frac{2C\ell\ln(2\ell/\varepsilon)\ln(1/\alpha)}{(1-\rho)\varepsilon}}$$
$$= \left(\frac{(1-\rho^2)\varepsilon}{\ell^{5/2}}\right)^{D'\frac{C\ln(1/\alpha)}{\gamma}}$$

for $D' > 0$ as needed. Hence, we obtain

$$\frac{2C\ell^{5/2}}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C\ln 1/\alpha}} = 2C \cdot \frac{\ell^{5/2}}{1-\rho^2} \cdot \left(\frac{(1-\rho^2)\varepsilon}{\ell^{5/2}}\right)^{D'} \leq 2C\varepsilon^{D'} \leq \varepsilon/2,$$

which establishes (91) for this choice of τ. $\qquad\qquad\square$

References

[AM13]   Per Austrin and Elchanan Mossel. Noise correlation bounds for uniform low degree functions. *Arkiv för Matematik*, 51(1):29–52, 2013. 6

[Bec75]   William Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102(1):159–182, 1975. 40

[Bon70]   Aline Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Annales de l'institut Fourier*, 20(2):335–402, 1970. 40

[Bor82]   Christer Borell. Positivity improving operators and hypercontractivity. *Mathematische Zeitschrift*, 180(3):225–234, 1982. 3

[Bor85]   Christer Borell. Geometric bounds on the Ornstein–Uhlenbeck velocity process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 70(1):1–13, 1985. 3, 31

[CDP15]   Wei-Kuo Chen, Nikos Dafnis, and Grigoris Paouris. Improved Hölder and reverse Hölder inequalities for Gaussian random vectors. *Advances in Mathematics*, 280:643–689, 2015. 31, 54

[DFR08]   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. 5

[FK91]   Harry Furstenberg and Yitzhak Katznelson. A density version of the Hales-Jewett theorem. *Journal d'Analyse Mathématique*, 57(1):64–119, 1991. 3

[FLS18]   Jacob Fox, László Miklós Lovász, and Lisa Sauermann. A polynomial bound for the arithmetic k-cycle removal lemma in vector spaces. *Journal of Combinatorial Theory, Series A*, 160:186–201, 2018. 6

[FR18]   Ehud Friedgut and Oded Regev. Kneser graphs are like Swiss cheese. *Discrete Analysis*, 2, 2018. 5, 6

[Fur77]   Harry Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *Journal d'Analyse Mathématique*, 31(1):204–256, 1977. 3

[GL15]   Venkatesan Guruswami and Euiwoong Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. In *SODA*, pages 822–836, 2015. 6

[Gow01]   W. T. Gowers. A new proof of Szemerédi's theorem. *Geometric & Functional Analysis GAFA*, 11(3):465–588, 2001. 3

[Gow07]   W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics*, 166(3):897–946, 2007. 3

[Gre05a]   Ben Green. Finite field models in additive combinatorics. In Bridget S. Webb, editor, *Surveys in Combinatorics*, pages 1–27. 2005. 3, 4

[Gre05b]   Ben Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geometric & Functional Analysis GAFA*, 15(2):340–376, 2005. 6

[Gro75]   Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. 40

[HHM16]   Jan Hązła, Thomas Holenstein, and Elchanan Mossel. Lower bounds on same-set inner product in correlated spaces. In *APPROX-RANDOM*, volume 60 of *LIPIcs*, pages 34:1–34:11, 2016. 7

[Hąz18]   Jan Hązła. On very restricted arithmetic progressions in symmetric sets in finite field model. arXiv:1811.09947, 2018. 6

[KSV09]   Daniel Král', Oriol Serra, and Lluís Vena. A combinatorial proof of the removal lemma for groups. *Journal of Combinatorial Theory, Series A*, 116(4):971–978, 2009. 6

[Led14]   Michel Ledoux. Remarks on Gaussian noise stability, Brascamp-Lieb and Slepian inequalities. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2011–2013*, pages 309–333, 2014. 31, 54

[LPW08]   David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008. 14

[Mes95]   Roy Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *Journal of Combinatorial Theory, Series A*, 71(1):168–172, 1995. 3

[MOO10]   Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010. 40

[MOR⁺06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E. Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006. 3

[Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010. 6, 7, 8, 10, 11, 12, 27, 30, 31

[MOS13] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013. 3, 30

[Mos17] Elchanan Mossel. Gaussian bounds for noise correlation of resilient functions. arXiv:1704.04745, 2017. 6

[Nel73] Edward Nelson. The free Markoff field. *Journal of Functional Analysis*, 12(2):211–227, 1973. 40

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 10, 27

[Rot53] Klaus F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, s1-28(1):104–109, 1953. 3, 4

[RS04] Vojtěch Rödl and Jozef Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures & Algorithms*, 25(1):1–42, 2004. 3

[RS06] Vojtěch Rödl and Jozef Skokan. Applications of the regularity lemma for uniform hypergraphs. *Random Structures & Algorithms*, 28(2):180–194, 2006. 3

[Rud87] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, Inc., 3rd edition, 1987. 47

[Sha10] Asaf Shapira. A proof of Green's conjecture regarding the removal properties of sets of linear equations. *Journal of the London Mathematical Society*, 81(2):355–373, 2010. 6

[Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27(1):199–245, 1975. 3

[TV06] Terence Tao and Van H. Vu. *Additive Combinatorics*. Cambridge University Press, 2006. 3

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

40

AUTHORS

Jan Hązła
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA
jhazla [at] mit [dot] edu
https://idss.mit.edu/staff/jan-hazla/

Thomas Holenstein
Google
Zurich, Switzerland
thomas [dot] holenstein [at] google [dot] com

Elchanan Mossel
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA
elmos [at] mit [dot] edu
https://math.mit.edu/~elmos/