DISS. ETH NO. 23537

Same-Set Hitting with Applications to Parallel Repetition

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

JAN HĄZŁA
MSc in Computer Science
Jagiellonian University in Kraków, Poland

born on September 21, 1987

citizen of
the Republic of Poland

accepted on the recommendation of

Prof. Dr. Thomas Holenstein, examiner
Prof. Dr. Elchanan Mossel, co-examiner
Prof. Dr. Angelika Steger, co-examiner

2016

Contents

Abstract

This thesis consists of two parts investigating topics from discrete mathematics with computer science motivation.

In the first part we study a problem in the *analysis of discrete functions*, which we call *same-set hitting*. Analysis of discrete functions is the study of functions $f : \Omega^n \to \mathbb{R}$ for a finite input alphabet Ω. An important part of this field is the analysis of Boolean functions $f : \{0,1\}^n \to \{0,1\}$. It dates back to the beginnings of the 20th century and has found many applications across computer science and mathematics.

Our problem is motivated by connections to hypercontractivity (which is an important tool in the analysis of discrete functions), hardness of approximation and additive combinatorics. It can be stated as follows:

Let $\mathcal{P}$ be a probability distribution over a finite alphabet Ω^ℓ with all ℓ marginals equal. Let $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}$ be random vectors, where $\underline{X}^{(j)} = (X_1^{(j)}, \ldots, X_n^{(j)})$. Furthermore, suppose that for each coordinate $i \in [n]$ the tuples $(X_i^{(1)}, \ldots, X_i^{(\ell)})$ are identically and independently distributed according to $\mathcal{P}$. We say that the distribution $\mathcal{P}$ is same-set hitting if there exists a function $c_{\mathcal{P}}()$ independent of n such that for every set $S \subseteq \Omega^n$ with $\Pr[\underline{X}^{(j)} \in S] = \mu > 0$:

$$\Pr\left[\underline{X}^{(1)} \in S \wedge \ldots \wedge \underline{X}^{(\ell)} \in S\right] \geq c_{\mathcal{P}}(\mu) > 0 \,.$$

The matter we address is: which probability distributions are same-set hitting? Our main result answers this question in case $\ell = 2$, as well as when $\ell > 2$ and the distribution $\mathcal{P}$ has *bounded correlation* $\rho(\mathcal{P}) < 1$.

The second part of this work discusses bounds on the value of *parallel repetition of multi-prover games*, another problem motivated by hardness of approximation.

An open problem in the field of parallel repetition concerns the difference between the games with two provers, for which we know exponentially

small bounds, and those with three and more provers. In the latter case it is unknown if general exponential bounds exist.

Exploring the differences between those two cases, and motivated by a certain connection to the density Hales-Jewett theorem from additive combinatorics, we study a different kind of parallel repetition bounds, which depend only on the number of repetitions and the question set of a game (so-called *forbidden subgraph bounds*).

We ask the question: Which question sets admit such a bound that decreases exponentially with the number of repetitions?

We demonstrate that because of the connection to the density Hales-Jewett theorem, exponential forbidden subgraph bounds cannot be established for certain sets for three and more provers. However, it is not known if the same holds in case of two provers. This is in contrast to classical bounds, where (as just mentioned) two-prover exponential bounds are known, but multiple-prover case is an open problem.

We use the concept of same-set hitting from the first part of the thesis to obtain some new exponential forbidden subgraph bounds. In particular, interpreting an r-prover questions set as an r-uniform hypergraph, we get such bounds for two-prover question sets that have *treewidth* at most two and for multi-prover sets that are α-*acyclic*.

Zusammenfassung

Diese Arbeit besteht aus zwei Teilen, die Themen der diskreten Mathematik behandeln, welche durch Informatik motiviert sind.

Im ersten Teil untersuchen wir ein Problem in der Analyse diskreter Funktionen, das wir *Gleichemengetreffen* nennen. Die Analyse diskreter Funktionen befasst sich mit Funktionen $f : \Omega^n \to \mathbb{R}$ über einem endlichen Eingabealphabet Ω. Ein wichtiger Teil dieses Gebiets ist die Analyse Boolescher Funktionen $f : \{0,1\}^n \to \{0,1\}$. Dieses Feld reicht zurück zu den Anfängen des 20. Jahrhunderts und fand seither viele Anwendungen in Informatik und Mathematik.

Unser Problem wird durch Beziehungen zu Hyperkontraktivität, einem wichtigen Werkzeug der Analyse diskreter Funktionen, zur Schwere von Approximationsproblemen und zur additiven Kombinatorik motiviert. Es kann folgendermassen gestellt werden:

Sei $\mathcal{P}$ eine Wahrscheinlichkeitsverteilung über einem endlichen Alphabet Ω^ℓ, wobei alle ℓ Randverteilungen gleich sind. Seien $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}$ mit $\underline{X}^{(j)} = (X_1^{(j)}, \ldots, X_n^{(j)})$ zufällige Vektoren, so dass für Koordinaten $i \in [n]$ die Tupel $(X_i^{(1)}, \ldots, X_i^{(\ell)})$ entsprechend der Verteilung $\mathcal{P}$ unabhängig identisch verteilt sind. Wir sagen, dass die Verteilung $\mathcal{P}$ gleichemengetreffend ist, wenn es eine Funktion $c_{\mathcal{P}}()$ gibt, die unabhängig von n ist, so dass für jede Menge $S \subseteq \Omega^n$ mit $\Pr[\underline{X}^{(j)} \in S] = \mu > 0$ gilt:

$$\Pr\left[\underline{X}^{(1)} \in S \wedge \ldots \wedge \underline{X}^{(\ell)} \in S\right] \geq c_{\mathcal{P}}(\mu) > 0 \,.$$

Dabei betrachten wir die folgende Fragestellung: welche Verteilungen sind gleichemengetreffend? Unser Hauptresultat beantwortet diese Frage für den Fall $\ell = 2$, sowie wenn $\ell > 2$ und die Verteilung $\mathcal{P}$ *begrenzte Korrelation* $\rho(\mathcal{P}) < 1$ hat.

Der zweite Teil dieser Arbeit betrachtet Schranken für den Wert *paralleler Wiederholung von Multibeweiserspielen*, ein weiteres Problem, das durch die Schwere von Approximationsproblemen motiviert wird.

Ein ungelöstes Problem des Bereichs der paralleler Wiederholung betrifft den Unterschied zwischen Spielen mit zwei Beweisern, für welche exponentiell kleine Schranken bekannt sind, und diesen mit drei und mehr Beweisern. Für letztere ist die Existenz genereller exponentieller Schranken nicht bekannt.

Um die Unterschiede dieser beiden Fälle zu erforschen, und motiviert durch eine bestimmte Beziehung zum Density Hales-Jewett Satz aus der additiven Kombinatorik, untersuchen wir eine andere Art von Schranken für parallele Wiederholung, die nur von der Anzahl der Wiederholungen und der Fragemenge des Spieles abhängen (sogenannte *Schranken verbotener Teilgraphen*).

Wir interessieren uns dafür, welche Fragemengen eine solche Schranke haben, die exponentiell mit der Anzahl der Wiederholungen sinkt.

Wir zeigen, dass auf Grund der Verbindung zum Density Hales-Jewett Satz, exponentielle Schranken verbotener Teilgraphen für gewisse Mengen im Fall von drei und mehr Beweisern nicht existieren können. Jedoch ist nicht bekannt, ob das auch für zwei Beweiser geht. Das ist im Gegensatz zu klassischen Schranken, für die (wie oben erwähnt) im Fall von zwei Beweisern exponentielle Schranken bekannt sind, aber der Fall von mehreren Beweisern ungelöst ist.

Wir benutzen das Konzept des Gleichemengetreffens aus dem ersten Teil dieser Arbeit, um einige neue exponentielle Schranken verbotener Teilgraphen zu zeigen. Insbesondere, weil man eine Fragemenge für r Beweiser als einen r-uniformen Hypergraphen interpretieren kann, zeigen wir entsprechende Schranken für Fragemengen für zwei Beweiser, die *Baumweite* höchstens zwei haben, sowie für Fragemengen für mehrere Beweiser, die α-*azyklisch* sind.

Acknowledgements

First of all I would like to express my deepest gratitude to my supervisor, Thomas Holenstein. Many of my friends can attest that one of my favorite hobbys is complaining (which is understandable, since I consider myself a typical Pole in this respect). Therefore the best I can compliment Thomas is: I can hardly complain about anything during my time in his group. He was a perfect, considerate boss.

I am grateful to my other examiners: Elchanan Mossel for a fruitful collaboration, his kindness and helpful career advice, and Angelika Steger for her useful feedback on the thesis and generously enabling me to stay at ETH hassle-free after Thomas had left.

I would like to acknowledge the whole of the Institute of Theoretical Computer Science at ETH Zurich. I have never worked (and probably never will) for such a well-organised and reasonable organisation. This was not least due to the great work of all the administrative staff I had interacted with: Marianna Berger, Beate Bernhard, Claudia Günthart, Andrea Salow, Denise Spicher and Sile Hasler.

I would like to thank my long-time officemates and fellow students: Chandan Dubey and Robin Künzler for all great times we had together at work and privately. In the same way I want to thank my Master's students: Nemanja Škorić and Felix Weissenberger. Spending time with both of them was a pleasure. I would also like to thank the entire Angelika's group for "adopting" me for the final stretch of my PhD and for all excessively long coffee breaks and enjoyable (if not obviously productive) discussions. Ich bin auch dankbar aller, die (zu viel) ihre Zeit damit verbrachten, mit mir die Zusammenfassung zu übersetzen: Felix, Frank Mousset und Andreas Noever. Finally, I am grateful to all friends and colleagues I met during my time at ETH and in Zurich.

Last but not least, chciałbym podziękować tym, którzy przyczynili się do tego doktoratu po polsku: wszystkim, z którymi miałem przyjemność zetknąć się podczas studiów na Uniwersytecie Jagiellońskim (szczególnie promotorowi mojej pracy magisterskiej, Jakubowi Kozikowi oraz Tomkowi Rużyckiemu), a także Przyjaciołom z Gniezna, Krakowa, Zurychu i skądinąd. Chciałbym

również podziękować z całego serca mojej Rodzinie za bezwarunkowe wsparcie i miłość.

Chapter 1

Introduction

This thesis consists of two parts. The first part (Chapter 2) discusses a property of discrete probability distributions called *same-set hitting*. In the second part (Chapter 3) we apply related techniques to the problem of *parallel repetition of multi-prover games*.

In the two following sections we give respective introductions to those two topics. Section 1.1 discusses same-set hitting and Section 1.2 considers parallel repetition.

1.1 Same-Set Hitting

1.1.1 Background

The analysis of Boolean and discrete functions is a well established field at the intersection of discrete mathematics and theoretical computer science. It concerns itself with the study of *Boolean functions* $f : \{0,1\}^n \to \{0,1\}$ and, more broadly, *discrete functions* $f : \Omega^n \to \mathbb{R}$ for a finite alphabet Ω. In particular, it analyzes various aspects of behavior of those functions when their inputs are sampled from different probability distributions.

The idea of studying Boolean functions in the context of computer science is usually credited to Nakashima [Nak35], Shannon [Sha37] and Shestakov [She38]. Arguably the most famous tool of the analysis of Boolean functions, the discrete Fourier expansion, was introduced by Walsh [Wal23] and Paley [Pal32] with a motivation arising from the study of $L^2([0,1])$ space in mathematical analysis.

A very important concept in the analysis of discrete functions is the *influence* of a function $f : \Omega^n \to \mathbb{R}$ on a given coordinate $i \in [n] := \{1, \ldots, n\}$. The influence is a non-negative value that indicates how much impact the i-th coordinate has on the value of a function. It was discovered independently multiple times in the contexts of law and social sciences [Pen46, Ban65, Col71].

In particular, Banzhaf [Ban65] used it to argue that a certain local voting system in the US gave zero influence to some of the involved municipalities. It was first considered in the computer science setting by Ben-Or and Linial [BL85].

The methods of the analysis of discrete functions have led to many impressive applications, including the famous Arrow's theorem [Arr50] on the impossibility of a perfect voting system, the hardness of approximation of 3-SAT results by Håstad [Hås01] and non-trivial lower bounds on the maximum influence of a function given by the theorem of Kahn, Kalai and Linial [KKL88]. For an extensive introduction to the field together with historical notes (on which this short overview is mostly based) we refer to the textbook by O'Donnell [O'D14].

The theory of *hypercontractivity* is a powerful tool that belongs to the analysis of discrete functions. The original hypercontractivity theorem (for the uniform distribution over Boolean space $\{0,1\}^n$) has a rather complicated history [Bon70, Nel73, Gro75, Bec75] and there exist multiple follow-up works extending it to more alphabets and probability distributions and sharpening the bounds in the theorem [Tal94, Ole03, Wol07]. Hypercontractivity has found many applications in theoretical computer science, notably for the aforementioned KKL theorem, as well as for approximability results using semi-definite programs (e.g., [ARV09, KV15]).

It turns out that hypercontractivity has a dual theory called *reverse hypercontractivity*. It was first proved by Borell [Bor82], but expanded only recently [MOR$^+$06, MOS13]. It has already found applications, for example to the *non-interactive correlation distillation* [MOR$^+$06] which deals with the problem of agreeing on a shared random bit in the presence of noise, to a quantitative version of Arrow's theorem [Mos12] or to inapproximability results for certain types of hypergraph colorings [GL15].

The precise theorem statements for hypercontractivity and reverse hypercontractivity are somewhat technical and require introducing several concepts first. In short, those theorems are inequalities between different p-norms of certain discrete functions (for details see Appendix A, in particular Sections A.4 and A.8). However, there is a simple corollary of reverse hypercontractivity that is elementary to state. Furthermore, the applications we mentioned use this corollary rather than the full theorem. We call the property of probability distributions given by this corollary *set hitting*. However, not every distribution is set hitting. In the first part of this thesis, we introduce a weaker (but still useful) property, which is exhibited by a broader class of discrete distributions: *same-set hitting*.

1.1.2 Basic example

To obtain some intuition on the problem we are studying, consider the following example. For a set $S \subseteq \{0,1,2\}^n$, define its measure as $\mu(S) := |S|/3^n$. 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 $\{X_i, X_i + 1 \pmod 3\}$. Our goal is to show that there exists a function $c : (0,1) \to (0,1)$ such that for every n and every $S \subseteq \Omega^n$ with $\mu(S) = \mu$,

$$\Pr[\underline{X} \in S \wedge \underline{Y} \in S] \geq c(\mu) > 0 \tag{1}$$

holds. In particular, the probability in (1) is bounded away from 0 by an expression which depends only on μ and not on n.

1.1.3 Set hitting and same-set hitting

More generally, let Ω be a finite alphabet. We consider *ℓ-step probability distributions over* Ω, that is distributions $\mathcal{P}$ over Ω^ℓ for some $\ell \geq 2$, where we think of subsequent coordinates as steps in a random process.

Furthermore, assume we are given $n \in \mathbb{N}$. We consider ℓ vectors $\underline{X}^{(1)}$, $\ldots, \underline{X}^{(\ell)}$, where $\underline{X}^{(j)} = (X_1^{(j)}, \ldots, X_n^{(j)})$ 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 2.1 for an overview of the notation).

We say that a distribution $\mathcal{P}$ is *set hitting*[1] if for every $\mu \in (0,1]$ there exists $c(\mu) \in (0,1]$ such that for every ℓ-tuple of sets $S^{(1)}, \ldots, S^{(\ell)} \subseteq \Omega^n$ with $\Pr[\underline{X}^{(j)} \in S^{(j)}] \geq \mu$ for every $j \in [\ell]$ we have

$$\Pr\left[X^{(1)} \in S^{(1)} \wedge \ldots \wedge X^{(\ell)} \in S^{(\ell)}\right] \geq c(\mu) \,.$$

As we already mentioned, set hitting turns out to be a useful property in the context of some applications [MOR$^+$06, GL15]. Therefore, it makes sense to ask which probability distributions are set hitting. It is easy to see that not every distribution has this property. For example, since in the distribution from Section 1.1.2 we have $\mathcal{P}(0,2) = 0$, we can take $S := \{\underline{x} : x_1 = 0\}$ and $T := \{\underline{y} : y_1 = 2\}$. This gives us $\Pr[\underline{X} \in S] = \Pr[\underline{Y} \in T] = 1/3$ and, at the same time,

$$\Pr\left[\underline{X} \in S \wedge \underline{Y} \in T\right] = 0 \,.$$

[1] The definition we use later is a bit more general (but more difficult to read), see Section 2.1.2.

It is easy to extend this example to any distribution that (informally speaking) does not have a full support. It turns out that this is the only obstacle, and the proof is an easy application of reverse hypercontractivity by Mossel, Oleszkiewicz and Sen [MOS13]:

Theorem 1.1 (see Lemma 8.3 in [MOS13])**.** *A probability space $\mathcal{P}$ is set hitting if and only if:*

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

where $\mathrm{supp}(X)$ *denotes the support of a random variable* X*.*

Even though a distribution given in Section 1.1.2 is not set hitting, one could still hope for the following, weaker, but still meaningful property:

We say that $\mathcal{P}$ is *same-set hitting* if for every $\mu \in (0,1]$ there exists $c(\mu) \in (0,1]$ such that for every *single* set $S \subseteq \Omega^n$ with $\Pr[\underline{X}^{(j)} \in S] \geq \mu$ we have:

$$\Pr\left[X^{(1)} \in S \wedge \ldots \wedge X^{(\ell)} \in S\right] \geq c(\mu)\ . \tag{3}$$

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

1.1.4 Our contributions

We present the original contributions of this thesis to same-set hitting. Most of them are contained in [HHM16]. We need a couple more definitions first.

We say that an ℓ-step probability distribution $\mathcal{P}$ over an alphabet Ω *has equal marginals* if for every $j \in [\ell]$ and every $x \in \Omega$:

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

As explained in Section 2.2.4, same-set hitting is interesting only for distributions with equal marginals. Whenever we discuss such distributions, we assume w.l.o.g that Ω is equal to the support of the marginal.

Let us also define:

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

Below we present our contributions in four related subareas:

Two steps

Observe that $\alpha(\mathcal{P}) > 0$ is a necessary condition for same-set hitting. Indeed, if $\mathcal{P}(x^*, \ldots, x^*) = 0$ for some $x^* \in \Omega$, then $S := \{\underline{x} : x_1 = x^*\}$ is a constant-measure set with

$$\Pr\left[X^{(1)} \in S \wedge \ldots \wedge X^{(\ell)} \in S\right] = 0 \,.$$

In Section 2.4 (Theorem 2.29 and Corollary 2.30) we show that in case of two steps this condition is also sufficient. Namely, we show that a two-step probability distribution with equal marginals $\mathcal{P}$ is same-set hitting if and only if $\alpha(\mathcal{P}) > 0$. This provides a full classification of same-set hitting for two-step distributions.

Note that if $\beta(\mathcal{P}) > 0$, then same-set hitting follows from Theorem 1.1. The case $\beta(\mathcal{P}) = 0$, in particular concerning the probability space from Section 1.1.2, is our original contribution.

Multiple steps

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

We provide the following partial results. Firstly, by an inductive argument based on the two-step theorem, in Section 2.5 (Theorem 2.43) we show that multi-step distributions induced by Markov chains are same-set hitting.

Secondly, in Sections 2.2 and 2.3 (Theorem 2.10) we prove that $\mathcal{P}$ is same-set hitting if $\alpha(\mathcal{P}) > 0$ and its *correlation* $\rho(\mathcal{P})$ is smaller than 1. Intuitively (but not quite correctly), the opposite condition $\rho(\mathcal{P}) = 1$ means that for some step j the value of $X_i^{(j)}$ is deterministic given values of all other steps $X_i^{(1)}, \ldots, X_i^{(j-1)}, X_i^{(j+1)}, \ldots, X_i^{(\ell)}$.

Formally, the 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 $S, T \neq \emptyset$, $S \neq \Omega$, and

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

in other words we can partition Ω and $\Omega^{\ell-1}$ into two non-empty pairs that are fully correlated with each other.

For example, taking $\Omega := \mathbb{Z}_3$ with three steps and the distribution $\mathcal{P}$ uniform over $\{000, 111, 222, 012, 120, 201\}$, one can see that $\rho(\mathcal{P}) = 1$ by taking j to be the first step, $S := \{0\}$ and $T := \{00, 12\}$. On the other hand, the space from Section 1.1.2 has $\rho(\mathcal{P}) < 1$. For the full definition of $\rho(\mathcal{P})$, see Definition 2.8.

As a matter of fact, Theorem 2.10 is our main result, with the two-step and Markov theorems relying on it. Section 2.2 contains the proof sketch and the discussion of the assumptions of the theorem. We give the full proof in Section 2.3.

The proof of Theorem 2.10 relies on an adapted version of the low-influence theorem from [Mos10]. This theorem in turn is based on a refined version of invariance principle, which is another important tool in the analysis of discrete functions. The landmark application of the invariance principle is the "Majority is stablest" theorem [MOO10], which in turn gives the optimality of the Goemans-Williamson approximation algorithm for Max-Cut under the unique games conjecture [GW95, KKMO07].

The relation between the theorem in [Mos10] and our modified version is explained in Section 2.2.3. The full proof of the low-influence theorem is attached for completeness in Appendix A.

Finally, in Section 2.8 we explore the conjecture that $\alpha(\mathcal{P}) > 0$ and equal marginals are sufficient for same-set hitting, even when $\rho(\mathcal{P}) = 1$. We are unable to prove the general result, but we provide a toy version that establishes same-set hitting for dictators, linear functions and thresholds.

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. Specifically, in Section 2.6 (Theorem 2.45) we show that if $\rho(\mathcal{P}) < 1$, then a distribution $\mathcal{P}$ is set hitting for functions with $\Omega(1)$ expectation and $o(1)$ largest Fourier coefficient.

Polynomial same-set hitting

We also consider a stronger notion of *polynomial same-set hitting*, where we require $c(\mu)$ from the bound (3) to be at least μ^C for some $C \geq 0$.

The bounds that [MOS13] obtain when proving Theorem 1.1 actually imply that all distributions that are set hitting are also polynomially set hitting. However, our method of establishing same-set hitting gives much worse bounds: they are triply exponentially small in case of the multi-step theorem.

With that respect we provide the following contribution: In Section 2.7 (Theorem 2.51) we show how to obtain polynomial same-set hitting for all *symmetric* two-step distributions, i.e., the ones where $\mathcal{P}(x, y) = \mathcal{P}(y, x)$.

1.1.5 Related work

Understanding set hitting by a number of consecutive steps of a process has been of intense study also in additive combinatorics (where almost always

$\rho = 1$).

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

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

Some of the distributions $\mathcal{P}_{Z,\ell}$ can be shown to be same-set hitting using 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.*

This follows a long line of work, started by Szemerédi lemma [Sze75], its proof by Furstenberg using the ergodic theorem [Fur77] as well as finite group and multidimensional versions, see, e.g., [Rot53, FK91, Gow01].

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 less symmetric spaces. This suggests that proving the conjecture fully in $\rho = 1$ case might be a difficult undertaking.

The case of $\rho < 1$ has also been studied in the context of extremal combinatorics and hardness of approximation. In particular, Mossel [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 2.10. Additionally, Theorem 2.45 can be seen as a strengthening of [Mos10].

Furthermore, Austrin and Mossel [AM13] establish the result equivalent to Theorem 2.45 assuming in addition to $\rho(\mathcal{P}) < 1$ also that the steps of $\mathcal{P}$ are pairwise independent (they also prove results for the case $\rho(\mathcal{P}) = 1$ with pairwise independence but these involve only bounded degree functions).

Finally, we note that another relevant paper in the case of $\ell = 2$ and symmetric $\mathcal{P}$ is by Dinur, Friedgut and Regev [DFR08], who give a characterization of non-hitting sets. However, due to a different framework, their results are not directly comparable to ours.

Although we do not have a direct application, it is conceivable that our work might be useful in inapproximability. 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.

1.2 Parallel Repetition

1.2.1 Multi-prover games

An *r-prover game* is a protocol, in which r *provers* have a joint objective of making another entity, the *verifier*, accept. The execution of such a game looks as follows: the verifier first samples r questions $q^{(1)}, \ldots, q^{(r)} \in Q^{(1)} \times \ldots \times Q^{(r)}$. Those questions are sampled uniformly from some *question set* $\overline{Q} \subseteq Q^{(1)} \times \ldots \times Q^{(r)}$.

Then, she sends the questions to the provers: the j-th prover receives $q^{(j)}$ and sends back an answer $a^{(j)}$ (from a finite answer alphabet) that depends only on $q^{(j)}$. Finally, the verifier accepts or rejects based on the evaluation of a *verification predicate* $V(q^{(1)}, \ldots, q^{(r)}, a^{(1)}, \ldots, a^{(r)})$.

A *strategy* $(\mathcal{S}^{(1)}, \ldots, \mathcal{S}^{(r)})$ for the provers consists of r functions, the j-th of which maps questions to answers for the j-th prover. The *value* of a game is

$$\mathrm{val}(\mathcal{G}) := \max_{\mathcal{S}^{(1)},\ldots,\mathcal{S}^{(r)}} \Pr\left[V\left(q^{(1)}, \ldots, q^{(r)}, \mathcal{S}^{(1)}(q^{(1)}), \ldots, \mathcal{S}^{(r)}(q^{(r)})\right) = 1\right] ,$$

where the maximum is over all strategies and the probability over the uniform choice of $q^{(1)}, \ldots, q^{(r)} \in \overline{Q}$. A game $\mathcal{G}$ is called *trivial* if $\mathrm{val}(\mathcal{G}) = 1$.

A formal definition is provided in Section 3.1.1. One might consider allowing other distributions over a question set $\overline{Q}$ than the uniform one. This would not make much difference in our setting, as discussed in Section 3.1.4.

1.2.2 Parallel repetition

The *n-fold parallel repetition* $\mathcal{G}^n$ of an r-prover game $\mathcal{G}$ is a game where the verifier samples n independent question tuples, sends n questions to each prover, receives n answers from each prover, and accepts if all n instances of the verification predicate for $\mathcal{G}$ accept.

Since the provers can play an optimal strategy for $\mathcal{G}$ independently in each coordinate, it is easy to see that $\mathrm{val}(\mathcal{G}^n) \geq \mathrm{val}(\mathcal{G})^n$, in particular if $\mathcal{G}$ is trivial, then $\mathcal{G}^n$ is trivial as well. However, since the i-th answer of a prover can depend on all of his questions, as opposed to just the i-th one, it is possible that $\mathrm{val}(\mathcal{G}^n)$ attains a higher value.

As a simple example, in Section 3.1.2 we show a family of r-prover games with $\mathrm{val}(\mathcal{G}^r) = \mathrm{val}(\mathcal{G}) = 1/2$.

The question, some aspects of which we address in this thesis, is how strong this phenomenon can be for a non-trivial game. In particular, does the value $\mathrm{val}(\mathcal{G}^n)$ have to converge to 0 as n goes to infinity and, if yes, what is the rate of convergence.

1.2.3 Background

Two-prover games in the context of theoretical computer science were first introduced by Ben-Or, Goldwasser, Kilian and Wigderson [BGKW88]. Extensive works on their parallel repetition were produced. Here we present only the ones that are most relevant to this thesis. For a more extensive survey we refer to [Fei95] and [Raz10].

Fortnow, Rompel and Sipser [FRS88] were the first to treat the value of two-prover repeated games by incorrectly claiming $\mathrm{val}(\mathcal{G}^n) = \mathrm{val}(\mathcal{G})^n$. This was followed by an errata [FRS90] showing an example with $\mathrm{val}(\mathcal{G}^2) > \mathrm{val}(\mathcal{G})^2$.

One important motivation for showing that the value of a repeated two-prover game decreases fast with the number of repetitions is for proving NP-hardness of approximation. Roughly speaking, one can think of an instance of an optimization problem as a two-prover game $\mathcal{G}$, where $\mathrm{val}(\mathcal{G})$ corresponds to the value of the instance normalized to $[0, 1]$. Then, assuming it is computationally hard to distinguish between instances with value 1 and with value $1-\epsilon$ for some small $\epsilon > 0$, the parallel repetition can be used to show that it is also hard to distinguish between values 1 and δ, where $\delta \ll 1 - \epsilon$. This is because the n-fold repetition maps trivial games to trivial games, but (hopefully) decreases the value of non-trivial games. A famous example of an application of this technique is due to Håstad [Hås01].

The most important parallel repetition upper bound for two-prover games discovered by Raz [Raz98] and improved by Holenstein [Hol09] gives

$$\mathrm{val}(\mathcal{G}^n) \leq \exp\left(-\Omega\left(\epsilon^3 n / \log\left|\overline{A}\right|\right)\right) \tag{4}$$

for a two-prover game $\mathcal{G}$ with $\mathrm{val}(\mathcal{G}) = 1 - \epsilon$ and answer set $\overline{A} = A^{(1)} \times A^{(2)}$. This bound, as well as numerous others that depend on the same parameters (e.g., [Rao11, BRR$^+$09]), is usually proved by employing information theory. In the multi-prover case, such a bound is known for free games (i.e., those, where the questions to the provers are sampled independently, see, e.g., [CWY15]). However, the only known fully general bound is by Verbitsky [Ver96]:

$$\mathrm{val}(\mathcal{G}^n) \leq \omega_r^{\mathrm{DHJ}}(n)\ , \tag{5}$$

where $\mathcal{G}$ is a non-trivial game with the question set of size r, and $\omega_r^{\mathrm{DHJ}}(n)$ is the threshold from the famous *density Hales-Jewett theorem* [FK91] from additive combinatorics (see Section 3.1.3 for details):

$$\omega_r^{\mathrm{DHJ}}(n) := \text{maximum measure of a subset of } [r]^n \text{ without a combinatorial line.}$$

$\omega_r^{\mathrm{DHJ}}(n)$ is known to go to 0 as n goes to infinity, but only very slowly. For example, we have $\omega_2^{\mathrm{DHJ}}(n) = \Theta(1/\sqrt{n})$ and $1/\exp(\sqrt{\log n}) \leq \omega_3^{\mathrm{DHJ}}(n) \leq 1/\sqrt{\log^* n}$.

Furthermore, it was shown [HHR16] that the inequality (5) is sometimes tight for *three* and more provers. It is open if it is ever tight in the case of two provers.

If the answer was positive, it would constitute an interesting difference between two and more provers. One could hope that this sheds some light on why bound like (4) seems to be harder to prove for multiple provers. However, it should be noted that even the positive answer does not logically preclude (4) from being true in the multi-prover case (since (4) and (5) are incomparable with each other).

This potential difference between two- and multiple-prover cases is a motivation for our research in the second part of this thesis. We consider so-called *forbidden subgraph bounds*, discussed in the section below.

1.2.4 Forbidden subgraph bounds

We are interested in upper bounds on $\mathrm{val}(\mathcal{G}^n)$ that depend only on the question set $\overline{Q}$ and the number of repetitions n. Let $\omega_{\overline{Q}}(n) := \max_{\mathcal{G}} \mathrm{val}(\mathcal{G}^n)$, where the maximum is over all non-trivial games $\mathcal{G}$ with question set[2] $\overline{Q}$. We say that the question set $\overline{Q}$ *admits parallel repetition* if $\lim_{n\to\infty} \omega_{\overline{Q}}(n) = 0$. Furthermore, we will say that $\overline{Q}$ *admits exponential parallel repetition* if there exists $C_{\overline{Q}} < 1$ such that

$$\omega_{\overline{Q}}(n) \leq (C_{\overline{Q}})^n \ .$$

We will be interested in which $\overline{Q}$ admit exponential parallel repetition. Note that if one proved that all two-prover question sets admit exponential parallel repetition, it would mean that (5) can never be tight in the two-prover case (since $\omega_2^{\mathrm{DHJ}}(n) \geq \Omega(1/\sqrt{n})$).

The bounds on $\omega_{\overline{Q}}(n)$ are called forbidden subgraph bounds, with the name explained in [FV02].

1.2.5 Our contributions

We present the original contributions of this thesis in the area of parallel repetition. They consist of some positive and negative results on exponential parallel repetition for different question sets $\overline{Q}$ and are mostly contained in [HHR16].

[2] Note that the number of provers r is implicitly determined by $\overline{Q}$.

We note here that such a set $\overline{Q}$ for an r-prover game can be naturally interpreted as an r-uniform, r-partite hypergraph. In particular, for a two-prover game it can be thought of as a bipartite simple graph.

On the positive side, in Section 3.2 (Theorem 3.21) we define a class of hypergraphs that are *constructible by conditioning* and prove that they admit exponential parallel repetition.

The construction and the proof are inspired by results of Chapter 2. In particular, one of the main ingredients in the proof consists in establishing of same-set hitting of a certain probability distribution.

Based on this, in Section 3.3 (Theorem 3.36) we show that all bipartite graphs with treewidth at most two are constructible and therefore admit exponential parallel repetition. This is an improvement over previous results [Ver95, Wei13] that showed exponential parallel repetition for trees and cycles.

Furthermore, in Section 3.4 (Theorem 3.42) we prove that all multi-prover question hypergraphs that are α-*acyclic* admit parallel repetition. α-acyclicity is a generalization of the notion of graph acyclicity and therefore this result can be seen as a generalization of the exponential parallel repetition for trees [Ver95]. This is a considerable improvement over previous best bound in this case, which was (5). We are not aware of any previous results of this kind in the multi-prover setting, except for free games.

As for the negative results, in Section 3.5 (Theorem 3.55) we use the previously mentioned equivalence of parallel repetition with the density Hales-Jewett theorem [HHR16] to conclude that there exist question sets for three and more provers that do *not* admit exponential parallel repetition. We also use this result to construct an interesting family of multi-prover games with good lower bounds on their parallel repetition value.

It remains open if all two-prover question sets admit exponential parallel repetition. In Section 3.6 (Theorem 3.59) we show that at least our notion of graph constructability is not enough to prove it. We do this by exhibiting a specific bipartite graph on 12 vertices which is not constructible by conditioning. The proof is computer-assisted and turns out to be unexpectedly complicated. We explain where the difficulty lies in Section 3.6.1.

We note that even when we generalize previous work, our proofs do not seem to be generalizations of earlier ones, but rather they establish a genuinely new technique inspired by the work presented in Chapter 2. Finally, we observe that where previous work is available, our bounds are mostly quantitatively worse. This seems to be a trade-off for the generality of our method.

1.2.6 Related work

Verbitsky [Ver96] used the density Hales-Jewett theorem to show that every question set, regardless of the number of provers, admits parallel repetition. This result is shown for completeness in Section 3.1.3.

As concerns the exponential parallel repetition, Cai, Condon and Lipton [CCL92] showed it for free games (their proof is for the two-prover case, but using a result on hypergraphs by Erdős [Erd64] it generalizes to multiple provers), with further works [Fei91, Pel95] improving the bounds.

Less is known for non-free games: Verbitsky [Ver95] proved exponential parallel repetition for two-prover games when the question set is a tree and Weissenberger [Wei13] for a cycle[3].

There is little work on multi-prover case. An exponential bound like (4) for free games is considered folklore and there are some results in related models of games with *entangled* and *no-signaling* strategies, see, e.g., [CWY15, BFS14]. Furthermore, exponential parallel repetition for so called *anchored* multi-prover games is shown by Bavarian, Vidick and Yuen [BVY15].

Most of the work done on parallel repetition lower bounds is based on two-prover examples by Feige and Verbitsky [FV02] and Raz [Raz11] with little known in the multi-prover case.

1.3 Preliminaries

In this section we introduce basic preliminaries and notation used throughout the thesis.

Perhaps unfortunately for the reader, many of our results feature doubly-indexed collections of elements. We adopt the following conventions:

- Most of the time we consider two dimensions, one of which corresponds to n independent coordinates. The other one can denote, e.g., r provers or ℓ steps of a random process. Most of the time n is meant to be large compared to the other dimension.

- We index the n-dimension with i in the subscript and the other dimension with j in parentheses in the superscript. We denote aggregation over i by underline and over j by overline. For example:

$$\begin{aligned} \overline{\underline{V}} &= (\underline{V}^{(1)}, \dots, \underline{V}^{(j)}, \dots, \underline{V}^{(r)}) = (\overline{V}_1, \dots, \overline{V}_i, \dots, \overline{V}_n) \\ \underline{V}^{(j)} &= (V_1^{(j)}, \dots, V_i^{(j)}, \dots, V_n^{(j)}) \\ \overline{V}_i &= (V_i^{(1)}, \dots, V_i^{(j)}, \dots, V_i^{(r)}) \end{aligned}$$

[3] As a matter of fact, his result is even stronger: it gives an upper bound that depends only on $\mathrm{val}(\mathcal{G})$, but not on question or answer alphabet size.

- We call the element collections aggregated over i (like $\underline{v}^{(j)}$) *vectors* and the element collections aggregated over j (like $\overline{v}_i$) *tuples*.

We let $\mathbb{N} := \{0, 1, 2, \ldots\}$. For $k \in \mathbb{N}_{>0}$, we let $[k] := \{1, \ldots, k\}$. For sets A, B we sometimes write $A \cup B$ as $A \dot{\cup} B$ to emphasize that $A \cap B = \emptyset$. For an event $\mathcal{E}$ we denote its indicator function by $\mathbb{1}_{\mathcal{E}}$. The powerset of X is denoted by 2^X. In accordance with the computer science tradition, by log we mean the logarithm of base two.

Let X be a set and $n \in \mathbb{N}_{>0}$. For a vector $\underline{x} \in X^n$ and $y \in X$ we define the *y-weight of $\underline{x}$* as $w_y(\underline{x}) := |\{i \in [n] : x_i = y\}|$.

For a random variable X, its *variance* is $\mathrm{Var}[X] := \mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$. For random variables X, Y their *covariance* is

$$\mathrm{Cov}[X, Y] := \mathrm{E}\left[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\right] .$$

For a discrete random variable Y over a support Ω and a random variable X we define another random variable $\mathrm{E}[X \mid Y]$ that is coupled with Y in the natural way, i.e., whenever $Y = y \in \Omega$, we have $\mathrm{E}[X \mid Y] = \mathrm{E}[X \mid Y = y]$. We also define $\mathrm{Var}[X \mid Y]$ by analogy.

Theorem 1.3 (Chernoff-Hoeffding bound, Theorem 1 in [Hoe63])**.** *Let random variable $X_1, \ldots, X_n$ be i.i.d. in $\{0, 1\}$, with $\mathrm{E}[X_i] = p$ and let $\epsilon > 0$. Then,*

$$\Pr\left[\sum_{i=1}^{n} X_i \geq n(p + \epsilon)\right] \leq \exp\left(-2\epsilon^2 n\right) .$$

Chapter 2

Same-Set Hitting

2.1 Notation and Preliminaries

We start with introducing necessary preliminaries: Section 2.1.1 introduces our notation, while Sections 2.1.2 to 2.1.4 give the definitions we will work with throughout this chapter.

2.1.1 Notation

We will now introduce our setting and notation specific to same-set hitting. We refer the reader to Figure 2.1 for an overview.

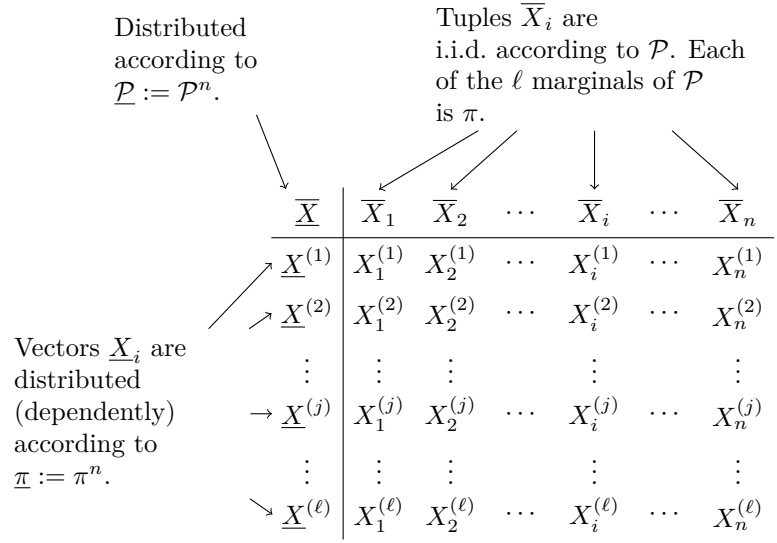
We always assume that we have n independent coordinates. In each coordinate i we pick ℓ values $X_i^{(j)}$ for $j \in [\ell] = \{1, \dots, \ell\}$ at random using some distribution. Each value $X_i^{(j)}$ is chosen from the same fixed set Ω, and the distribution of the tuple $\overline{X}_i = (X_i^{(1)},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

This gives us values $X_i^{(j)}$ for $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, \ell\}$. Thus, we have ℓ vectors $\underline{X}^{(1)}, \dots, \underline{X}^{(\ell)}$, where $\underline{X}^{(j)} = (X_1^{(j)}, \dots, X_n^{(j)})$ 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 2.2.4 we assume that all of $X_i^{(1)}, \dots, X_i^{(\ell)}$ have the same marginal distribution, which we call π. We assume that Ω is the support of π.

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

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



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

$\rho(\mathcal{P})$: See Definition 2.8

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

Figure 2.1: 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 $\underline{\mathcal{P}}$.

As visible in Figure 2.1, 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 $\overline{\underline{X}} = (\overline{X}_1, \ldots, \overline{X}_n) = (\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)})$.

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., $\overline{X}^{\setminus j} := (X^{(1)}, \ldots, X^{(j-1)}, X^{(j+1)}, \ldots, X^{(\ell)})$.

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

2.1.2 Formal definitions

In this section we state the definitions necessary for the formal statements of our theorems. The definitions of set hitting and same-set hitting provided here are slightly more general (though actually equivalent) than the ones presented in the introduction.

Definition 2.1. Let $\mu, \delta \in (0, 1]$. We say that an ℓ-step distribution $\mathcal{P}$ *is* (μ, δ)*-set hitting*, if, whenever functions $f^{(1)}, \ldots, f^{(\ell)} : \underline{\Omega} \to [0, 1]$ satisfy $\mathrm{E}[f^{(j)}(\underline{X}^{(j)})] \geq \mu$ for every $j \in [\ell]$, we have:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq \delta\,. \tag{6}$$

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. ◊

Definition 2.2. Let $\mu, \delta \in (0, 1]$. We say that an ℓ-step distribution $\mathcal{P}$ *is* (μ, δ)*-same-set hitting*, if, whenever a function $f : \underline{\Omega} \to [0, 1]$ satisfies $\mathrm{E}[f(\underline{X}^{(j)})] \geq \mu$ for every $j \in [\ell]$, we have:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq \delta\,. \tag{7}$$

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. ◊

Definition 2.3. 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 $\mu \in (0, 1]$. ◊

Definition 2.4. Let $\mathcal{P}$ be an ℓ-step distribution and $j \in [\ell]$. The j-th marginal of $\mathcal{P}$ is defined as:

$$\mathcal{P}^{(j)}(x) := \sum_{x^{(1)},\ldots,x^{(j-1)},x^{(j+1)},\ldots,x^{(\ell)} \in \Omega} \mathcal{P}(x^{(1)}, \ldots, x^{(j-1)}, x, x^{(j+1)}, \ldots, x^{(\ell)})$$

for $x \in \Omega$. ◊

Definition 2.5. We say that $\mathcal{P}$ has *equal marginals*, if $\mathcal{P}^{(j)} = \mathcal{P}^{(j')}$ for every $j, j' \in [\ell]$. ◊

As mentioned, for same-set hitting we will always make the assumption that $\mathcal{P}$ has equal marginals. We then denote the marginal distribution by π. We refer to Section 2.2.4 for a discussion of this requirement.

Definition 2.6. Let $\mathcal{P}$ be an ℓ-step distribution. We let

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

◊

Remark 2.7. If $\alpha(\mathcal{P}) = 0$ then $\mathcal{P}$ is not same-set hitting. To see this, suppose that x^* is such that $\mathcal{P}(x^*, x^*, \ldots, x^*) = 0$. We set f to be the indicator function of $S := \{(x_1, \ldots, x_n) \mid x_1 = x^*\}$. Clearly, the probability in (7) equals 0. ◊

2.1.3 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 2.8. Let $\mathcal{P}$ be a single-coordinate distribution and let $A, B \subseteq [\ell]$. We define the respective *correlation* as

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

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

2.1.4 Influence

A crucial notion in the proof of Theorem 2.10 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 2.9. 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 $\underline{X}$.

2.2 Multi-Step Theorem: Statement, Proof Sketch and Discussion

This is the first of two sections that address Theorem 2.10. We first state the full theorem formally, then present a proof sketch of a simplified version and finally discuss the assumptions and how the theorem from [Mos10] is used in the proof.

2.2.1 Statement

Theorem 2.10. *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 $\mathrm{E}[f(\underline{X}^{(j)})] = \mu > 0$:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c\left(\alpha(\mathcal{P}), \rho(\mathcal{P}), \ell, \mu\right) , \tag{8}$$

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

Furthermore, whenever $\alpha(\mathcal{P}) > 0$ and $\rho(\mathcal{P}) < 1$, 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:

$$c(\alpha, \rho, \ell, \mu) := 1/\exp\left(\exp\left(\exp\left((1/\mu)^D\right)\right)\right) . \tag{9}$$

Corollary 2.11. *Let $\mathcal{P}$ be an ℓ-step distribution with equal marginals. If $\alpha(\mathcal{P}) > 0$ and $\rho(\mathcal{P}) < 1$, then $\mathcal{P}$ is same-set hitting.*

2.2.2 Proof sketch

In this section we briefly outline the proof of Theorem 2.10. For simplicity, we assume that the probability space is the one from Section 1.1.2, 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., $S_{\text{thr}} = \{\underline{x} : |\{i : x_i = 0\}| \geq n/3\}$.

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 $\epsilon > 0$ we have $\Pr[\mathcal{E}] \geq \epsilon$ and, at the same time, $\Pr[\underline{X} \in S \mid \mathcal{E}] \geq (1+\epsilon)\Pr[\underline{X} \in S]$.

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

After we are done, one can see that since no conditioning can increase the measure of S to $(1+\epsilon)\mu$, it must also be that no conditioning can *decrease* this measure to $(1-\epsilon')\mu$ for some other (constant) $\epsilon' > 0$.

Step 2 — eliminating high influences

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

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

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

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

Assume w.l.o.g. that $\mathrm{Inf}_i(T) \geq \tau$ for some $i \in [n]$. Given $z \in \{0,1,2\}$, let

$$T_z := \{(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) : (x_1, \ldots, x_n) \in T \wedge x_i = z\} \ .$$

Note that T_z is a subset of Ω^{n-1} rather than Ω^n. Therefore, the sets T_z can have non-empty intersections (while they would form a partition of T if considered as subsets of Ω^n). Furthermore, let $T_z^* := T_z \cup T_{z+1 \pmod 3}$.

The reduction is achieved by observing that there exists $z \in \{0,1,2\}$ such that, after conditioning on $X_i = Y_i = z$, the sum $\mu(S_z) + \mu(T_z^*)$ is strictly greater than the sum $\mu(S) + \mu(T)$:

$$\Pr[\underline{X} \in S_z \mid X_i = z] + \Pr[\underline{Y} \in T_z^* \mid Y_i = z] \geq \Pr[\underline{X} \in S] + \Pr[\underline{Y} \in T] + c(\tau) \ . \tag{10}$$

We choose to delete the coordinate i and replace S with $S' := S_z$ and T with $T' := T_z^*$. Equation (10) implies that after a constant number of such operations, neither S nor T has any remaining high-influence coordinates.

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

Step 3 – applying low-influence theorem from [Mos10]

Once we are left with two low-influence, somewhat-large sets S and T, we obtain $\Pr[\underline{X} \in S \wedge \underline{Y} \in T] \geq c(\mu) > 0$ by a straightforward application of a version of the theorem from [Mos10]. The details are discussed in the next section.

2.2.3 Low influence theorem from [Mos10]

A crucial ingredient in the proof of Theorem 2.10 is a slightly modified version of Theorem 1.14 from [Mos10]. The theorem says that $\rho(\mathcal{P}) < 1$ implies that the distribution $\mathcal{P}$ is set hitting for low-influence functions:

Theorem 2.12. *Let $\overline{\underline{X}}$ be a random vector distributed according to $(\overline{\underline{\Omega}}, \underline{\mathcal{P}})$ such that $\mathcal{P}$ has equal marginals, $\rho(\mathcal{P}) \leq \rho < 1$ and $\min_{x \in \Omega} \pi(x) \geq \alpha > 0$.*

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

$$\max_{i \in [n], j \in [\ell]} \mathrm{Inf}_i(f^{(j)}(\underline{X}^{(j)})) \leq \tau \ , \tag{11}$$

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

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq \left(\prod_{j=1}^{\ell} \mu^{(j)}\right)^{\ell/(1-\rho^2)} - \epsilon \ . \tag{12}$$

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

$$\tau := \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}}\right)^{C \frac{\ell \ln(\ell/\epsilon) \ln(1/\alpha)}{(1-\rho)\epsilon}} \ . \tag{13}$$

Theorem 2.12 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:

$$\alpha^* := \min_{(x^{(1)}, \ldots, x^{(\ell)}) \in \mathrm{supp}(\mathcal{P})} \mathcal{P}(x^{(1)}, \ldots, x^{(\ell)}) \ , \tag{14}$$

while our bound depends only on the smallest marginal probability:

$$\alpha = \min_{x \in \Omega} \pi(x) \ . \tag{15}$$

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 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 (15) and (14)). This better bound is then propagated in the proof of the invariance principle.

- Another change is unrelated to the dependency on the smallest marginal. For 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.

The (somewhat long) proof of Theorem 2.12 can be found in Appendix A.

2.2.4 Assumptions of the theorem

Equal distributions: unnecessary

In Theorem 2.10 we assume 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 $\rho(\mathcal{P}_i) \leq \rho$.

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 1.1, $\mathcal{P}$ is set hitting, and therefore also same-set hitting.

In case $\beta(\mathcal{P}) = 0$, we exhibit an example which shows that the expectation $\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right]$ 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:

$$S_1 = \{\underline{x} : x_1 = 1 \wedge |w_1(\underline{x}) - n/3| \leq 0.01n\} \tag{16}$$

$$S_2 = \{\underline{x} : x_1 = 0 \wedge |w_1(\underline{x}) - 2n/3| \leq 0.01n\}\ , \tag{17}$$

recalling that $w_1(\underline{x})$ is the number of ones in $\underline{x}$.

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 2.10 holds. However, because of the first coordinate we have $\Pr[\forall j : \underline{X}^{(j)} \in S] \leq \Pr[\underline{X}^{(1)} \in S_2 \vee \underline{X}^{(2)} \in S_1]$, and the right hand side is easily seen to be exponentially small.

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

The case $\rho = 1$: open question

Theorem 2.10 requires that $\rho < 1$ in order to give a meaningful bound. It is unclear whether this is an artifact of our proof or if it is necessary. In particular, consider the three step distribution $\mathcal{P}$ which picks a uniform triple from $\{000, 111, 222, 012, 120, 201\}$. One easily checks that $\rho(\mathcal{P}) = 1$ and that all marginals are uniform. We do not know if this distribution is same-set hitting.

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

Lemma 2.13. *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 2.12 fails).

Proof. We let

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

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. □

2.3 Proof of Multi-Step Theorem

The goal of this section is a proof of Theorem 2.10.

2.3.1 Properties of the correlation

Recall Definition 2.8. 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 2.14. Let $\mathcal{P}$ be a single-coordinate distribution and let $j \in [\ell]$. We call a collection of random variables $(\overline{X}^{\setminus j} = (X^{(1)}, \ldots, X^{(j-1)}, X^{(j+1)}, \ldots, X^{(\ell)}), Y, Z)$ 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

$$\pi(y)K(y, z) = \Pr[Y = y \wedge Z = z] = \Pr[Y = z \wedge Z = y] = \pi(z)K(z, y)\ , \quad (18)$$

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 2.15 (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,*

$$\lambda_2(K) = \max_{\substack{f:\Omega\to\mathbb{R} \\ \mathrm{E}[f(Y)]=0 \\ \mathrm{Var}[f(Y)]=1}} \mathrm{E}\left[f(Y)f(Z)\right]\ . \quad (19)$$

Lemma 2.16. *Let $\mathcal{P}$ be a single-coordinate distribution and let $(\overline{X}^{\setminus j}, Y, Z)$ 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 $\overline{X}^{\setminus j}$.

Consider first two functions f and g as in Definition 2.8 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 2.8. We will show that

$$\mathrm{Cov}\left[f(Y), g(\overline{X})\right]^2 \leq \lambda_2(K) \ , \tag{20}$$

and that there exists a choice of f and g that achieves equality in (20).

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

$$\begin{aligned}\mathrm{E}[f(Y)f(Z)] &= \sum_{\overline{x},y,z} \Pr[\overline{X} = \overline{x}] \Pr[Y = y \mid \overline{X} = \overline{x}] \Pr[Z = z \mid \overline{X} = \overline{x}] f(y) f(z) \\ &= \mathrm{E}[h(\overline{X})^2] \ . \end{aligned} \tag{21}$$

Now, by Cauchy-Schwarz, (21) and Lemma 2.15 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 (19) and $g := c \cdot h$ for some $c > 0$. □

For later use, we make the following implication of Lemma 2.16.

Corollary 2.17. *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 $f : \Omega \to \mathbb{R}$,*

$$\mathrm{E}[(f(Y) - f(Z))^2] \geq 2(1 - \rho^2)\,\mathrm{Var}[f(Y)]. \tag{22}$$

Proof. Assume w.l.o.g. that $\mathrm{E}[f(Y)] = 0$. By Lemmas 2.15 and 2.16,

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

□

2.3.2 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 2.18. 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

$$(\mathcal{R}f)(x_1, \dots, x_n) := f(y_1, \dots, y_n) \tag{23}$$

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 2.19. 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 $(1-\epsilon)\,\mathrm{E}[f(\underline{X})] \leq \mathrm{E}[\mathcal{R}f(\underline{X})] \leq (1+\epsilon)\,\mathrm{E}[f(\underline{X})]$. ◊

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

Definition 2.20. 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 $\mathrm{E}[\mathcal{R}f(\underline{X})] \leq (1+\epsilon)\,\mathrm{E}[f(\underline{X})]$. ◊

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

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

$$\mathrm{E}[f(\underline{X})] = \sum_{\mathcal{R}'} p(\mathcal{R}') \cdot \mathrm{E}\left[\mathcal{R}' f(\underline{X})\right] \ , \tag{24}$$

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,

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

and hence:

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

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

Lemma 2.22. *Let $\overline{\underline{X}}$ 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 $\epsilon \in (0,1], k \in \mathbb{N}$. Then, there exists a restriction $\mathcal{R}$ such that $g := (\mathcal{R}f)$ is ϵ-resilient up to size k and

$$\mathrm{E}[g(\underline{X}^{(1)})] \geq \mu \ , \tag{25}$$

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g(\underline{X}^{(j)})\right] \ , \tag{26}$$

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

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

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

Since the expectation of f only increases, we get (25). Finally, once the process stops, the resulting function is ϵ-resilient due to Lemma 2.21 (note that $\alpha(\pi) \geq \alpha$).

It remains to argue that (26) holds for the resulting function. Note first that the expectation cannot exceed 1, and hence the process will be repeated at most $p := \ln(1/\mu)/\ln(1+\epsilon') \leq \frac{2\ln(1/\mu)}{\epsilon'}$ 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 $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}$ are all equal to the value of $\mathcal{R}$ in the restricted coordinates of $\mathcal{R}$. 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

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

□

2.3.3 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 2.23. 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

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

$\Diamond$

The operator $\mathcal{M}[i,y,z]$ is useful for two reasons. First, if $\mathrm{Inf}_i(f^{(j)})$ is "large", then $\mathrm{E}\left[\mathcal{M}[i,y,z]f^{(j)}(\underline{X}^{(j)})\right] \geq \mathrm{E}[f^{(j)}(\underline{X}^{(j)})]+c$ 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 $\overline{x}^{\setminus j^*} = (x^{(1)},\ldots,x^{(j^*-1)},x^{(j^*+1)},\ldots,x^{(\ell)}), y, z \in \Omega$ both conditional probabilities $\Pr[X_i^{(j^*)} = y \mid \overline{X}_i^{\setminus j^*} = \overline{x}^{\setminus j^*}]$ and $\Pr[X_i^{(j^*)} = z \mid \overline{X}_i^{\setminus j^*} = \overline{x}^{\setminus j^*}]$ are "somewhat large" (larger than some constant). We imagine now that $\overline{X}_i^{(\setminus j^*)} = \overline{x}^{\setminus j^*}$ and that we also picked all values $\underline{X}_{\setminus i}^{(j^*)} = (X_1^{(j^*)},\ldots,X_{i-1}^{(j^*)},X_{i+1}^{(j^*)},\ldots,X_n^{(j^*)})$. 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 2.24. *Let $\overline{\underline{X}}$ be a random vector distributed according to $(\overline{\underline{\Omega}}, \overline{\underline{\mathcal{P}}})$. Fix $i \in [n]$, $j^* \in [\ell]$ and $\overline{x}^{\setminus j^*} = (x^{(1)},\ldots,x^{(j^*-1)},x^{(j^*+1)},\ldots,x^{(\ell)}), y, z \in \Omega$. Suppose that:*

$$\mathcal{P}(\overline{x}^{\setminus j^*}, y) \geq \beta , \tag{27}$$

$$\mathcal{P}(\overline{x}^{\setminus j^*}, z) \geq \beta . \tag{28}$$

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

$$g^{(j)} := \begin{cases} \mathcal{R}[i,x^{(j)}]f^{(j)} & \text{if } j \neq j^*, \\ \mathcal{M}[i,y,z]f^{(j)} & \text{if } j = j^*. \end{cases} \tag{29}$$

Then:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq \beta \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g^{(j)}(\underline{X}^{(j)})\right] . \tag{30}$$

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,

$$A = \begin{cases} y & \text{if } f\left(\underline{X}_{\setminus i}^{(j^*)}, y\right) > f\left(\underline{X}_{\setminus i}^{(j^*)}, z\right), \\ z & \text{otherwise.} \end{cases} \tag{31}$$

Consider now the event $\mathcal{E}$ which occurs if $\overline{X}_i = (\overline{x}^{\setminus j}, A)$. We will use $\mathbb{1}(\mathcal{E})$ to denote the function which is 1 if event $\mathcal{E}$ happens and 0 otherwise. Then, we get

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

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 $\overline{\underline{X}}_{\setminus i}$ the functions $g^{(j)}(\underline{X}^{(j)})$ are constant. Finally, the last inequality follows because by (27) and (28), for every choice of $\overline{\underline{X}}_{\setminus i} = (\overline{X}_1, \ldots, \overline{X}_{i-1}, \overline{X}_{i+1}, \ldots, \overline{X}_n)$ event $\mathcal{E}$ has probability at least β. □

The obvious idea for the next step would be to find values $\overline{x}^{\setminus j}, y, z$ 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 $f^{(j^*)}$ with $\mathcal{M}[i,y,z]f^{(j^*)}$, Lemma 2.24 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 $(\overline{X}^{\setminus j^*}, Y, Z)$ at random, as a double sample on coordinate j^* (cf. Definition 2.14). Let:

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

We will 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 2.25. *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

$$\mathrm{E}\left[\mathcal{M}[i,Y,Z]f(\underline{X})\right] \geq \mathrm{E}[f(\underline{X})] + \tau(1-\rho^2(\mathcal{P}))\,, \tag{32}$$

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 2.17,

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

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

$$\mathrm{E}\Big[\Big|f(\underline{X}_{\setminus i}, Y) - f(\underline{X}_{\setminus i}, Z)\Big|\Big] \geq 2(1-\rho^2)\,\mathrm{Inf}_i(f(\underline{X}_{\setminus i}, Y)) = 2\tau(1-\rho^2)\,. \tag{33}$$

Since Y and Z are symmetric (i.e., they define a reversible Markov chain, cf. remarks after Definition 2.14) and by (33),

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

as claimed. □

Lemma 2.26. *Let a random vector $\overline{X}$ be distributed according to $(\overline{\Omega}, \overline{\mathcal{P}})$ 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 $(\overline{X}^{(j^*)}, Y, Z)$ from $\mathcal{P}$ and let:

$$G^{(j)} := \begin{cases} \mathcal{R}[i, X^{(j)}] f^{(j)} & \text{if } j \neq j^* \\ \mathcal{M}[i, Y, Z] f^{(j^*)} & \text{if } j = j^*. \end{cases} \tag{34}$$

Then:

$$\mathrm{E}\left[\sum_{j=1}^{\ell} \mathrm{E}[G^{(j)}(\underline{X}^{(j)}) \mid G^{(j)}]\right] \geq \sum_{j=1}^{\ell} \mathrm{E}\left[f^{(j)}(\underline{X}^{(j)})\right] + \tau \cdot (1 - \rho^2)\,. \tag{35}$$

Note that (34) defines the functions $G^{(j)}$ as random variables which is why we use capital letters.

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

$$\mathrm{E}\left[\mathrm{E}[G^{(j)}(\underline{X}^{(j)}) \mid G^{(j)}]\right] = \mathrm{E}[f^{(j)}(\underline{X}^{(j)})]\,, \tag{36}$$

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

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

but this is exactly Lemma 2.25. □

Lemma 2.27. *Let $\overline{X}$ be a random vector distributed according to $(\overline{\Omega}, \overline{\mathcal{P}})$ and 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

$$g^{(j)} := \begin{cases} \mathcal{R}[i, x^{(j)}] f^{(j)} & \text{if } j \neq j^* \\ \mathcal{M}[i, y, z] f^{(j)} & \text{if } j = j^* \end{cases} \tag{37}$$

satisfy

$$\sum_{j=1}^{\ell} \mathrm{E}[g^{(j)}(\underline{X}^{(j)})] \geq \sum_{j=1}^{\ell} \mathrm{E}[f^{(j)}(\underline{X}^{(j)})] + \tau(1-\rho^2)/2\,, \tag{38}$$

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq \frac{\tau(1-\rho^2)}{2\ell|\Omega|^{\ell+1}} \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g^{(j)}(\underline{X}^{(j)})\right]\,. \tag{39}$$

While (38) is immediate from Lemma 2.26, we have to do a little bit of work to guarantee (39).

Proof. Choose $(\overline{X}^{\setminus j^*}, Y, Z)$ as a double sample from $\mathcal{P}$ and let $G^{(j)}$ be defined as in (34).

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 2.26, 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 $\Pr[\mathcal{E}] \leq |\Omega|^{\ell+1}\beta \leq \tau(1-\rho^2)/2\ell < 1$ and hence it makes sense to consider $\mathrm{E}[A \mid \neg\mathcal{E}]$. Bearing in mind the above and that $A \in [-\ell, \ell]$,

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

As a consequence, let us choose $(\overline{x}^{\setminus j^*}, y, z)$ such that $A \geq \tau(1-\rho^2)/2$ and $\mathcal{E}$ does not happen. (38) is now immediate, while for (39) 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 2.24. □

We can now repeat the process from Lemma 2.27 multiple times to get the result of this section.

Corollary 2.28. *Let $\overline{\underline{X}}$ be a random vector distributed according to $(\overline{\underline{\Omega}}, \underline{\mathcal{P}})$ 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 $\epsilon \in [0,1]$ 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* $j \in [\ell]$*:* $\mathrm{E}[g^{(j)}(\underline{X}^{(j)})] \geq (1-\epsilon)\,\mathrm{E}[f^{(j)}(\underline{X}^{(j)})]$.

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 2.27, 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 immediately from Lemma 2.27.

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. □

2.3.4 Finishing the proof

Proof of Theorem 2.10. Let us assume that $\mu \in (0, 0.99]$, the computations being only easier if this is not the case. To establish (9), 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 2.22, Corollary 2.28 and Theorem 2.12.

Given $f : \underline{\Omega} \to [0,1]$ with $\mathrm{E}[f(\underline{X}^{(1)})] = \mu$, first apply Lemma 2.22 to f with $\epsilon := 1/2$ 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.
- $\mathrm{E}[g(\underline{X}^{(1)})] \geq \mu$.
- $$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} g(\underline{X}^{(j)})\right], \tag{40}$$

where:

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

Next, apply Corollary 2.28. 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 2.28:

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

Therefore, Corollary 2.28 is applicable and yields $h^{(1)}, \ldots, h^{(\ell)} : \underline{\Omega} \to [0,1]$ such that:

- $\max_{j\in[\ell]} \max_{i\in[n]} \mathrm{Inf}_i(h^{(j)}(\underline{X}^{(j)})) \leq \tau$.
- $\forall j \in [\ell] : \mathrm{E}[h^{(j)}(\underline{X}^{(j)})] \geq \mu/2$.
- $$\mathrm{E}\left[\prod_{j=1}^{\ell} g(\underline{X}^{(j)})\right] \geq \beta \cdot \mathrm{E}\left[\prod_{j=1}^{\ell} h^{(j)}(\underline{X}^{(j)})\right] \ , \tag{41}$$

  where:

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

Finally, we need to apply Theorem 2.12. To this end, set:

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

and verify (13):

$$\begin{aligned} \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}}\right)^{O\left(\frac{\ln(\ell/\epsilon)\ln(1/\alpha)}{(1-\rho)\epsilon}\right)} &\geq \Omega\left(\epsilon\right)^{(O(1)+\ln 1/\epsilon)\cdot O(1/\epsilon)} \\ &\geq 1/\exp\left(\left(O(1)+\ln 1/\epsilon\right)^2 \cdot O\left(1/\epsilon\right)\right) \\ &\geq 1/\exp\left((1/\epsilon)^{O(1)}\right) \geq 1/\exp\left((1/\mu)^{O(1)}\right) \ . \end{aligned}$$

Hence, from Theorem 2.12:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} h^{(j)}(\underline{X}^{(j)})\right] \geq \epsilon/2 \geq \mu^{O(1)} \ . \tag{42}$$

(40), (41) and (42) 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. □

2.4 Proof for Two Steps

In this section we prove the classification of same-set hitting for two-step probability distributions:

Theorem 2.29. *Let Ω be a finite set and $\mathcal{P}$ a probability distribution over Ω^2 with equal marginals π. Let pairs (X_i, Y_i) be i.i.d. according to $\mathcal{P}$ for $i \in \{1, \ldots, n\}$.*

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

$$\mathrm{E}[f(\underline{X})f(\underline{Y})] \geq c\left(\alpha(\mathcal{P}), \mu\right) \ , \tag{43}$$

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

We remark that Theorem 2.29 does not depend on $\rho(\mathcal{P})$ in any way. This is in contrast to the case $\ell > 2$. To prove Theorem 2.29 we use the multi-step Theorem 2.10 and then handle the $\rho(\mathcal{P}) = 1$ case by convex decomposition.

Corollary 2.30. *A two-step probability distribution with equal marginals $\mathcal{P}$ is same-set hitting if and only if $\alpha(\mathcal{P}) > 0$.*

Proof. The "if" part follows from Theorem 2.29. The "only if" can be seen by taking f to be an appropriate dictator. □

Our goal in the rest of this section is to prove Theorem 2.29 assuming Theorem 2.10.

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.

2.4.1 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 pairwise disjoint sequence of $x_i \in \Omega$.

Definition 2.31. We call a probability distribution $\mathcal{C}$ over Ω an (s,p)-*cycle* if

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

◊

Lemma 2.32. *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 2.3.1). 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 $v_k := (1, \exp(\alpha_k \imath), \ldots, \exp((s-1)\alpha_k \imath))$ (where $\imath$ 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 2.16. □

2.4.2 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 $\mathcal{P} = \sum_{k=1}^{r} \beta_{(k)} \mathcal{P}_{(k)}$ and each distribution $\mathcal{P}_{(k)}$ is same-set hitting, then also $\mathcal{P}$ is same-set hitting.

Since we also need to apply this result in Section 2.7.2, the proof is more general than otherwise necessary.

Definition 2.33. Let $\mathcal{P}$ be a multi-step, single-coordinate probability distribution with equal marginals and let $\mathcal{P} = \sum_{k=1}^{r} \beta_{(k)} \cdot \mathcal{P}_{(k)}$ be its convex decomposition into r probability distributions, each of them with equal marginals.

For every $(z_1, \ldots, z_n) = \underline{z} \in [r]^n$ we let $\underline{\mathcal{P}}_{(\underline{z})}$ be the tensorized multi-step distribution such that its i-th coordinate is distributed independently according to the distribution $\mathcal{P}_{(z_i)}$. ◊

Definition 2.34. Let $\mathcal{P} = \sum_{k=1}^{r} \beta_{(k)} \cdot \mathcal{P}_{(k)}$ be a convex decomposition of an ℓ-step distribution with equal marginals into distributions with equal marginals and let $c : [0,1] \to [0,1]$ be a function such that $c(\mu) > 0$ for $\mu > 0$.

We say that a decomposition is *c-same-set hitting* if, for every function $f : \underline{\Omega} \to [0,1]$ and every vector $\underline{z} \in [r]^n$:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f\left(\underline{X}^{(j)}_{(\underline{z})}\right)\right] \geq c\left(\mu_{(\underline{z})}\right) \ ,$$

where the random vectors $\overline{\underline{X}}_{(\underline{z})}$ are distributed according to $\underline{\mathcal{P}}_{(\underline{z})}$ and

$$\mu_{\underline{z}} := \mathrm{E}\left[f\left(\underline{X}^{(1)}_{(\underline{z})}\right)\right] \ .$$

◊

It turns out that a distribution that has a same-set hitting convex decomposition is itself same-set hitting:

Lemma 2.35. *Let $\mathcal{P}$ be an ℓ-step distribution with equal marginals such that $\mathcal{P}$ has a c-same-set hitting convex decomposition.*

Then, $\mathcal{P}$ is same-set hitting. In particular, if $\overline{\underline{X}}$ is a random vector distributed according to $\mathcal{P}$, then for every function $f : \underline{\Omega} \to [0,1]$ with the expectation $\mathrm{E}\left[f(\underline{X}^{(1)})\right] = \mu > 0$:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq \mu/2 \cdot c(\mu/2) > 0 \ . \tag{44}$$

Furthermore, if $c(\mu) = \mu^{\alpha}$ for some $\alpha \geq 1$ we have

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq \mu^{\alpha} \ . \tag{45}$$

Proof. Let us write the relevant decomposition as $\mathcal{P} = \sum_{k=1}^{r} \alpha_{(k)} \mathcal{P}_{(k)}$. 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 $\Pr[Z_i = k] = \alpha_{(k)}$.
- For every vector $\underline{z} \in [r]^n$, conditioned on $\underline{Z} = \underline{z}$ the random vector $\overline{\underline{X}}$ is distributed according to $\underline{\mathcal{P}}_{(\underline{z})}$.

Recall that $\mu_{(\underline{z})} = \mathrm{E}[f(\underline{X}^{(1)}) \mid \underline{Z} = \underline{z}]$. 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

$$\Pr\left[\mu_{(\underline{Z})} \geq \mu/2\right] \geq \mu/2 \,. \tag{46}$$

Since the decomposition of $\mathcal{P}$ is c-same-set hitting, (46) implies

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

As for (45), we use Jensen's inequality on function μ^α:

$$\begin{aligned}\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] &= \mathrm{E}\left[\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)}) \mid \underline{Z}\right]\right] \\ &\geq \mathrm{E}\left[\mu_{(\underline{Z})}^{\alpha}\right] \geq \mathrm{E}\left[\mu_{(\underline{Z})}\right]^{\alpha} = \mu^{\alpha} \,.\end{aligned}$$

□

Definition 2.36. Let $\mathcal{P} = \sum_{k=1}^{r} \beta_{(k)} \cdot \mathcal{P}_{(k)}$ be a convex decomposition of an ℓ-step distribution with equal marginals into distributions with equal marginals.

We say that it is an (α, ρ)-convex decomposition if $\alpha(\mathcal{P}_{(k)}) \geq \alpha$ and $\rho(\mathcal{P}_{(k)}) \leq \rho$ for every $k \in [r]$. ◊

Lemma 2.37. *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 a random vector $\overline{\underline{X}}$ distributed according to $\mathcal{P}$ and 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. Due to Theorem 2.10[1], an (α, ρ)-convex decomposition of $\mathcal{P}$ is also $c(\alpha, \rho, \ell, \mu)$-same-set hitting for some universal function $c()$. By equation (44)

[1] Strictly speaking, Theorem 2.10 requires the distributions to be the same for each coordinate, which is not the case in our setting. However, this is not a problem, see Section 2.2.4.

in Lemma 2.35,

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

□

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

Definition 2.38. 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 2.39. *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. □

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 2.40. *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 2.39 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 $r \leq t^2$. Let $\beta_k := \min(w_k, \alpha(\mathcal{P})/t^2)$ 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 2.32, $\rho(\mathcal{C}_k) \geq 1 - 3\alpha(\mathcal{P})^5$ and, since every (s, p)-cycle has equal marginals, we obtained an (α, ρ)-convex decomposition of $\mathcal{P}$. □

2.4.4 Putting things together

Proof of Theorem 2.29. From Lemmas 2.40 and 2.37. □

Remark 2.41. One can see that see that, as in Theorem 2.10, 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

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

◊

2.5 Multiple Steps of a Markov Chain

In this section 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 2.42. 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*[2] if for every $j \in \{2, \ldots, \ell\}$ and $x^{(1)}, \ldots, x^{(j)} \in \Omega$ we have

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

$\Diamond$

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 2.29 to prove a bound that does not depend on $\rho(\mathcal{P})$:

Theorem 2.43. *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 $\mathrm{E}[f(\underline{X}^{(1)})] = \mu > 0$:

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] \geq c\left(\alpha(\mathcal{P}), \ell, \mu\right) \,, \tag{47}$$

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 $\mathrm{E}[f(\underline{X}^{(1)})] = \mu > 0$.

The proof is by induction on ℓ. For $\ell = 2$, apply Theorem 2.29 directly. For $\ell > 2$, define the function $g : \underline{\Omega} \to [0,1]$ as

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

Applying Theorem 2.29 for the distribution of the last two steps,

$$\mathrm{E}[g(\underline{X}^{(1)})] = \mathrm{E}[g(\underline{X}^{(\ell-1)})] = \mathrm{E}[f(\underline{X}^{(\ell-1)}) f(\underline{X}^{(\ell)})] \geq c(\alpha, \mu) > 0 \,. \tag{48}$$

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

Now we have

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f(\underline{X}^{(j)})\right] = \mathrm{E}\left[\left(\prod_{j=1}^{\ell-2} f(\underline{X}^{(j)})\right) g(\underline{X}^{(\ell-1)})\right] \tag{49}$$

$$\geq \mathrm{E}\left[\prod_{j=1}^{\ell-1} g(\underline{X}^{(j)})\right] \tag{50}$$

$$\geq c\left(\alpha, \ell-1, c(\alpha,\mu)\right) = c(\alpha,\ell,\mu) > 0, \tag{51}$$

where (49) holds since $\mathcal{P}$ is generated by Markov chains, (50) is due to $f \geq g$ pointwise and (51) is an application of the induction and (48). □

Remark 2.44. 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 $3(\ell-1)$. ◊

2.6 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 the following:

Theorem 2.45. *Let $\underline{X}$ 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

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq c(\mathcal{P}, \mu^{(1)}, \ldots, \mu^{(\ell)}) > 0\,. \tag{52}$$

We remark that for this theorem we do *not* require equal marginals. In the rest of the section we prove Theorem 2.45. First, from Corollary 2.28 and Theorem 2.12 it is easy to establish[3] the following:

[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 2.12 then depends on $\min_{j\in[\ell], x\in\mathrm{supp}(X^{(j)})} \pi^{(j)}(x)$).

Theorem 2.46. *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 $\epsilon \in [0,1)$.*

Then, for all $\mu^{(1)}, \ldots, \mu^{(\ell)} \in (0,1]$ there exists $k(\mathcal{P}, \epsilon, \mu^{(1)}, \ldots, \mu^{(\ell)}) \in \mathbb{N}$ 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

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq c(\mathcal{P}, \epsilon, \mu^{(1)}, \ldots, \mu^{(\ell)}) > 0 \,. \tag{53}$$

Definition 2.47. 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

$$f^{\subseteq S}(\underline{x}) := \mathrm{E}[f(\underline{x}_S, \underline{X}_{\overline{S}})] \,, \tag{54}$$

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]:

Claim 2.48. *Let π be a single-step distribution and $f : \underline{\Omega} \to \mathbb{R}$ and $S \subseteq [n]$. If $\underline{X}$ is distributed according to $\underline{\pi}$ and $\phi_0, \ldots, \phi_{m-1}$ form a Fourier basis for π and $f = \sum_{\sigma \in \mathbb{N}^n_{<m}} \hat{f}(\sigma)\phi_\sigma$, 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 2.49. *Let $\underline{X}$ be distributed according to a single-step distribution $\underline{\pi}$ with $\min_{x \in \Omega} \pi(x) \geq \alpha$ and let $\epsilon \in [0,1]$, $k \in \mathbb{N}$.*

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

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

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

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

But this gives

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

as required. □

Using Lemma 2.49 we can weaken the assumption in Theorem 2.46 such that it only requires that all Fourier coefficients of degree at most k are small:

Proof of Theorem 2.45. By Theorem 2.46, there must be some number $k := k(\mathcal{P}, \mu^{(1)}, \ldots, \mu^{(\ell)})$ such that if $f^{(1)}, \ldots, f^{(\ell)}$ are all 1/2-resilient up to size k, then (52) 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 2.48, if $\max_{\sigma:0<|\sigma|\leq k} |\hat{f}^{(j)}(\sigma)| \leq \gamma$, then for any $S \subseteq [n]$ with $|S| = k$ we have $\mathrm{Var}\left[(f^{(j)})^{\subseteq S}(\underline{X}^{(j)})\right] \leq 2^k\gamma^2$. With that in mind it is easy to choose γ such that Lemma 2.49 can be applied to each $f^{(j)}$. □

2.7 Polynomial Same-Set Hitting by Convexity

The property of set hitting establishes a lower bound on $\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right]$ 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 2.10 is triply exponentially small, and the bound from Theorem 1.2 is not even primitive recursive.

Recall the concept of polynomial set hitting (Definition 2.3). As a matter of fact, [MOS13] (cf. Theorem 1.1) 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.

In this section we present two techniques to establish polynomial same-set hitting. We start with a simple, folklore method that works for some symmetric distributions. Then, we harness reverse hypercontractivity from [MOS13] to handle all two-step, symmetric (i.e., with $\mathcal{P}(x, y) = \mathcal{P}(y, x)$) distributions.

2.7.1 "Meet in the middle"

We make an illustration of the first technique by applying it for binary distributions:

Theorem 2.50. *Fix $p \in [0, 1/2]$. Let $\mathcal{P}$ be a distribution over $\Omega = \{0, 1\}$ such that X_i is uniform and $Y_i = 1 - X_i$ with probability p and $X_i = Y_i$ otherwise.*

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

$$\mathrm{E}\left[f(\underline{X})f(\underline{Y})\right] \geq \mu^2 \ .$$

Proof. Let $q := (1 - \sqrt{1-2p})/2$. Let $Z_i, A_i, B_i \in \{0,1\}$ such that Z_i is uniform in $\{0,1\}$ and A_i and B_i be independent of each other with $\Pr[A_i = 1 - Z_i] = \Pr[B_i = 1 - Z_i] = q$. Since A_i and B_i are uniform and

$$\Pr[A_i \neq B_i] = 2q(1-q) = p \ ,$$

the distribution of (A_i, B_i) is the same as of (X_i, Y_i). Since conditioned on Z_i, the random variables A_i and B_i are independent of each other, and by Jensen's inequality,

$$\begin{aligned}\mathrm{E}\left[f(\underline{A})f(\underline{B})\right] &= \mathrm{E}\left[\mathrm{E}\left[f(\underline{A})f(\underline{B}) \mid \underline{Z}\right]\right] = \mathrm{E}\left[\mathrm{E}\left[f(\underline{A})^2 \mid \underline{Z}\right]\right] \\ &\geq \mathrm{E}\left[\mathrm{E}\left[f(\underline{A}) \mid \underline{Z}\right]^2\right] \geq \mathrm{E}\left[\mathrm{E}\left[f(\underline{A}) \mid \underline{Z}\right]\right]^2 = \mathrm{E}\left[f(\underline{A})\right]^2 = \mu^2 \ .\end{aligned}$$

□

Unfortunately, it is not generally true that a random variable like Z_i can be defined. In particular, this idea already does not work for the Boolean space defined above with the choice of $p \in (1/2, 1]$.

2.7.2 Symmetric two-step distributions

In this section we prove polynomial same-set hiting for two-step symmetric distributions:

Theorem 2.51. *Let a two-step probability distribution* $\mathcal{P}$ *be symmetric, i.e.,* $\mathcal{P}(x,y) = \mathcal{P}(y,x)$ *for all* $x, y \in \Omega$. *Additionally, let*

$$\alpha' := \alpha'(\mathcal{P}) := \min_{x \in \Omega} \mathcal{P}(x,x)/\pi(x) > 0 \ .$$

Then, $\mathcal{P}$ *is polynomially same-set hitting. In particular, if* $(\underline{X}, \underline{Y})$ *is a random vector distributed according to* $\mathcal{P}$*, then for every function* $f : \underline{\Omega} \to [0,1]$ *such that* $\mathrm{E}[f(\underline{X})] = \mu$*:*

$$\mathrm{E}\left[f(\underline{X}) \cdot f(\underline{Y})\right] \geq \mu^{8/\alpha'} \ .$$

Note that a two-step distribution is symmetric if and only if it has equal marginals. Furthermore, $\alpha'(\mathcal{P}) > 0$ if and only if $\alpha(\mathcal{P}) > 0$.

In the rest of this section we prove Theorem 2.51. In order to do that, we need the following quantitative version of Theorem 1.1:

Theorem 2.52 ([MOS13], Lemma 8.3)**.** *Let $(\underline{X}, \underline{Y})$ be a random vector distributed according to a two-step tensorized distribution with equal marginals $\underline{\mathcal{P}} = (\mathcal{P}_1, \ldots, \mathcal{P}_n)$ over an alphabet $\underline{\Omega} = (\Omega_1, \ldots, \Omega_n)$ and let*

$$\gamma := \min_{i=1}^{n} \min_{x,y \in \Omega_i} \mathcal{P}_i(x,y)/\pi_i(x) > 0 \,. \tag{55}$$

Then, for all functions $f, g : \underline{\Omega} \to [0,1]$ such that $\mathrm{E}[f(\underline{X})], \mathrm{E}[g(\underline{Y})] \geq \mu \geq 0$ we have

$$\mathrm{E}\left[f(\underline{X}) \cdot g(\underline{Y})\right] \geq \mu^{4/\gamma} \,.$$

Note that Theorem 2.52 does not assume that the coordinates have the same distribution.

Proof of Theorem 2.51. We make a convex decomposition of $\mathcal{P}$ into distributions with support sizes one and two:

$$\mathcal{P} = \sum_{x \in \Omega} \beta_{(x)} \cdot \mathcal{P}_{(x)} + \sum_{\substack{e = \{x,y\} \in \binom{\Omega}{2} \\ \mathcal{P}(x,y) > 0}} \beta_{(e)} \cdot \mathcal{P}_{(e)} \,.$$

Distributions $\mathcal{P}_{(e)}$ for $e = \{x, y\}$ are defined such that

$$\begin{aligned}
\beta_{(e)}\mathcal{P}_{(e)}(x,y) &:= \beta_{(e)}\mathcal{P}_{(e)}(y,x) := \mathcal{P}(x,y) \,, \\
\beta_{(e)}\mathcal{P}_{(e)}(x,x) &:= \mathcal{P}(x,y) \cdot \mathcal{P}(x,x)/\pi(x) \,, \\
\beta_{(e)}\mathcal{P}_{(e)}(y,y) &:= \mathcal{P}(x,y) \cdot \mathcal{P}(y,y)/\pi(y) \,.
\end{aligned}$$

Distributions $\mathcal{P}_{(x)}$ get the rest of the weight, i.e.,

$$\beta_{(x)}\mathcal{P}_{(x)}(x,x) := \mathcal{P}(x,x) - \sum_{y \in \Omega, y \neq x} \mathcal{P}(x,y) \cdot \mathcal{P}(x,x)/\pi(x) \,.$$

One checks that $\beta_{(x)}\mathcal{P}_{(x)}(x,x) \geq 0$ and therefore we defined a valid convex decomposition of $\mathcal{P}$. Furthermore, it is easy to see that all distributions $\mathcal{P}_{(e)}$ and $\mathcal{P}_{(x)}$ have equal marginals.

Next, we turn to evaluating γ defined in (55). To this end, we see that for every $e = \{x, y\}$:

$$\frac{\mathcal{P}_{(e)}(x,y)}{\pi_{(e)}(x)} = \frac{\mathcal{P}_{(e)}(x,y)}{\mathcal{P}_{(e)}(x,y) + \mathcal{P}_{(e)}(x,x)} = \frac{1}{1 + \mathcal{P}(x,x)/\pi(x)} \geq \frac{1}{2} \geq \frac{\alpha'}{2}$$

and

$$\frac{\mathcal{P}_{(e)}(x,x)}{\pi_{(e)}(x)} = \frac{\mathcal{P}_{(e)}(x,x)}{\mathcal{P}_{(e)}(x,y)+\mathcal{P}_{(e)}(x,x)} = \frac{\mathcal{P}(x,x)/\pi(x)}{1+\mathcal{P}(x,x)/\pi(x)} \geq \frac{\mathcal{P}(x,x)}{2\pi(x)} \geq \frac{\alpha'}{2} .$$

Consequently, applying Theorem 2.52 with $\lambda \geq \alpha'/2$ we conclude that our convex decomposition is $\mu^{8/\alpha'}$-same-set hitting (cf. Definition 2.34). By equation (45) from Lemma 2.35, we have

$$\mathrm{E}\left[f(\underline{X}) \cdot f(\underline{Y})\right] \geq \mu^{8/\alpha'} ,$$

as claimed. □

2.8 Conjecture with $\rho = 1$ for Simple Functions

One can conjecture that an ℓ-step distribution with equal marginals $\mathcal{P}$ is same-set hitting if and only if $\alpha(\mathcal{P}) > 0$. Since a variant of Szemerédi's theorem for finite groups (cf. Theorem 1.2) is a special case of this conjecture, its proof might be difficult to find. In this section we offer a modest sanity check instead. Specifically, we show that $\alpha(\mathcal{P}) > 0$ implies same-set hitting for some simple families of functions: dictators, linear functions and thresholds. We state those results in Lemmas 2.53, 2.54 and 2.56.

Lemma 2.53. *Let $\overline{\underline{X}}$ be a random vector distributed according to $(\Omega, \mathcal{P})$. Let $S \subseteq \underline{\Omega}$ be such that*

$$S = \{\underline{x} : x_i = y\}$$

for some $i \in [n]$ and $y \in \Omega$. Then,

$$\Pr\left[\forall j \in [\ell] : \underline{X}^{(j)} \in S\right] \geq \alpha(\mathcal{P}) .$$

In particular, if $\alpha(\mathcal{P}) > 0$, then $\mathcal{P}$ is same-set hitting for S.

Proof. By definition of $\alpha(\mathcal{P})$. □

Lemma 2.54. *Let $\overline{\underline{X}}$ be a random vector distributed according to $(\Omega, \mathcal{P})$, where $\Omega = \mathbb{Z}_r$ for some $r \in \mathbb{N}_{>0}$. Assume that $\alpha(\mathcal{P}) > 0$ and let $\beta' := \min_{\overline{x}:\mathcal{P}(\overline{x})>0} \mathcal{P}(\overline{x})$. Let $S \subseteq \underline{\Omega}$ be such that*

$$S = \left\{\underline{x} : \sum_{i=1}^{n} x_i = z\right\}$$

for some $z \in [r]$. Then,

$$\Pr\left[\forall j \in [\ell] : \underline{X}^{(j)} \in S\right] \geq (\beta')^{r^\ell} \ .$$

In particular, $\mathcal{P}$ is same-set hitting for S.

Proof. For $x \in \Omega$, let $x^\ell := (x, \ldots, x)$. If $n < r^\ell$, we have

$$\Pr\left[\forall j \in [\ell] : \underline{X}^{(j)} \in S\right] \geq$$
$$\Pr\left[\overline{X}_1 = z^\ell \wedge \overline{X}_2 = \ldots = \overline{X}_n = 0^\ell\right] \geq \alpha(\mathcal{P})^n \geq (\beta')^{r^\ell} \ .$$

If $n \geq r^\ell$, let $n' := n - r^\ell$. We are going to show that

$$\Pr\left[\forall j \in [\ell] : \underline{X}^{(j)} \in S \mid \overline{X}_1, \ldots, \overline{X}_{n'}\right] \geq (\beta')^{r^\ell} \ . \tag{56}$$

In other words, we claim that regardless of an assignment to $\overline{X}_1, \ldots, \overline{X}_{n'}$, there is always at least $(\beta')^{r^\ell}$ conditional probability of hitting S in all steps.

For $i \in \{0, \ldots, n\}$ and $j \in [\ell]$, let $Y_i^{(j)} := \sum_{i'=1}^{i} X_{i'}^{(j)}$ (recall that the arithmetic is in $\mathbb{Z}_r$). Note that the set S is hit in all steps if and only if $\overline{Y}_n = z^\ell$.

Consider a Markov chain where the state space is $[r]^\ell$ and for every $\overline{x}, \overline{y} \in [r]^\ell$ we set the transition probability from $\overline{x}$ to $\overline{x+y} := (x^{(1)} + y^{(1)}, \ldots, x^{(\ell)} + y^{(\ell)})$ to $\mathcal{P}(\overline{y})$. Starting in the state 0^ℓ, an i-step random walk on this Markov chain models the distribution of $\overline{Y}_i$.

Note that for every state $\overline{x}$ that is accessible from 0^ℓ in this Markov chain, it is also possible to get back from $\overline{x}$ to 0^ℓ. This is because for every pair of states $\overline{x}$, $\overline{x+y}$ with non-zero transition probability, one can get back from $\overline{x+y}$ to $\overline{x}$ by sampling $\overline{y}$ for $r-1$ more times in a row.

Consequently, for every $\overline{x}, \overline{y} \in [r]^\ell$ accessible from 0^ℓ, there must be a way to get from $\overline{x}$ to $\overline{y}$ in $k \leq r^\ell$ steps with probability at least $(\beta')^k$. Since it is possible to sample 0^ℓ in the remaining $r^\ell - k$ coordinates, there is also a way to get from $\overline{x}$ to $\overline{y}$ in exactly r^ℓ steps with probability at least $(\beta')^{r^\ell}$.

But then, this proves (56) (note that since $\alpha(\mathcal{P}) > 0$, z^ℓ is accessible from 0^ℓ in one step). □

For the threshold case we need a reverse Chernoff bound:

Theorem 2.55 (Lemma 4 in [KY15])**.** *Let $\underline{X}$ be an i.i.d. random vector over $\{0,1\}^n$ with $\Pr[X_i = 1] = p$ and let $p^* := \min(p, 1-p)$. For every $\epsilon \in (0, p^*/2]$ and $n \geq 3/\epsilon^2$:*

$$\Pr\left[\sum_{i=1}^{n} X_i \geq (p+\epsilon)n\right] \geq \exp\left(-9\epsilon^2 n/p^*\right) \ .$$

Lemma 2.56. *Let $\overline{\underline{X}}$ be a random vector distributed according to $(\Omega, \mathcal{P})$ with equal marginals such that $\alpha := \alpha(\mathcal{P}) > 0$. There exists a constant $c_\alpha > 0$ such that for every $y \in \Omega$ with $\pi(y) = p$ and for a set*

$$S = \{\underline{x} : w_y(\underline{x}) \geq pn\}$$

we have

$$\Pr\left[\forall j \in [\ell] : \underline{X}^{(j)} \in S\right] \geq c_\alpha \ .$$

Proof. Consider the random vector $\overline{\underline{X}}$ as a string of length n over alphabet $\overline{\Omega} = \Omega^\ell$ and let $q = \mathcal{P}(y^\ell)$. The high-lever overview of the proof is as follows: We show that in the vector $\overline{\underline{X}}$ with constant probability:

- The number of coordinates where y^ℓ is sampled is slightly more than the expectation qn.
- For every $\overline{x} \neq y^\ell$, the number of coordinates where $\overline{x}$ is sampled is roughly the same as expected.

Those two conditions ensure that S is hit in all ℓ steps.

We proceed with a detailed proof. By Theorem 2.55 with $\epsilon = 2/\sqrt{n}$ we have (for n big enough)

$$\Pr\left[w_{y^\ell}\left(\overline{\underline{X}}\right) \geq qn + 2\sqrt{n}\right] \geq \exp(-36/q^*) \geq c_\alpha > 0 \ .$$

Let $A := w_{y^\ell}\left(\overline{\underline{X}}\right)$ and assume $A = a \geq qn + \sqrt{n}$. Fix $\overline{x} \neq y^\ell$ with $\mathcal{P}(\overline{x}) = q_{\overline{x}} > 0$. By Chernoff bound,

$$\begin{aligned}\Pr\left[w_{\overline{x}}\left(\overline{\underline{X}}\right) > \left(q_{\overline{x}} + 1/|\Omega|^\ell\right) n \mid A = a\right] &\leq \Pr\left[w_{\overline{x}}\left(\overline{\underline{X}}\right) > \left(q_{\overline{x}} + 1/|\Omega|^\ell\right) n\right] \\ &\leq \exp\left(-2n/|\Omega|^{2\ell}\right) \ .\end{aligned}$$

Define the event $\mathcal{E} :\equiv \left(\forall \overline{x} \neq y^\ell : w_{\overline{x}}(\overline{\underline{X}}) \leq \left(q_{\overline{x}} + 1/|\Omega|^\ell\right) n\right)$. By union bound,

$$\Pr\left[\mathcal{E} \mid A \geq qn + 2\sqrt{n}\right] \geq 1 - |\Omega|^\ell \exp\left(-2n/|\Omega|^{2\ell}\right) \geq 1/2 \ ,$$

where the last inequality holds for n big enough. Finally, note that the event $(A \geq qn + 2\sqrt{n}) \wedge \mathcal{E}$ implies $\forall j \in [\ell] : \underline{X}^{(j)} \in S$. Therefore, we have

$$\Pr\left[\forall j \in [\ell] : \underline{X}^{(j)} \in S\right] \geq c_\alpha$$

for n big enough. The case of small n is easily handled by estimating

$$\Pr\left[\forall j \in [\ell] : \underline{X}^{(j)} \in S\right] \geq \mathcal{P}(y^\ell)^n \geq c_\alpha \ .$$

□

Remark 2.57. All three lemmas above can be generalized to a certain extent at the expense of more complicated proofs. ◊

Chapter 3

Parallel Repetition

3.1 Preliminaries

We start this chapter with a preliminary section to set-up the stage for our results. In Section 3.1.1 we give formal definitions of our most important concepts. In Section 3.1.2 we show a family of non-trivial games whose value does not decrease after a constant number of repetitions. Section 3.1.3 presents the argument of Verbitsky that density Hales-Jewett theorem implies that all question sets admit parallel repetition, while Section 3.1.4 argues that our assumption that a question is sampled from $\overline{Q}$ uniformly is without loss of generality. All those results are known, but are included here for completeness.

3.1.1 Multi-prover games and parallel repetition

In this section we provide formal definitions of the concepts we are concerned with in this chapter: multi-prover games and parallel repetition.

Definition 3.1 (Multi-prover games)**.** An *r-prover game* $\mathcal{G} = (\overline{Q}, \overline{A}, V)$ consists of the following elements:

- $\overline{Q} \subseteq Q^{(1)} \times \ldots \times Q^{(r)}$ is a finite *question set*.

  Note that $\overline{Q}$ does not have to consist of all possible tuples. However, we will always assume that there are no "impossible questions", i.e., that for each element of a *question alphabet* $q^{(j)} \in Q^{(j)}$ there exists at least one question tuple $\overline{q} \in \overline{Q}$ with $q^{(j)}$ as its j-th element.

- $\overline{A} = A^{(1)} \times \ldots \times A^{(r)}$ is a finite *answer set*.

- $V : \overline{Q} \times \overline{A} \to \{0, 1\}$ is a *verification predicate*.

A *strategy* $\overline{\mathcal{S}} = (\mathcal{S}^{(1)}, \ldots, \mathcal{S}^{(r)})$ for a game $\mathcal{G}$ is a tuple of functions, where $\mathcal{S}^{(j)} : Q^{(j)} \to A^{(j)}$.

Let $\overline{q} = (q^{(1)}, \ldots, q^{(r)})$ be a random variable sampled uniformly from $\overline{Q}$. For a strategy $\overline{\mathcal{S}}$ let $\overline{\mathcal{S}}(\overline{q}) := (\mathcal{S}^{(1)}(q^{(1)}), \ldots, \mathcal{S}^{(r)}(q^{(r)}))$.

We define the *value* of a game $\mathcal{G}$ as

$$\mathrm{val}(\mathcal{G}) := \max_{\overline{\mathcal{S}}} \Pr\left[V\left(\overline{q}, \overline{\mathcal{S}}(\overline{q})\right) = 1\right] .$$

We say that a game is *trivial* if its value is 1.

We also say it is *free* if $\overline{Q} = Q^{(1)} \times \ldots \times Q^{(r)}$. Note that in our setting this is equivalent to the property that the provers' questions are distributed independently. ◊

Definition 3.2 (Parallel repetition)**.** The *n-fold parallel repetition* $\mathcal{G}^n$ of an r-prover game $\mathcal{G} = (\overline{Q}, \overline{A}, V)$ is the game $\mathcal{G}^n = (\underline{Q}, \underline{A}, \underline{V})$ where

- The question set for the j-th prover $\underline{Q}^{(j)} := \left(Q^{(j)}\right)^n$ is the n-fold product of the original $Q^{(j)}$. Consequently, the question set $\underline{Q}$ is the n-fold product of $\overline{Q}$.
- In the same way, the answer set for the j-th prover $\underline{A}^{(j)}$ is the n-fold product of $A^{(j)}$.
- The verification predicate $\underline{V}$ accepts if and only if all of its n single instances accept:

$$\underline{V}\left(\underline{\overline{q}}, \underline{\overline{a}}\right) = 1 \iff \forall i \in [n] : V\left(\overline{q}_i, \overline{a}_i\right) = 1 .$$

◊

Claim 3.3. *For every r-prover game $\mathcal{G}$ and every $n \in \mathbb{N}$:* $\mathrm{val}(\mathcal{G})^n \leq \mathrm{val}(\mathcal{G}^n) \leq \mathrm{val}(\mathcal{G})$*. In particular, $\mathcal{G}$ is trivial if and only if $\mathcal{G}^n$ is trivial.*

Definition 3.4 (Forbidden subgraph bounds)**.** For an r-prover question set $\overline{Q}$ we define the *parallel repetition threshold* $\omega_{\overline{Q}}(n) := \max_{\mathcal{G}} \mathrm{val}(\mathcal{G}^n)$, where the maximum is over all non-trivial games $\mathcal{G}$ with question set $\overline{Q}$.

We say that $\overline{Q}$ *admits parallel repetition* if $\lim_{n\to\infty} \omega_{\overline{Q}}(n) = 0$. We say that $\overline{Q}$ *admits exponential parallel repetition* if there exists $C_{\overline{Q}} < 1$ such that for every $n \in \mathbb{N}$:

$$\omega_{\overline{Q}}(n) \leq (C_{\overline{Q}})^n .$$

◊

3.1.2 Parallel repetition example

We present an example family of games with $\mathrm{val}(\mathcal{G}^r) = \mathrm{val}(\mathcal{G})$. The example is taken from [Fei95] and credited to Ran Raz, but we discuss it here for the convenience of the reader.

Definition 3.5. For $r \geq 3$, $m \geq 2$ the r-prover game $\mathcal{G}_{r,m}$ is defined as follows:

The question alphabet for the j-th prover is $Q^{(j)} := \mathbb{Z}_m$. The answer alphabet for the j-th prover consists of pairs $A^{(j)} := [r] \times \mathbb{Z}_m$, where we interpret the first element as a pointer to one of the provers.

Given questions $x^{(1)}, \ldots, x^{(r)}$ and answers $(p^{(1)}, y^{(1)}), \ldots, (p^{(r)}, y^{(r)})$ the provers win if:

- $p^{(1)} = \ldots = p^{(r)} =: p$. That is, we have a *special prover* that everyone points to.
- $x^{(p)} = y^{(p)}$.
- $\sum_{j=1}^{r} y^{(j)} = 0 \pmod{m}$.

$\Diamond$

Claim 3.6. $\mathrm{val}(\mathcal{G}_{r,m}) = 1/m$.

Proof. To achieve value $1/m$, given a question $x^{(j)}$, the j-th prover can output $(1, x^{(j)})$. Then, the provers win if and only if $\sum_{j=1}^{r} x^{(j)} = 0 \pmod{m}$, which happens with probability $1/m$.

On the other hand, fix a strategy for the provers and consider a question tuple for which they win. Let the special prover for this question tuple be p and the question to the special prover $x^{(p)}$. If we exchange the special question in this tuple to any other value, the provers must lose. Therefore, given a tuple for which the provers win, we found $m - 1$ tuples for which they lose. Since one can see that this association is injective, it must be that $\mathrm{val}(\mathcal{G}_{r,m}) \leq 1/m$. $\square$

Claim 3.7. $\mathrm{val}(\mathcal{G}^r_{r,m}) = 1/m$.

Proof. Given r numbers $x_1^{(j)}, \ldots, x_r^{(j)}$, the j-th prover responds with $(1, x_j^{(j)})$, $\ldots, (r, x_j^{(j)})$. The provers succeed in the repeated game if and only if it holds that $\sum_{j=1}^{r} x_j^{(j)} = 0 \pmod{m}$, which happens with probability $1/m$. $\square$

3.1.3 All question sets admit parallel repetition

In this section we present the proof from [Ver96] showing that all question sets admit parallel repetition. We include it for completeness and because it is a good illustration of the forbidden subgraph method.

The proof uses the famous density Hales-Jewett theorem:

Definition 3.8 (Combinatorial line)**.** Let $r, n \in \mathbb{N}_{>0}$. A *combinatorial pattern* over $[r]^n$ is a string

$$(b_1, \ldots, b_n) = \underline{b} \in ([k] \cup \{\star\})^n \setminus [r]^n \ ,$$

where $\star$ is a special symbol called the *wildcard*. Note that a pattern has to contain at least one wildcard.

For $q \in [r]$ we let $\underline{b}(q) \in [r]^n$ to be the string formed from $\underline{b}$ by substituting all occurrences of the wildcard with q.

A *combinatorial line* associated with a pattern $\underline{b}$ is the set $\{\underline{b}(1), \ldots, \underline{b}(r)\}$. ◊

Example 3.9. For $r = 3$, $n = 5$, an example pattern is $\underline{b} = 12\star2\star$.

The corresponding combinatorial line is $L = \{12121, 12222, 12323\}$. ◊

Definition 3.10. Let $r \in \mathbb{N}_{>0}$. For a set $S \subseteq [r]^n$ we define its *measure* as $\mu(S) := |S|/r^n$.

We then let the *density Hales-Jewett threshold* $\omega_r^{\mathrm{DHJ}}(n)$ to be the maximum measure of a subset of $[r]^n$ that does not contain a combinatorial line. ◊

The density Hales-Jewett theorem states that sets of constant measure contain combinatorial lines for n big enough:

Theorem 3.11 (Density Hales-Jewett theorem, [FK91], for the proof see also [Pol12])**.** *Let $r \in \mathbb{N}_{>0}$. Then,* $\lim_{n\to\infty} \omega_r^{\mathrm{DHJ}}(n) = 0$.

Theorem 3.12 ([Ver96])**.** *Let $\overline{Q}$ be a question set with $|Q| = r$. Then:*

$$\omega_{\overline{Q}}(n) \leq \omega_r^{\mathrm{DHJ}}(n) \ . \tag{57}$$

In particular, $\overline{Q}$ admits parallel repetition.

Proof. The "in particular" part follows from (57) and Theorem 3.11. Therefore, we only need to prove (57).

To this end, fix $\overline{Q}$, $n \in \mathbb{N}_{>0}$ and a game $\mathcal{G}$ with question set $\overline{Q}$. Note that it is enough to show that if $\mathrm{val}(\mathcal{G}^n) > \omega_r^{\mathrm{DHJ}}(n)$, then $\mathcal{G}$ is trivial.

Fix a strategy $\overline{\underline{\mathcal{S}}} = (\underline{\mathcal{S}}^{(1)}, \ldots, \underline{\mathcal{S}}^{(r)})$ for the repeated game that achieves a value greater than $\omega_r^{\mathrm{DHJ}}(n)$. Let $H \subseteq \overline{Q}$ be the set of question tuples for the repeated game on which the provers win using $\overline{\underline{\mathcal{S}}}$. By Theorem 3.11, H contains a combinatorial line with an associated pattern $\overline{\underline{b}} = (\overline{b}_1, \ldots, \overline{b}_n)$.

We construct a strategy for $\mathcal{G}$ as follows: given a question $q \in Q^{(j)}$, the j-th prover proceeds as follows: It computes $\underline{a}^{(j)} := \underline{\mathcal{S}}^{(j)}(q_1^{(j)}, \ldots, q_n^{(j)})$, where $q_i^{(j)} = b_i^{(j)}$ if $\overline{b}_i \in \overline{Q}$ and $q_i^{(j)} = q$ if $\overline{b}_i = \star$. Then, it returns $a_i^{(j)}$, where i is the first coordinate of $\overline{\underline{b}}$ with $\overline{b}_i = \star$.

Since H contains the combinatorial line of $\overline{\underline{b}}$, for every question tuple $\overline{q} \in \overline{Q}$ the provers win the repeated game on all coordinates. In particular, they win on the i-th coordinate, but the game played there is exactly $\mathcal{G}$. □

3.1.4 Reduction of general parallel repetition to uniform case

We show how parallel repetition for a question distribution that is not necessarily uniform over a question set $\overline{Q}$ reduces to the uniform case. The proof is taken from [FV02].

Theorem 3.13. *Let $\overline{Q}$ be an r-prover question set and let $\mathcal{Q}$ be a probability distribution with support $\overline{Q}$ such that $\epsilon := \min_{\overline{q} \in \overline{Q}} \mathcal{Q}(\overline{q})$ and*

$$\alpha := \frac{\epsilon}{1/\left|\overline{Q}\right|} = \epsilon \left|\overline{Q}\right| \ .$$

Furthermore, assume that a function $f : \mathbb{N}_{>0} \to [0,1]$ is such that for every non-trivial game $\mathcal{H}$ uniform over $\overline{Q}$:

$$\mathrm{val}(\mathcal{H}^n) \leq f(n) \ .$$

Then, for every non-trivial game $\mathcal{G}$ such that its questions are sampled according to $\mathcal{Q}$ we have

$$\mathrm{val}(\mathcal{G}^n) \leq \exp\left(-\alpha^2 n/2\right) + f\left(\alpha n/2\right) \ .$$

Proof. First, note that we can write $\mathcal{Q} = \alpha U_{\overline{Q}} + (1-\alpha)\mathcal{Q}'$, where $U_{\overline{Q}}$ is the uniform distribution over $\overline{Q}$ and $\mathcal{Q}'$ some other probability distribution. Consequently, we can define an i.i.d. random binary vector $\underline{B} = (B_1, \ldots, B_n)$ coupled with an execution of $\mathcal{G}^n$ such that $B_i = 1$ if the i-th question is sampled from $U_{\overline{Q}}$ and $B_i = 0$ if it was sampled from $\mathcal{Q}'$.

Consider an execution of $\mathcal{G}^n$ with a modified verifier. The new verifier first checks if $w_1(\underline{B}) \geq \alpha n/2$, i.e., if the number of coordinates with $B_i = 1$ is at least half of the expectation αn. She accepts if this check fails. If the

first check succeeds, the new verifier accepts if the single-coordinate verifier accepts on all coordinates with $B_i = 1$.

Let us call this modified game $(\mathcal{G}^n)'$. Clearly, $\mathrm{val}(\mathcal{G}^n) \leq \mathrm{val}((\mathcal{G}^n)')$. Furthermore, let $\mathcal{G}^*$ be a game with the same verifier as $\mathcal{G}$ but uniform over $\overline{Q}$. Note that $\mathcal{G}^*$ is non-trivial. Observe that conditioned on a choice of $\underline{B}$, the game $(\mathcal{G}^n)'$ is the same as $\mathcal{G}^*$ repeated $w_1(\underline{B})$ times. Consequently, and using Chernoff bound,

$$\begin{aligned}\mathrm{val}(\mathcal{G}^n) &\leq \mathrm{val}((\mathcal{G}^n)') = \mathrm{E}\left[\mathrm{val}((\mathcal{G}^n)') \mid \underline{B}\right] \leq \Pr\left[w_1(\underline{B}) < \alpha n/2\right] + f(\alpha n/2) \\ &\leq \exp(-\alpha^2 n/2) + f(\alpha n/2)\ .\end{aligned}$$

□

3.2 Constructability Implies Parallel Repetition

Observe that we can identify the question set of an r-prover game with an r-uniform, r-partite hypergraph. In this section we first define a class of constructible hypergraphs and then establish that all constructible question sets admit exponential parallel repetition. The main result of this section is Theorem 3.21.

3.2.1 Constructing hypergraphs by conditioning

We define constructability in the general case, but for intuition the reader is invited to think about bipartite graphs (i.e., $r = 2$).

Definition 3.14. Let $r \geq 2$. We consider r-uniform, r-partite finite hypergraphs $G = (Q^{(1)}, \ldots, Q^{(r)}, \overline{Q})$, where $\overline{Q} \subseteq Q^{(1)} \times \ldots \times Q^{(r)}$. We will often abuse the notation by identifying the hypergraph with its edge set $\overline{Q}$.

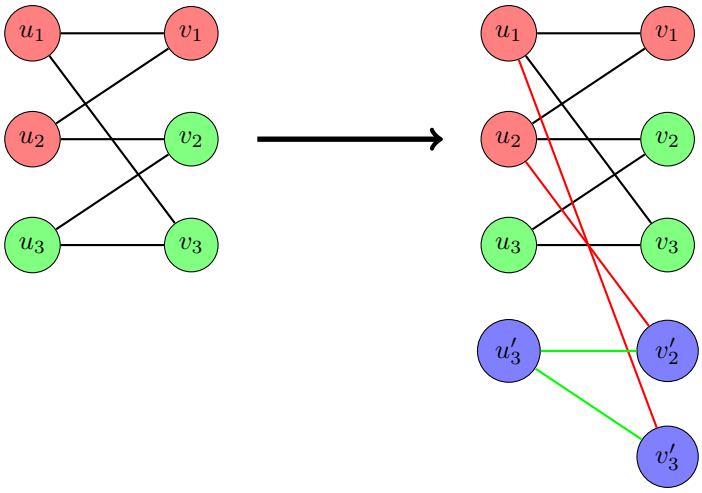
Given two hypergraphs $(Q^{(1)}, \ldots, Q^{(r)}, \overline{Q})$ and $(P^{(1)}, \ldots, P^{(r)}, \overline{P})$ we say that $f = (f^{(1)}, \ldots, f^{(r)})$, $f^{(j)} : Q^{(j)} \to P^{(j)}$ is a *homomorphism* from $\overline{Q}$ to $\overline{P}$ if $\overline{q} = (q^{(1)}, \ldots, q^{(r)}) \in \overline{Q}$ implies $f(\overline{q}) := (f^{(1)}(q^{(1)}), \ldots, f^{(r)}(q^{(r)})) \in \overline{P}$.

If f is a homomorphism from $\overline{Q}$ to $\overline{Q}$ we can just say that f is a homomorphism of $\overline{Q}$. We also define the *homomorphism space* $\mathrm{Hom}(\overline{Q}, \overline{P})$ as the set of homomorphisms from $\overline{Q}$ to $\overline{P}$. ◊

Definition 3.15. Given a hypergraph $(Q^{(1)}, \ldots, Q^{(r)}, \overline{Q})$ and sets $P^{(j)} \subseteq Q^{(j)}$ we define its *section hypergraph* $(P^{(1)}, \ldots, P^{(r)}, \overline{P})$, where $\overline{P} \subseteq \overline{Q}$ consists of those hyperedges whose vertices are all in $P^{(1)} \cup \ldots \cup P^{(r)}$. ◊

In the graph case the section hypergraph corresponds to an induced subgraph.

Figure 3.1: Doubling a bipartite graph. Fixed vertices in red, old vertices in green, new vertices in blue.



Definition 3.16. Let $r \geq 2$. We recursively define the class of r-partite hypergraphs that are *constructible by conditioning*:

1. A hyperedge $\left(\{q^{(1)}\}, \ldots, \{q^{(r)}\}, \{(q^{(1)}, \ldots, q^{(r)})\}\right)$ is constructible.

2. If a hypergraph $(P^{(1)} \dot\cup Q^{(1)}, \ldots, P^{(r)} \dot\cup Q^{(r)}, \overline{P})$ is constructible, then $(P^{(1)} \dot\cup Q^{(1)} \dot\cup R^{(1)}, \ldots, P^{(r)} \dot\cup Q^{(r)} \dot\cup R^{(r)}, \overline{P} \cup \overline{Q})$ is also constructible, where:

   - $R^{(j)} := \{q' : q \in Q^{(j)}\}$ is a set of copies of vertices from $Q^{(j)}$. We say that the vertices in $P^{(j)}$ are *fixed*, vertices from $Q^{(j)}$ are *old* and vertices from $R^{(j)}$ are *new*.
   - We say that a hyperedge is fixed if all of its vertices are fixed. For a hyperdge $\overline{q} \in \overline{P}$ that is not fixed we define $\overline{q}'$ as the hyperedge formed from $\overline{q}$ by replacing all of its old vertices by their respective copies.

     Then, $\overline{Q} := \left\{\overline{q}' : \overline{q} \in \overline{P}, \overline{q} \text{ is not fixed}\right\}$.

   In this case we say that $\overline{P} \cup \overline{Q}$ was constructed from $\overline{P}$ by *doubling* $Q^{(1)} \cup \ldots \cup Q^{(r)}$.

   Figure 3.1 can be consulted for an example in the graph case.

3. If a hypergraph $(P^{(1)} \dot\cup Q^{(1)}, \ldots, P^{(r)} \dot\cup Q^{(r)}, \overline{Q})$ is constructible and, at the same time, $(P^{(1)}, \ldots, P^{(r)}, \overline{P})$ is a section hypergraph of $\overline{Q}$ such that there exists a homomorphism from $\overline{Q}$ to $\overline{P}$ which is identity on $P^{(1)} \cup \ldots \cup P^{(r)}$, then $\overline{P}$ is also constructible.

   In such case we say that $\overline{Q}$ *collapses onto* $\overline{P}$.

$\Diamond$

Observe that the doubling operation never produces hyperedges incident to both old and new vertices.

To give some intuition on the conditioning operations we state two simple properties.

Claim 3.17. *Let $r \geq 2$. Every r-partite hypergraph can be collapsed onto one of its hyperedges.*

Definition 3.18. The *complete* hypergraph on $Q^{(1)}, \ldots, Q^{(r)}$ is given as $(Q^{(1)}, \ldots, Q^{(r)}, Q^{(1)} \times \ldots \times Q^{(r)})$. $\Diamond$

Claim 3.19. *Let $r \geq 2$ and $Q^{(1)}, \ldots, Q^{(r)}$ be finite sets. The complete hypergraph on $Q^{(1)}, \ldots, Q^{(r)}$ is constructible.*

Proof. Let $k^{(j)} := |Q^{(j)}|$, $Q^{(j)} = \{q_1^{(j)}, \ldots, q_{k^{(j)}}^{(j)}\}$ and $P^{(j)} := \{q_1^{(j)}\}$. Start the construction with a single hyperedge $(P^{(1)}, \ldots, P^{(r)}, P^{(1)} \times \ldots \times P^{(r)})$.

The complete hypergraph is constructed in r stages. In the j-th stage vertex $q_1^{(j)}$ is doubled $k^{(j)} - 1$ times with all other vertices fixed. Observe that after the j-th stage the current hypergraph is the complete hypergraph on $Q^{(1)}, \ldots, Q^{(j)}, P^{(j+1)}, \ldots, P^{(r)}$. $\square$

Remark 3.20. As a matter of fact, if $|Q^{(1)}| = \ldots = |Q^{(r)}| = 2^k$, then it is not difficult to see that the complete hypergraph on $Q^{(1)}, \ldots, Q^{(r)}$ can be constructed with rk doublings. $\Diamond$

3.2.2 Theorem statement

Our goal is:

Theorem 3.21. *Let $\overline{Q}$ be an r-prover question set that is constructible by conditioning using k doublings (and an arbitrary number of collapses). Let $M := \left|\overline{Q}\right|$. Then,*

$$\omega_{\overline{Q}}(n) \leq 3 \exp\left(-n/M^{2^{k+1}}\right) .$$

In particular, $\overline{Q}$ admits exponential parallel repetition.

A very rough proof outline is as follows: first, exponential parallel repetition is equivalent to exponential decrease of the threshold for good homomorphism vectors (cf. Definition 3.22). Second, we show that the existence of good homomorphism vectors is implied by existence of a probability distribution over $\mathrm{Hom}\left(\overline{Q},\overline{Q}\right)$ with certain same-set hitting properties. Third, we prove that such distribution exists for every constructible $\overline{Q}$.

We elaborate those three steps in the next subsections.

3.2.3 Good question sets

Definition 3.22. Let $\overline{Q}$ be an r-prover question set and let $\underline{\overline{Q}} := \overline{Q}^n$ be its n-fold parallel repetition. Let $S \subseteq \underline{\overline{Q}}$ with $\mu(S) = |S|/\left|\underline{\overline{Q}}\right|^n$ and let $\underline{f} = (f_1, \ldots, f_n)$ be a vector of n homomorphisms of $\overline{Q}$. We say that $\underline{f}$ is *good* for S if:

- For every $\overline{q} \in \overline{Q}$ we have that $\underline{f}(\overline{q}) := (f_1(\overline{q}), \ldots, f_n(\overline{q})) \in S$.
- There exists $i \in [n]$ such that f_i is identity.

We say that the question set $\overline{Q}$ is (n,ϵ)-good if for every $S \subseteq \underline{\overline{Q}}$ with $\mu(S) \geq \epsilon$ there exists a vector of homomorphisms that is good for S. ◊

Observe that a vector of n identities $\underline{f} = (\mathrm{Id}, \ldots, \mathrm{Id})$ is good for the whole space $\underline{\overline{Q}}$ and hence every question set $\overline{Q}$ is $(n,1)$-good.

Remark 3.23. Note that if $\overline{Q}$ is (n,ϵ)-good, then it is also $(n+1,\epsilon)$-good. This is because given $S \subseteq \overline{Q}^{n+1}$ we can set f_{n+1} to be a constant homomorphism such that the (relative) measure $\mu(S)$ does not decrease conditioned on $\overline{q}_{n+1} = f_{n+1}(\overline{q})$. Then we can get $f_1, \ldots, f_n$ from the assumption that $\overline{Q}$ is (n,ϵ)-good. ◊

Definition 3.24. Let $\overline{Q}$ be a question set and $n \in \mathbb{N}_{>0}$. We define the *goodness threshold* as

$$\omega_{\overline{Q}}^{\mathrm{good}}(n) := \inf\left\{\epsilon : \overline{Q} \text{ is } (n,\epsilon)\text{-good}\right\} .$$

We say that $\overline{Q}$ is *good* if $\lim_{n\to\infty} \omega_{\overline{Q}}^{\mathrm{good}}(n) = 0$. ◊

The value of the goodness threshold $\omega_{\overline{Q}}^{\mathrm{good}}(n)$ is an upper bound on the parallel repetition threshold $\omega_{\overline{Q}}(n)$:

Lemma 3.25. *Let $\overline{Q}$ be a question set. Then,*

$$\omega_{\overline{Q}}(n) \leq \omega_{\overline{Q}}^{\mathrm{good}}(n) .$$

Proof. Assume otherwise, i.e., that there exists a game $\mathcal{G}$ with question set $\overline{Q}$ and $\mathrm{val}(\mathcal{G}) < 1$ such that $\mathrm{val}(\mathcal{G}^n) > \omega_{\overline{Q}}^{\mathrm{good}}(n)$. We construct a perfect strategy for $\mathcal{G}$, which is a contradiction.

Fix an optimal strategy for $\mathcal{G}^n$ and let $S \subseteq \underline{\overline{Q}}$ with $\mu(S) > \omega_{\overline{Q}}^{\mathrm{good}}(n)$ be the set of question vectors in the repeated game for which the players win.

Let $\underline{f}$ be a vector of homomorphisms of $\overline{Q}$ that is good for S and let i be a coordinate where f_i is identity.

A strategy for the game $\mathcal{G}$ for the j-th prover is as follows: Given $q^{(j)} \in Q^{(j)}$, obtain $\underline{f}^{(j)}(q^{(j)}) = (f_1^{(j)}(q^{(j)}), \ldots, f_n^{(j)}(q^{(j)}))$. Then, consider the answer of the j-th prover on $\underline{f}^{(j)}(q^{(j)})$ in the strategy for $\mathcal{G}^n$. Finally, output the i-th coordinate of that answer.

Since for every $\overline{q} = (q^{(1)}, \ldots, q^{(r)}) \in \overline{Q}$ we have that $\underline{f}(\overline{q}) = (\underline{f}^{(1)}(q^{(1)}), \ldots, \underline{f}^{(r)}(q^{(r)})) \in S$, when applying the above strategy the provers are always winning on all coordinates of $\mathcal{G}^n$. Since $f_i(\overline{q}) = \overline{q}$, their answers on the i-th coordinate are winning for $\overline{q}$ in the game $\mathcal{G}$. Therefore, $\mathrm{val}(\mathcal{G}) = 1$, a contradiction. □

Remark 3.26. Lemma 3.25 is essentially what is called in the literature the forbidden subgraph method. However, our formulation with homomorphisms is different than the usual one, e.g., in [FV02].

Verbitsky [Ver95, FV02] showed that the forbidden subgraph method is universal, i.e., for a connected question set $\overline{Q}$ there is $\omega_{\overline{Q}}(n) = \omega_{\overline{Q}}^{\mathrm{good}}(n)$. ◊

3.2.4 Proving that $\overline{Q}$ is good with probabilistic method

Lemma 3.27. *Let $\overline{Q}$ be a question set and let $\mathcal{H}$ be a distribution over $\mathrm{Hom}(\overline{Q}, \overline{Q})$ such that:*

1. *If $\underline{f} = (f_1, \ldots, f_n)$ is sampled such that f_i is i.i.d. in $\mathcal{H}$, then:*

$$\forall S \subseteq \underline{\overline{Q}} : \Pr\left[\forall \overline{q} \in \overline{Q} : \underline{f}(\overline{q}) \in S\right] \geq c(\mu(S)) \,,$$

   where $c(\mu) > 0$ if $\mu > 0$.

2. *$\mathcal{H}(\mathrm{Id}) > 0$.*

Then, $\overline{Q}$ is good. Furthermore, if $\mathcal{H}(\mathrm{Id}) \geq \epsilon > 0$ and $c(\mu) \geq \mu^C / C$ for some $C \geq 1$, then $\omega_{\overline{Q}}^{\mathrm{good}}(n) \leq 3\exp(-\epsilon n / C)$.

Proof. Let $\epsilon := \mathcal{H}(\mathrm{Id})$ and $\mu \in (0, 1]$. For $S \subseteq \underline{\overline{Q}}$ with $\mu(S) = \mu$, define the event:

$$\mathcal{E} :\equiv \forall \overline{q} \in \overline{Q} : \underline{f}(\overline{q}) \in S \wedge \exists i \in [n] : f_i = \mathrm{Id} \,.$$

Since $\Pr[\forall \overline{q} \in \overline{Q} : \underline{f}(\overline{q}) \in S] \geq c(\mu)$ and $\Pr[\exists i : f_i = \mathrm{Id}] = 1 - (1-\epsilon)^n$, by union bound, if:

$$\Pr[\forall i : f_i \neq \mathrm{Id}] = (1-\epsilon)^n \leq c(\mu)/2 \,, \tag{58}$$

then $\Pr[\mathcal{E}] > 0$. Therefore, if we choose n such that (58) holds, then $\overline{Q}$ is (n,μ)-good. Since for arbitrary $\mu \in (0,1]$ we found that $\overline{Q}$ is (n,μ)-good for n big enough, $\overline{Q}$ must be good.

Furthermore, if $c(\mu) \geq \mu^C/C$, setting:

$$\mu := (2C(1-\epsilon)^n)^{1/C} \leq (2C)^{1/C} \cdot \exp(-\epsilon n/C) \leq 3\exp(-\epsilon n/C) \,,$$

we see that:

$$(1-\epsilon)^n = \mu^C/2C \leq c(\mu)/2 \,,$$

and therefore $\omega_{\overline{Q}}^{\mathrm{good}}(n) \leq 3\exp(-\epsilon n/C)$. □

3.2.5 Same-set hitting homomorphism spaces

Lemma 3.28. *Let $\overline{P}$ be an r-partite hypergraph constructible using k doublings (and an arbitrary number of collapses) and let $\overline{Q}$ be another r-partite hypergraph.*

Then, there exists a distribution $\mathcal{H}$ over $\mathrm{Hom}(\overline{P},\overline{Q})$ such that:

1. *If $\underline{f} = (f_1,\ldots,f_n)$ is sampled such that f_i is i.i.d. in $\mathcal{H}$, then:*

$$\forall S \subseteq \overline{Q} : \Pr\left[\forall \overline{p} \in \overline{P} : \underline{f}(\overline{p}) \in S\right] \geq \mu(S)^C \,,$$

   where $C = 2^k$.

2. $\min_{f \in \mathrm{Hom}(\overline{P},\overline{Q})} \mathcal{H}(f) \geq 1/M^C$, *where $C = 2^k$ and $M = \left|\overline{Q}\right|$.*

This lemma is connected to the same-set hitting from Chapter 2 in the following way: Let f be a random homomorphism sampled according to $\mathcal{H}$ and let $\overline{P} = \{\overline{p}^{(1)},\ldots,\overline{p}^{(\ell)}\}$. We can think of $\mathcal{H}$ as a k-step probability distribution with the steps given by $f(\overline{p}^{(1)}),\ldots,f(\overline{p}^{(\ell)})$. Then, the first condition in Lemma 3.28 is equivalent to saying that $\mathcal{H}$ is polynomially same-set hitting.

Later we will apply Lemma 3.28 with $\overline{P} = \overline{Q}$.

Proof. The proof proceeds by induction on the structure of $\overline{P}$. To achieve the constant C as claimed, we need to show the base case with $C = 1$ and then argue that a collapse preserves C and that a doubling increases C at most twice.

1. If $\overline{P}$ is a single hyperedge, then $\mathrm{Hom}(\overline{P}, \overline{Q})$ is isomorphic to $\overline{Q}$. Setting $\mathcal{H}(f_{\overline{q}}) := 1/M$ for $\overline{q} \in \overline{Q}$ one can easily see that both 1 and 2 are satisfied with $C = 1$.

2. Assume that $\overline{P}$ was constructed by doubling a hypergraph $\overline{P}_0$. Let A be the set of fixed vertices, B the old vertices and B' the new vertices (regardless of the player they belong to). Therefore the vertex set of $\overline{P}_0$ is $A \cup B$ and the vertex set of $\overline{P}$ is $A \cup B \cup B'$.

   We are going to write homomorphisms $f \in \mathrm{Hom}(\overline{P}_0, \overline{Q})$ as $f = (f_A, f_B)$ and $f \in \mathrm{Hom}(\overline{P}, \overline{Q})$ as $f = (f_A, f_B, f_{B'})$.

   Observe that

$$\begin{aligned}\mathrm{Hom}(\overline{P}, \overline{Q}) = \big\{ (f_A, f_B, f_{B'}) : (f_A, f_B) \in \mathrm{Hom}(\overline{P}_0, \overline{Q}) \\ \wedge\, (f_A, f_{B'}) \in \mathrm{Hom}(\overline{P}_0, \overline{Q}) \big\}\,, \end{aligned} \tag{59}$$

   where we abused the notation in the expression $(f_A, f_{B'})$: this is justified from the definition of the doubling operation.

   By induction, there exists a distribution $\mathcal{H}_0$ on $\mathrm{Hom}(\overline{P}_0, \overline{Q})$ satisfying 1 and 2 for some $C_0 > 0$. Let $H = (H_A, H_B)$ be a random variable distributed according to $\mathcal{H}_0$. Define:

$$\begin{aligned}\mathcal{H}(f_A, f_B, f_{B'}) := \Pr[H_A = f_A] \cdot \Pr[H_B = f_B \mid H_A = f_A] \\ \cdot \Pr[H_B = f_{B'} \mid H_A = f_A]\,. \end{aligned} \tag{60}$$

   By (59), (60) defines a probability distribution. Furthermore:

$$\mathcal{H}(f_A, f_B, f_{B'}) \geq \mathcal{H}_0(f_A, f_B) \cdot \mathcal{H}_0(f_A, f_{B'}) \geq \epsilon^{2C_0}\,.$$

   As for condition 1, let $\overline{E}_A$ be the fixed hyperedges of $\overline{P}_0$ (i.e., those that have all their vertices in A) and $\overline{E}_B$ and $\overline{E}_{B'}$ be the hyperedges of $\overline{P}$ that have vertices incident to B and B', respectively. Note that $\overline{E}_A$ and $\overline{E}_B$ form a partition of $\overline{P}_0$ and $\overline{E}_A$, $\overline{E}_B$ and $\overline{E}_{B'}$ form a partition of $\overline{P}$.

   Recall that $\underline{f} = (f_1, \ldots, f_n)$ is a random vector with coordinates sampled i.i.d. from $\mathcal{H}$. We are going to decompose $\underline{f} = (\underline{f}_A, \underline{f}_B, \underline{f}_{B'})$ in the natural way. Fix $S \subseteq \overline{Q}^n$ and define the event $\mathcal{E} :\equiv \forall \overline{p} \in \overline{E}_A : \underline{f}(\overline{p}) \in S$.

We estimate, using Jensen's inequality in (61):

$$\begin{aligned}
&\Pr\left[\forall \overline{p} \in \overline{P} : \underline{f}(\overline{p}) \in S\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\mathbb{1}_{\mathcal{E}} \cdot \Pr\left[\forall \overline{p} \in \overline{E}_B \cup \overline{E}_{B'} : \underline{f}(\overline{p}) \in S \mid \underline{f}_A\right] \mid \underline{f}_A\right]\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\mathbb{1}_{\mathcal{E}} \cdot \Pr\left[\forall \overline{p} \in \overline{E}_B : \underline{f}(\overline{p}) \in S \mid \underline{f}_A\right]^2 \mid \underline{f}_A\right]\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\mathbb{1}_{\mathcal{E}} \cdot \Pr\left[\forall \overline{p} \in \overline{E}_B : \underline{f}(\overline{p}) \in S \mid \underline{f}_A\right]\right)^2 \mid \underline{f}_A\right]\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\mathbb{1}_{\mathcal{E}} \cdot \Pr\left[\forall \overline{p} \in \overline{E}_B : \underline{f}(\overline{p}) \in S \mid \underline{f}_A\right] \mid \underline{f}_A\right]^2\right] \\
&\geq \mathrm{E}\left[\mathrm{E}\left[\mathbb{1}_{\mathcal{E}} \cdot \Pr\left[\forall \overline{p} \in \overline{E}_B : \underline{f}(\overline{p}) \in S \mid \underline{f}_A\right] \mid \underline{f}_A\right]\right]^2 \qquad (61) \\
&= \Pr\left[\forall p \in P_0 : \underline{f}(p) \in S\right]^2 \geq \mu^{2C_0} \ .
\end{aligned}$$

3. The last case considers $\overline{P}$ constructed by collapsing some $\overline{P}_0$. Let A be the vertex set of $\overline{P}$ and $A \dot\cup B$ the vertex set of $\overline{P}_0$. Let $h \in \mathrm{Hom}(\overline{P}_0, \overline{P})$ be a homomorphism that defines this collapse.

   By induction, there exists a distribution $\mathcal{H}_0$ on $\mathrm{Hom}(\overline{P}_0, \overline{Q})$ satisfying properties 1 and 2 for some C_0. For $f \in \mathrm{Hom}(\overline{P}, \overline{Q})$, define

$$\mathcal{H}(f) := \sum_{\substack{g \in \mathrm{Hom}(\overline{P}_0, \overline{Q}) \\ g_A = f}} \mathcal{H}_0(g) \ .$$

   Since a restriction of a homomorphism is a homomorphism, $\mathcal{H}$ indeed is a probability distribution. Furthermore, since $\mathcal{H}(f) \geq \mathcal{H}(h \circ f) \geq \epsilon^{C_0}$, condition 2 is satisfied.

   Finally, let $\underline{f}_0$ be a vector of question homomorphisms that are sampled i.i.d. from $\mathcal{H}_0$ and recall that the vector $\underline{f}$ is sampled i.i.d. from $\mathcal{H}$. To establish condition 1, we check that

$$\begin{aligned}
\Pr[\forall \overline{p} \in \overline{P} : \underline{f}(\overline{p}) \in S] &= \Pr[\forall \overline{p} \in \overline{P} : \underline{f}_0(\overline{p}) \in S] \\
&\geq \Pr[\forall \overline{p} \in \overline{P}_0 : \underline{f}_0(\overline{p}) \in S] \geq \mu^{C_0} \ .
\end{aligned}$$

□

3.2.6 Putting things together

Proof of Theorem 3.21. Let $\overline{Q}$ be an r-prover question set constructible by conditioning using k doublings with $\left|\overline{Q}\right| = M$.

By Lemma 3.28 applied for $\overline{P} = \overline{Q}$, Lemma 3.27 and Lemma 3.25,

$$\omega_{\overline{Q}}(n) \leq \omega_{\overline{Q}}^{\text{good}}(n) \leq 3\exp\left(-n/2^k M^{2^k}\right) \leq 3\exp\left(-n/M^{2^{k+1}}\right) .$$

Since $3\exp(-\alpha n) \leq \exp(-\alpha n/2)$ for n big enough, this implies that $\overline{Q}$ admits exponential parallel repetition. □

As a corollary, we get exponential parallel repetition of free games:

Corollary 3.29. *Let $\mathcal{G}$ be an r-prover free game with 2^k questions available to each prover, so that the question set $\overline{Q}$ has size $M := 2^{kr}$. If $\mathrm{val}(\mathcal{G}) < 1$, then*

$$\mathrm{val}(\mathcal{G}^n) \leq 3\exp(-n/M^{2M}) .$$

Proof. By Remark 3.20, the question hypergraph of game $\mathcal{G}$ can be constructed using rk doublings. The bound then follows from Theorem 3.21:

$$\mathrm{val}(\mathcal{G}^n) \leq \omega_{\mathcal{Q}}(n) \leq 3\exp\left(-n/M^{2^{rk+1}}\right) = 3\exp\left(-n/M^{2M}\right) .$$

□

We note that quantitatively the bound for free games from Corollary 3.29 is much worse than the best known one by Feige, which is $\exp\left(-\Omega\left(n/M\log M\right)\right)$ [Fei91].

3.3 Constructing Graphs with Treewidth Two

We turn to presenting the power of our system for proving parallel repetition. In particular, we show that all two-prover graphs with treewidth at most two are constructible.

Since in this section we deal only with two provers, we use more standard notation where a bipartite graph is denoted as $G = (X, Y, E)$. We will sometimes refer to vertices from X as "on the left" and from Y as "on the right".

Our main result here is Theorem 3.36.

3.3.1 Warm-up: forests are constructible

We start with showing that all forests are constructible, recovering the parallel repetition result by Verbitsky [Ver95]. We will later use Lemma 3.31 in the construction of series-parallel graphs.

Firstly, we note that it is only interesting to consider constructability of connected graphs (note that to create a new connected component one can double all vertices of an existing connected component):

Claim 3.30. *A bipartite graph G is constructible by conditioning if and only if all its connected components are constructible.*

We can always add a "fresh" leaf to a constructible graph G:

Lemma 3.31. *If $G = (X \dot\cup \{u\}, Y, E)$ is constructible, then $G' = (X \dot\cup \{u\}, Y \dot\cup \{v\}, E \cup \{(u, v)\})$ is also constructible.*

Proof. Pick an arbitrary edge (u, w) originating from u. Fix u and double all the other vertices. Then collapse all new vertices on the left onto u and all new vertices on the right onto w' (i.e., the copy of w). □

From Claim 3.30, iterated application of Lemma 3.31 and Theorem 3.21 we have:

Theorem 3.32. *Let G be a tree. Then, G is constructible by conditioning.*

In particular, if G is interpreted as a two-prover question set, then it admits exponential parallel repetition.

3.3.2 Treewidth and series-parallel graphs

Definition 3.33 (Treewidth)**.** Let G be a simple graph. A *tree decomposition* of G is a tree T, where each node (also called a *bucket*) corresponds to a subset of the vertices of G, with the following properties:

- For each vertex v of G, the buckets in which v appears form a non-empty, connected subgraph of T.
- For each edge e of G, there exists a bucket that contains both endpoints of e.

The *width* of a tree decomposition of G is the size of the biggest bucket minus one. The *treewidth* of G denoted by $\mathrm{tw}(G)$ is the smallest possible width of a tree decomposition of G. ◊

We will not discuss treewidth here, referring the reader to any standard textbook on graph theory. We note that a connected graph has treewidth one if and only if it is a tree.

To characterise graphs with treewidth two, we need to introduce the notion of *generalized series-parallel graphs*.

Definition 3.34 (Series-parallel graphs)**.** Let $G = (X, Y, E)$ be a bipartite graph and $u, v \in X \cup Y$. We call a tuple (X, Y, E, u, v) an *oriented bipartite graph*. We call the vertex u the *top* and v the *bottom*.

We define the class of generalized bipartite series-parallel oriented (in short: series-parallel oriented) graphs recursively:

1. Let $G = (\{a\}, \{b\}, \{(a, b)\})$ be a single edge. Then, both (G, a, b) and (G, b, a) are series-parallel oriented graphs.

2. Let $G_1 = (X_1, Y_1, E_1, u, v)$ and $G_2 = (X_2, Y_2, E_2, v, w)$ be series-parallel oriented graphs such that $(X_1 \cup Y_1) \cap (X_2 \cup Y_2) = \{v\}$ and $v \in (X_1 \cap X_2) \cup (Y_1 \cap Y_2)$.

   Then, $G := (X_1 \cup X_2, Y_1 \cup Y_2, E_1 \cup E_2, u, w)$ is a series-parallel oriented graph.

   We say that G is a *series composition* of G_1 and G_2 with G_1 on top and G_2 at the bottom.

3. Let $G_1 = (X_1, Y_1, E_1, u, v)$ and $G_2 = (X_2, Y_2, E_2, v, w)$ be series-parallel graphs satisfying the same preconditions as for the series composition.

   Then, both $G := (X_1 \cup X_2, Y_1 \cup Y_2, E_1 \cup E_2, u, v)$ and $G' := (X_1 \cup X_2, Y_1 \cup Y_2, E_1 \cup E_2, v, w)$ are series-parallel graphs.

   We say that G and G' are a *generalized series composition* of G_1 and G_2. We say that G_1 is the *primary graph* of G and that G_2 is the primary graph of G'.

4. Let $G_1 = (X_1, Y_1, E_1, u, v)$ and $G_2 = (X_2, Y_2, E_2, u, v)$ be series-parallel oriented graphs such that $(X_1 \cup Y_1) \cap (X_2 \cup Y_2) = \{u, v\}$ and $\{u, v\} \subseteq (X_1 \cap X_2) \cup (Y_1 \cap Y_2)$.

   Then, $G := (X_1 \cup X_2, Y_1 \cup Y_2, E_1 \cup E_2, u, v)$ is also a series-parallel oriented graph.

   We call G a *parallel composition* of G_1 and G_2.

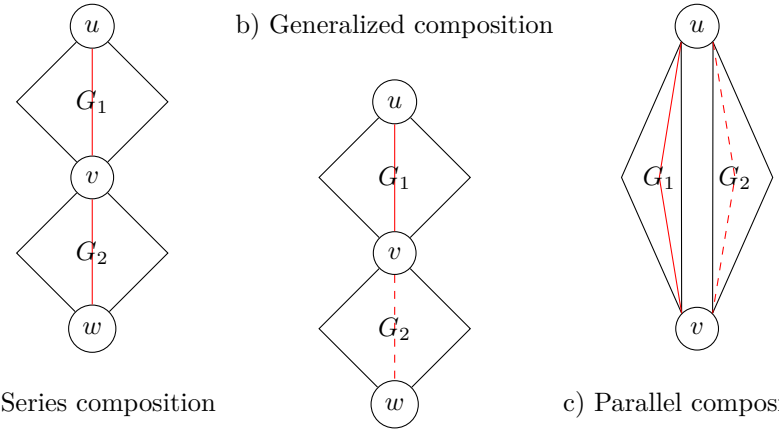
We say that a bipartite graph G is series-parallel if there exist vertices u, v such that (G, u, v) is an oriented series-parallel graph. ◊

We refer the reader to Figure 3.2 for intuitive understanding of the composition operations.

The requirement that the vertices by which the bipartite graphs are joined belong to the set $(X_1 \cap X_2) \cup (Y_1 \cap Y_2)$ ensures that they belong to the same side of the graph and therefore the bipartedness is preserved. On the other hand, observe that the top and the bottom can lie either on the same or the opposite sides of the bipartite graph.

In the literature the (not necessarily bipartite) graphs constructed with series and parallel composition are usually called series-parallel, and graphs that are constructed also with generalized composition are called generalized series-parallel. Incidentally, a connected graph is generalized series-parallel if and only if all its biconnected components are series-parallel (see, e.g., [Bod07]).

Figure 3.2: Illustration of the composition operations. The spines are drawn in continuous red. Note that $\mathcal{S}(G_2)$ is not part of $\mathcal{S}(G)$ in cases of generalized and parallel composition, hence the red dashed line.



From now on, by "series-parallel" we will always mean the generalized bipartite series-parallel graph from Definition 3.34.

We will use the following useful characterisation of graphs with treewidth at most two:

Theorem 3.35. *A connected bipartite graph G has treewidth at most two if and only if G is series-parallel.*

For a proof of Theorem 3.35 see [HHC99]. Their proof concerns the case of general (non-bipartite) graphs, but it is easy to see that a generalized series-parallel graph is bipartite if and only if it can be constructed with additional restrictions as in Definition 3.34.

3.3.3 Generalized series-parallel construction

The main theorem of this section is:

Theorem 3.36. *Every bipartite graph G with treewidth at most two is constructible by conditioning.*

In particular, if G is interpreted as a two-prover question set, then it admits exponential parallel repetition.

Due to Theorem 3.21, Claim 3.30 and Theorem 3.35, to establish Theorem 3.36 it is enough to show that series-parallel graphs are constructible. We spend the rest of this section to achieve that goal.

Definition 3.37. Let G be an oriented series-parallel graph. We define its (not oriented) subgraph $\mathcal{S}(G)$ and call it its *spine*. The definition follows the recursive pattern of Definition 3.34:

1. If G is a single edge, its spine is the whole of G.
2. If G is a series composition of G_1 and G_2, then $\mathcal{S}(G)$ consists of $\mathcal{S}(G_1)$ and $\mathcal{S}(G_2)$ taken together.
3. If G is a generalized composition of G_1 and G_2 with G_1 as the primary graph, then $\mathcal{S}(G)$ is equal to $\mathcal{S}(G_1)$.
4. If G is a parallel composition of G_1 and G_2 and $\mathcal{S}(G_1)$ has no more edges than $\mathcal{S}(G_2)$, then $\mathcal{S}(G)$ is equal to $\mathcal{S}(G_1)$. Otherwise, it is equal to $\mathcal{S}(G_2)$.

$\Diamond$

Observe that the spine is always an induced path between the top and the bottom of G. As a matter of fact, it is a shortest path from top to bottom in G. Furthermore, the length of the spine $L(G)$ is given as:

1. One, if G is a single edge.
2. $L(G_1) + L(G_2)$, if G is a series composition of G_1 and G_2.
3. $L(G_1)$, if G is a generalized composition of G_1 and G_2 with G_1 as the primary graph.
4. $\min(L(G_1), L(G_2))$, if G is a parallel composition of G_1 and G_2.

Finally, note that if G is a parallel composition of G_1 and G_2, then due to the bipartedness $L(G_1)$ and $L(G_2)$ must have the same parity.

Recall the graph construction operations from Definition 3.16. A series-parallel graph can always be collapsed onto its spine:

Lemma 3.38. *Let G be an oriented series-parallel graph. Then, G (treated as an unoriented graph) can be collapsed onto its spine.*

Proof. By induction on the series-parallel structure of G. If G is a single edge, it is clear. If G is a series composition of G_1 and G_2, then by induction G_1 and G_2 can be collapsed onto their respective spines.

If G is a generalized composition of G_1 and G_2, assume w.l.o.g. that G_1 is the primary graph and let v be the bottom vertex of G_1. Then, by induction, G_1 can be collapsed onto its spine $\mathcal{S}(G_1) = \mathcal{S}(G)$. On the other hand, all of G_2 can be collapsed onto the edge (v, w), where w is the neighbor of v in $\mathcal{S}(G_1)$.

In case (G, u, v) is a parallel composition of G_1 and G_2, assume w.l.o.g. that $\mathcal{S}(G_1)$ is not longer than $\mathcal{S}(G_2)$. Firstly, observe that the spine $\mathcal{S}(G_2)$ can be collapsed onto $\mathcal{S}(G_1) = \mathcal{S}(G)$: indeed, if we write the vertices of $\mathcal{S}(G_1)$ top-bottom as $(u_0 = u, u_1, \ldots, u_k = v)$ and analogously $\mathcal{S}(G_2)$ as $(v_0 = u, v_1, \ldots, v_{k+2\ell} = v)$, then the mapping:

$$f(u_i) := u_i$$
$$f(v_i) := \begin{cases} u_i & \text{if } i \leq k, \\ u_{k-(j \bmod 2)} & \text{if } i = k + j, \end{cases}$$

is a required homomorphism.

Finally, since by induction G_1 and G_2 can be collapsed onto their spines, and since the composition of homomorphisms is a homomorphism, G can be collapsed onto its spine. □

Recall that our objective is showing that every series-parallel graph is constructible.

Lemma 3.39. *Let (G, u, v) be an oriented series-parallel graph. Then, the spine $\mathcal{S}(G)$ can be extended to G using the doubling and collapsing operations. Furthermore, the construction preserves the following invariant:*

- *In every doubling step, the doubled vertices on the spine form its contiguous (possibly empty) subsegment.*

Since the spine of G can be constructed by repeated application of Lemma 3.31, Lemma 3.39 implies what we want. In the remainder we prove Lemma 3.39 after establishing a couple of technical preliminaries.

Remark 3.40. In the proof of Lemma 3.39 we will use the fact that whenever G is a composition of G_1 and G_2, the edges of G_1 and G_2 are disjoint.

This is not true if G is a parallel composition and there exists a direct edge from top to bottom in both G_1 and G_2, but any series-parallel G can be constructed without using this special case. ◊

Claim 3.41. *Let (G, u, v) be an oriented series-parallel graph which is a parallel composition. Then, there exists a series-parallel construction of (G, u, v) such that its final step is a parallel composition of G_1 and G_2 with the following properties:*

- $L(G_1) \leq L(G_2)$.
- *G_2 is a series composition.*

Proof. Firstly, note that whenever (G, u, v) is a generalized composition where the primary graph is a series or parallel composition, the order of those two compositions can be reversed without changing the final graph. Therefore, we can assume w.l.o.g. that whenever a graph is a generalized composition, its primary graph has spine of length one.

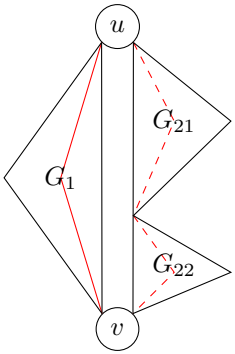
Let G be a parallel composition of H_1' and H_2'. If any of H_1' or H_2' is a parallel composition, recursively decompose them further until we are left with a collection of graphs $H_1, \ldots, H_k$ which are all series or generalized compositions or single edges.

Note that if we compose in parallel $H_1, \ldots, H_k$ in an arbitrary order, the end result will always be G.

Therefore, we can set G_2 as H_i with the longest spine and the parallel composition of the remaining H_i graphs as G_1. Due to Remark 3.40, the spine of G_2 must be longer than one, and therefore G_2 must be a series composition. □

Figure 3.3 illustrates the content of Claim 3.41.

Figure 3.3: The continuous red line is the spine of G. The dashed red line is the spine of G_2.



Proof of Lemma 3.39. Let (G, u, v) be an oriented series-parallel graph. We apply induction on the number of vertices of G and, secondarily, (in reverse) on the length of its spine.

1. If G is a single edge, there is nothing to prove (since $\mathcal{S}(G) = G$).

2. Assume that (G, u, v) is a series composition of (G_1, u, w) and (G_2, w, v). Recall that we need to extend $\mathcal{S}(G)$ to G. We do it in two stages, first extending $\mathcal{S}(G_1)$ to G_1 and then extending $\mathcal{S}(G_2)$ to G_2.

   By induction, we know how to extend $\mathcal{S}(G_1)$ to G_1. Now we will adapt this sequence of operations to the fact that also $\mathcal{S}(G_2)$ is present in the graph. We do it as follows:

   - Leave all the collapsing operations as they are (it is always possible to collapse onto a bigger graph).
   - For doubling operations that keep the vertex w fixed, keep all of $\mathcal{S}(G_2)$ fixed.
   - Finally, let us handle the doubling operations that double the vertex w. Let x be the neighbour of w on the spine $\mathcal{S}(G_1)$ and let y be x in case x is fixed and x' in case x is doubled. Note that the edge (y, w') is present in G_1 after doubling.

     To emulate this operation in G we double all of $\mathcal{S}(G_2)$ together with w and then collapse the new copy of $\mathcal{S}(G_2)$ onto the edge (y, w').

   Consult Figure 3.4 for the illustration of one of the cases.

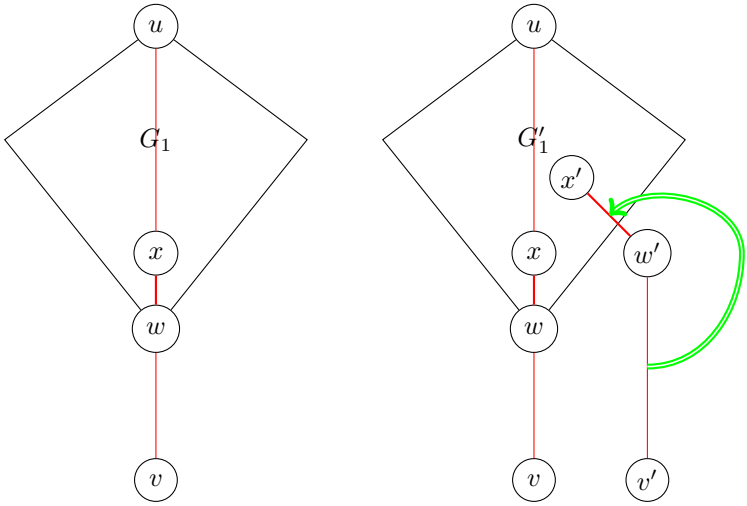
   It is easy to see that as a result of this emulation we extend $\mathcal{S}(G)$ to a series composition of G_1 and $\mathcal{S}(G_2)$.

   Now we proceed in the same way to extend $\mathcal{S}(G_2)$ to G_2. The only difference is that in case w is doubled we need to double and collapse all of G_1 instead of just $\mathcal{S}(G_1)$. This does not pose a problem though, since G_1 can be collapsed onto $\mathcal{S}(G_1)$ which then can be collapsed as previously.

   Finally, one easily checks that the "contiguous subsegment" invariant of Lemma 3.39 is preserved in this construction.

3. If (G, u, v) is a generalized composition, assume w.l.o.g. that it is a composition of the primary graph (G_1, u, v) and (G_2, v, w). Using Lemma 3.31 we can extend $\mathcal{S}(G_1)$ to $\mathcal{S}(G_1) \cup \mathcal{S}(G_2)$ and then proceed as in the series composition case.

Figure 3.4: Handling series decomposition in case w and x are doubled.



4. Assume that the graph (G, u, v) is a parallel composition of (G_1, u, v) and (G_2, u, v). By Claim 3.41, we can also assume that G_2 is a series composition of (G_3, u, w) and (G_4, w, v) and that $L(G_1) \leq L(G_2)$. In this point we address a subcase where additionally:

$$L(G_1) + L(G_3) < L(G_4) \,. \tag{62}$$

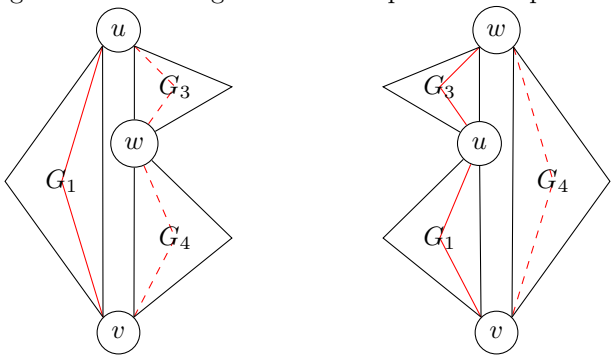
(G, u, v) is the parallel composition of (G_1, u, v) and the series composition of (G_3, u, w) and (G_4, w, v). Observe that we can also obtain (G, w, v) as the parallel composition of (G_4, w, v) and the series composition of (G_3, w, u) and (G_1, u, w). This is illustrated in Figure 3.5. Furthermore, due to (62) we have that $L(G, w, v) = L(G_3) + L(G_1) > L(G_1) = L(G, u, v)$.

To extend $\mathcal{S}(G, u, v) = \mathcal{S}(G_1)$ we proceed as follows: first, add $\mathcal{S}(G_3)$ on top of $\mathcal{S}(G_1)$ using Lemma 3.31. Then, extend $\mathcal{S}(G_3) \cup \mathcal{S}(G_1) = \mathcal{S}(G, w, v)$ to G using induction (which is applicable since the length of the spine increased).

Again, one easily checks that the contiguous subsegment invariant is preserved in this construction.

5. If G is a parallel composition and $L(G_1) + L(G_4) < L(G_3)$, we proceed symmetrically as in case 4.

Figure 3.5: Rotating G which is a parallel composition.



6. Finally, let (G, u, v) be a parallel composition and:

$$L(G_1) + L(G_3) \geq L(G_4)\ , \tag{63}$$
$$L(G_1) + L(G_4) \geq L(G_3)\ . \tag{64}$$

Again, we proceed in stages successively building (G_1, u, v), (G_3, u, w) and (G_4, w, v) using induction.

We start with $\mathcal{S}(G) = \mathcal{S}(G_1)$, which we need to extend to G. First, by induction we extend $\mathcal{S}(G_1)$ to G_1. Next, we add $\mathcal{S}(G_2) = \mathcal{S}(G_3) \cup \mathcal{S}(G_4)$ as follows: let $a := L(G_1)$ and $b := L(G_2)$. Recall that $b \geq a$ and that a and b have the same parity.

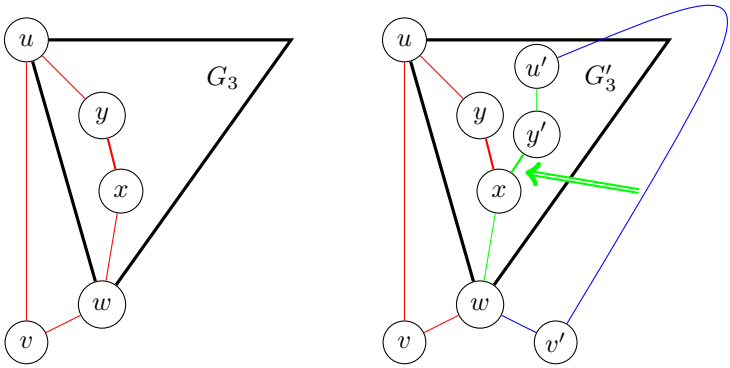
Using Lemma 3.31, add a path of length $(b-a)/2$ starting from u and let x be the endpoint of this path. Fix v and x and double all the other vertices. Finally, collapse the resulting copy of G_1 onto the path from v to u'.

In the next stage, we work with the sequence that extends $\mathcal{S}(G_3)$ to G_3. We need to adapt it to additional edges we have in the graph. This is done as follows:

- All collapsing operations stay the same.
- Doubling operations that keep both u and w fixed fix all the vertices of G_1 and $\mathcal{S}(G_4)$.
- In case at least one of u and w is doubled the arguments are very similar to each other. Therefore we present only the one where u is doubled and w is fixed. See Figure 3.6 for a graphical illustration.

  By inductive assumption, we know that a contiguous subpath of the spine $\mathcal{S}(G_3)$ is doubled. Assume that its doubled vertices go

Figure 3.6: Handling parallel decomposition when segment from u to y is doubled. For clarity, G_1 and G_4 are drawn as spines only. The blue path is collapsed onto the green path.



from u to y and the fixed ones from w to x (i.e., x and y are neighbours on the spine).

To emulate this case in G, double all vertices of G_1 and $\mathcal{S}(G_4)$ except of w. Next, collapse the new copy of G_1 onto its spine. Finally, collapse the resulting path $\mathcal{P}_1 := u' - v' - w$ onto the copy of $\mathcal{S}(G_3)$, i.e., $\mathcal{P}_2 := u' - y' - x - w$. This is possible due to (64): since the path $\mathcal{P}_1$ is at least as long as $\mathcal{P}_2$, $\mathcal{P}_1$ can be collapsed onto $\mathcal{P}_2$ as in the proof of Lemma 3.38.

Finally, we construct G_4 from $\mathcal{S}(G_4)$ in a very similar way. The only differences are that when emulating doubling we need to perform an additional collapse of G_3 onto $\mathcal{S}(G_3)$ and that we rely on the inequality (63) for the final collapse.

Again, one checks that the contiguous subsegment invariant is preserved throughout the whole process.

□

3.4 α-acyclic Hypergraphs Are Constructible

Recall that in Section 3.3.1 we showed that all forests are constructible by conditioning and therefore admit exponential parallel repetition. In this section we generalize that proof to hypergraphs, obtaining exponential parallel repetition for all *α-acyclic hypergraphs*:

Theorem 3.42. *Let $\overline{Q}$ be an r-uniform, r-partite α-acyclic hypergraph. Then, $\overline{Q}$ is constructible by conditioning.*

In particular, if $\overline{Q}$ is interpreted as an r-prover question set, then it admits exponential parallel repetition.

In the following we first define and characterize α-acyclicity, and then prove Theorem 3.42. The proof of constructibility uses a natural generalization of Lemma 3.31.

3.4.1 Hypergraphs and α-acyclicity

The following exposition of hypergraph acyclicity is mainly based on [Bra14].

Definition 3.43 (Basic hypergraph notions)**.** A (general) *hypergraph* H is a finite set of non-empty sets, which are called its *edges*. The set of vertices of a hypergraph $V(H)$ is the union of its edges.

For a set of vertices $S \subseteq V(H)$, the *induced hypergraph* $H(S)$ is defined as $H(S) := \{e \cap S \mid e \in H\} \setminus \{\emptyset\}$.

For a hypergraph H, we let its *minimization* $\mathcal{M}(H)$ to be $\mathcal{M}(H) := \{e \in H \mid \nexists f \in H : e \subsetneq f\}$, i.e., $\mathcal{M}(H)$ is the set of edges maximal for inclusion.

Let $m \geq 3$. We say that a tuple of pairwise distinct vertices $(t_1, \ldots, t_m)$ of a hypergraph H is a *cycle* if

$$\mathcal{M}\left(H\left(\{t_1, \ldots, t_m\}\right)\right) = \{\{t_1, t_2\}, \{t_2, t_3\}, \ldots, \{t_{m-1}, t_m\}, \{t_m, t_1\}\} \ .$$

We say that a hypergraph is *cycle-free* if it does not contain a cycle.

We say that two vertices $u, v \in V(H)$ are *neighbors* if there exists an edge $e \in H$ containing both u and v. We call a subset $S \subseteq V(H)$ a *clique* if every pair $u, v \in S$ are neighbors.

Finally, we say that a hypergraph H is *conformal* if each of its cliques is contained in an edge. ◊

Definition 3.44 (α-acyclicity)**.** A hypergraph H is *α-acyclic* if it is cycle-free and conformal. ◊

Example 3.45. The graph triangle $H_1 := \{\{u, v\}, \{v, w\}, \{w, u\}\}$ is neither cycle-free nor conformal. The graph square $H_2 := \{\{u, v\}, \{v, w\}, \{w, x\}, \{x, u\}\}$ is conformal, but not cycle-free.

The tetrahedron $H_3 := \{\{u, v, w\}, \{u, v, x\}, \{u, w, x\}, \{v, w, x\}\}$ is cycle-free, but not conformal. The beta triangle $H_4 := \{\{u, v\}, \{v, w\}, \{w, u\}, \{u, v, w\}\}$ is both cycle-free and conformal, that is α-acyclic.

The question graph $\overline{Q}_r$ from Definition 3.52 is not conformal, since the set of r vertices labeled with 0 is a clique that is not contained in an edge. On the other hand, adding the edge 0^r to $\overline{Q}_r$ makes it α-acyclic. ◊

Remark 3.46. In the setting introduced above, a simple graph is a hypergraph whose edges have all size two[1]. It is easy to see that a simple graph is a forest if and only if it is α-acyclic.

There exist several other notions of hypergraph acyclicity. For example, one might require the hypergraph to be only cycle-free, disregarding the conformality. On the other hand, there are multiple more restricted definitions, e.g., β-acyclicity, γ-acyclicity or Berge acyclicity (see [Bra14] for a survey).

It is somewhat interesting that we obtain exponential parallel repetition for a relatively unrestricted notion. ◊

For the proof we need a different characterization of α-acyclicity:

Definition 3.47 (GYO reduction, [Gra79, YÖ79])**.** We say that a hypergraph H' is obtained from H by an *included edge removal* if $H' = H \setminus \{e\}$ where $e \in H$ and $\exists f \in H : e \subseteq f \wedge e \neq f$.

We say that H' is obtained from H by a *singleton vertex removal* if $H' = H\left(V(H) \setminus \{u\}\right)$, where the degree of u is one, i.e., u is contained in a single edge of H.

We say that a hypergraph H is *GYO-reducible* if it can be reduced to the empty hypergraph by a sequence of included edge and singleton vertex removals. ◊

Theorem 3.48 (see Characterization 13 in [Bra14])**.** *A hypergraph is α-acyclic if and only if it is GYO-reducible.*

3.4.2 Constructability proof

To prove Theorem 3.42, we need to introduce one more elementary operation. Recall our definitions of graph constructability from Section 3.2.1.

Definition 3.49 (Simple constructability)**.** Let $(Q^{(1)}, \ldots, Q^{(r)}, \overline{Q})$ be an r-uniform, r-partite hypergraph and let $\overline{q} = (q^{(1)}, \ldots, q^{(r)}) \in \overline{Q}$ and $S \subseteq [r]$, $S \neq \emptyset$.

Let $\overline{q}'$ be a hyperedge where the positions not in S have the same vertices as in $\overline{q}$ and the positions in S have new vertices, as in the doubling operation.

We define the $(\overline{q}, S)$-extension of $\overline{Q}$ as $(P^{(1)}, \ldots, P^{(r)}, \overline{Q} \cup \{\overline{q}'\})$, where $P^{(j)} = Q^{(j)} \cup \{(q^{(j)})'\}$ or $P^{(j)} = Q^{(j)}$ depending on if $j \in S$.

We say that an r-uniform, r-partite hypergraph $\overline{Q}$ is *simply constructible* if it can be obtained from a single hyperedge by a sequence of $(\overline{q}, S)$-extensions. ◊

Lemma 3.50. *Every question set that is simply constructible is constructible by conditioning.*

[1] Neglecting the issue of isolated vertices.

Proof. We show that every $(\overline{q}, S)$-extension can be simulated using doubling and collapsing operations. It may be instructive to compare this with the proof of Lemma 3.31.

The simulation is achieved by one doubling and one collapse as follows: Let $\overline{q} = (q^{(1)}, \ldots, q^{(r)})$. First, fix all the vertices of $\overline{q}$ from the positions not in S and double all other vertices in the hypergraph $\overline{Q}$. If $j \in S$, then let $(q^{(j)})'$ be the copy of $q^{(j)}$.

Then, collapse all new vertices except for $(q^{(j)})'$: If $j \in S$, collapse each of them onto $(q^{(j)})'$, otherwise onto $q^{(j)}$. Check that this operation leaves all old edges intact and collapses all new edges onto $\overline{q}'$. Therefore, it is a valid collapse and its result is the $(\overline{q}, S)$-extension of $\overline{Q}$. □

Lemma 3.51. *Let $\overline{Q}$ be an r-uniform, r-partite hypergraph. $\overline{Q}$ is simply constructible if and only if it is GYO-reducible.*

Proof. If $\overline{Q}$ is simply constructible, it is easy to devise an appropriate sequence of removals: For every $(\overline{q}, S)$-extension (in reverse order) perform singleton vertex removals on new vertices of $\overline{q}'$ and then an included edge removal on the remaining part of $\overline{q}'$ (if any).

On the other hand, suppose that $\overline{Q}$ is GYO-reducible. As a preliminary point, let us assume that in case a vertex removal in the GYO reduction sequence causes two edges to collapse into one, we keep two identical copies of this edge instead (also counting them twice in the vertex degree calculation). Of course one of those two copies can be removed at any later time. It is easy to see that a hypergraph is GYO-reducible if and only if it is reducible with this modified procedure.

We proceed by induction. If $\overline{Q}$ is a collection of (pairwise disjoint) single hyperedges, then it is simply constructible and we are done. Otherwise, consider the first edge removal of some edge $\overline{q}$ in the reduction sequence of $\overline{Q}$.

Note that if we move all preceding vertex removals of the vertices contained in $\overline{q}$ to the front of the sequence and then move the removal of $\overline{q}$ directly thereafter, it is still a valid reduction sequence. But now after deleting $\overline{q}$ we obtain an r-uniform, r-partite hypergraph $\overline{P}$ that can be $(\overline{p}, S)$-extended to $\overline{Q}$, where $\overline{p}$ is one of the edges in which $\overline{q}$ was included and S corresponds to the previously deleted vertices of $\overline{q}$.

Since by induction $\overline{P}$ is simply constructible, $\overline{Q}$ is simply constructible as well. □

Proof of Theorem 3.42. Let $\overline{Q}$ be an r-uniform, r-partite α-acyclic hypergraph. By Theorem 3.48, it is GYO-reducible and by Lemma 3.51 it is simply constructible. Finally, by Lemma 3.50, it is constructible by conditioning. Moreover, by Theorem 3.21 the question set $\overline{Q}$ admits exponential parallel repetition. □

3.5 Lower Bounds on Multi-Prover Parallel Repetition

We turn to lower bounds on parallel repetition and on our methods. In this section we observe that for more than two provers the situation is dire: there exist question sets that do not admit exponential parallel repetition.

To this end, we need a result from [HHR16] establishing that parallel repetition of certain question sets implies the density Hales-Jewett theorem.

Definition 3.52. Let $r \geq 2$. We define an r-prover question set $\overline{Q}_r \subseteq \{0,1\}^r$ of size r, where the j-th question contains 1 in the j-th position and 0 in the remaining positions. In other words,

$$\overline{Q}_r := \{\overline{q} : w_1(\overline{q}) = 1\} \ .$$

$\Diamond$

Theorem 3.53 ([HHR16])**.** *Let $r \geq 3$, $n \geq 1$ and $S \subseteq [r]^n$ with $\mu(S) = |S|/r^n$ such that S does not contain a combinatorial line.*

There exists an r-prover game $\mathcal{G}_S$ with question set $\overline{Q}_r$ and answer alphabets $A^{(j)} = 2^{[n]} \times [n]$ such that:

- $\mathrm{val}(\mathcal{G}_S) \leq 1 - 1/r$.
- $\mathrm{val}(\mathcal{G}_S^n) \geq \mu(S)$.

Recall $\omega_r^{\mathrm{DHJ}}(n)$ from Definition 3.10. Theorem 3.53 immediately implies

Theorem 3.54 ([HHR16])**.** *Let $r \geq 3$. We have $\omega_r^{\mathrm{DHJ}}(n) \leq \omega_{\overline{Q}_r}(n)$.*

As another consequence, we have

Theorem 3.55. *Let $r \geq 3$. The question set $\overline{Q}_r$ does not admit exponential parallel repetition.*

Proof. Let n be divisible by r and let

$$S := \{\underline{x} \in [r]^n : w_1(\underline{x}) = \ldots = w_r(\underline{x}) = n/r\} \ .$$

It is clear that S does not contain a combinatorial line. At the same time, by Stirling's approximation, $\mu(S) \geq \Omega(1/n^{(r-1)/2})$ (where the constant in the $\Omega()$ notation depends on r) and therefore $\omega_r^{\mathrm{DHJ}}(n)$ cannot decrease exponentially.

By Theorem 3.54, $\omega_{\overline{Q}_r}(n)$ cannot decrease exponentially either. $\square$

Better lower bounds for $\omega_r^{\mathrm{DHJ}}(n)$ are known, with the best ones established by the Polymath project [Pol10].

Theorem 3.56 ([Pol10], Theorem 1.3)**.** *Let $\ell \geq 1$ and $r := 2^{\ell-1}+1$. There exists $C_\ell > 0$ such that for every $n \geq 2$ there exists a set $S \subseteq [r]^n$ with*

$$\mu(S) \geq \exp\left(-C_\ell \left(\log n\right)^{1/\ell}\right) ,$$

such that S does not contain a combinatorial line.

That is, for $r = 2^{\ell-1}+1$, we have

$$\omega_{\overline{Q}_r}(n) \geq \exp\left(-C_r \left(\log n\right)^{1/\lceil \log r \rceil}\right) , \tag{65}$$

where the $o(1)$ function depends only on r.

Inequality (65) is also interesting in the context of the two-prover parallel repetition lower bound by Feige and Verbitsky [FV02]. Recall that the upper bound of Raz (cf. (4)) exhibits a dependence on the answer set size. More specifically, it contains $\frac{1}{\log\left|\overline{A}\right|}$ term in the exponent. The example from [FV02] shows that if an exponential two-prover parallel repetition bound depends only on ϵ and $\left|\overline{A}\right|$, this term cannot be larger than $\frac{\log\log\left|\overline{A}\right|}{\log\left|\overline{A}\right|}$.

Our example implies that we can bring down this last term to $\frac{\left(\log\log\left|\overline{A}\right|\right)^{\epsilon}}{\log\left|\overline{A}\right|}$ for any $\epsilon > 0$, at the price of increasing the number of provers:

Theorem 3.57. *Let $\ell \geq 1$, $r := 2^{\ell-1}+1$. There exists a constant $C_\ell > 0$ such that for each $n \geq 2$ there exists an r-prover game $\mathcal{G}$ with question set $\overline{Q}_r$, $\mathrm{val}(\mathcal{G}) \leq 1 - 1/r$ and an answer set $\overline{A}$ with size $\left|\overline{A}\right| \in [2^{rn}, 2^{2rn}]$ such that*

$$\mathrm{val}(\mathcal{G}^n) \geq \exp\left(-C_\ell n \cdot \frac{\left(\log\log\left|\overline{A}\right|\right)^{1/\ell}}{\log\left|\overline{A}\right|}\right) . \tag{66}$$

Proof. Fix ℓ and n and take the r-prover game $\mathcal{G}_S$ from Theorem 3.53 for the set $S \subseteq [r]^n$ from Theorem 3.56. One verifies that $\mathcal{G}_S$ has question set $\overline{Q}_r$ and that the answer alphabet size is $\left|\overline{A}\right| = (2^n+n)^r \in [2^{rn}, 2^{2rn}]$.

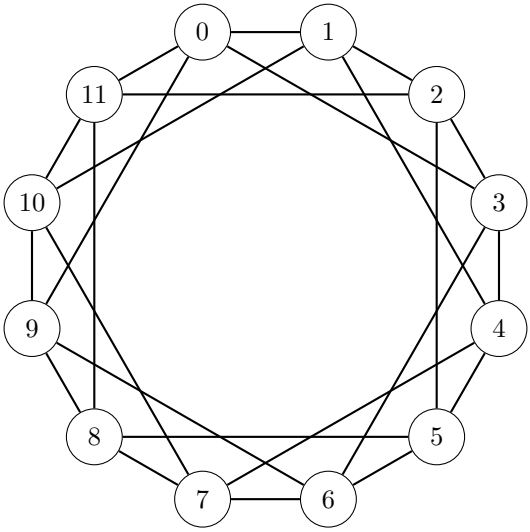
Since S has no combinatorial line, we have $\mathrm{val}(\mathcal{G}_S) \leq 1 - 1/r$ and $\mathrm{val}(\mathcal{G}_S^n) \geq \mu(S) \geq \exp\left(-C_\ell \left(\log n\right)^{1/\ell}\right)$.

Noting that $n \geq \log\left|\overline{A}\right|/2r$ and $\log n \leq \log\log\left|\overline{A}\right|$, we can establish (66):

$$\begin{aligned}\mathrm{val}(\mathcal{G}_S^n) &\geq \exp\left(-C_\ell \left(\log n\right)^{1/\ell}\right) = \exp\left(-C_\ell n \cdot \frac{\left(\log n\right)^{1/\ell}}{n}\right) \\ &\geq \exp\left(-C'_\ell n \cdot \frac{\left(\log\log\left|\overline{A}\right|\right)^{1/\ell}}{\log\left|\overline{A}\right|}\right) .\end{aligned}$$

□

Figure 3.7: A drawing of $\mathfrak{C}_{12}$.



As a final note, we reiterate that our lower bounds do *not* exclude the possibility of an information-theoretic (see (4)) parallel repetition bound. Furthermore, all results of this section concern games with at least three provers. We consider the two-prover case in the following sections.

3.6 Some Graphs Are Not Constructible

It is an open question if all two-prover distributions admit exponential parallel repetition. One way to prove that they do would be to show that all graphs are constructible by conditioning. However, in this section we show that that is not the case, hence another way must be found to resolve this open question:

Definition 3.58. Let $n \in \mathbb{N}$ be even and greater or equal to 8. We define the *cycle with shortcuts* $\mathfrak{C}_n$ as the following simple graph: $V(\mathfrak{C}_n) := \{0, \dots, n-1\}$ and $\{u, v\} \in E(\mathfrak{C}_n)$ if and only if $|u - v| \in \{1, 3, n-3, n-1\}$. $\Diamond$

See Figure 3.7 for a drawing of $\mathfrak{C}_{12}$. Observe that $\mathfrak{C}_n$ is bipartite. We show:

Theorem 3.59. *The cycle with shortcuts $\mathfrak{C}_{12}$ is not constructible by conditioning.*

Since any bipartite graph G joined with $\mathfrak{C}_{12}$ by a single vertex can be collapsed onto $\mathfrak{C}_{12}$, Theorem 3.59 implies the existence of an infinite family of graphs that are not constructible.

Our proof of Theorem 3.59 turns out to be quite involved and computer-assisted. Before we proceed with it, we explain why another natural proof idea fails.

3.6.1 Warm-up: constructing all induced subgraphs

A natural idea to prove Theorem 3.59 would be to show for a certain graph G that if it is not already present as an induced subgraph in another graph H, then no doubling of H can produce an induced instance of G. It turns out that this approach must fail, since for every bipartite graph G we can construct a graph H such that G is an induced subgraph of H.

Definition 3.60. Let $k \geq 1$. We define the *set graph* $\mathfrak{S}_k := (X, Y, E)$ as follows:

- $X := [k]$.
- $Y := \{S \subset [k] : S \neq \emptyset\}$.
- $E := \{(x, S) : x \in S\}$.

$\Diamond$

Theorem 3.61. *The set graph $\mathfrak{S}_k$ is constructible by conditioning with $2(k-1)$ doublings.*

Proof. The proof is by induction on k. The graph $\mathfrak{S}_1$ is just a single edge. To construct $\mathfrak{S}_{k+1}$, start with constructing $\mathfrak{S}_k$ with $2(k-1)$ doublings.

We make a preliminary point to avoid confusion. Note that the right hand-side vertices of $\mathfrak{S}_k$ are labeled with subsets of $[k]$ such that for a vertex labeled with S we have that its neighborhood is equal to its label: $N(S) = S$. We will now perform some doublings and label the new vertices with subsets that contain $k+1$. However, for a new vertex with a label S it is not evident that $N(S) = S$: this is what we have to prove.

After constructing $\mathfrak{S}_k$, perform a doubling as follows: double all vertices labeled with S such that $k \in S$ and label each new vertex as $S \cup \{k+1\}$.

Then, perform a second doubling: double k and, again, all vertices labeled with S such that $k \in S$ and $k+1 \notin S$. This time label the copy of k as $k+1$ and a copy of S as $S \setminus \{k\} \cup \{k+1\}$.

Note that after the doublings $Y = \{S \subseteq [k+1] : S \neq \emptyset\}$. For $S \in Y$ let $N(S) := \{x \in X : (x, S) \in E\}$ be the neighborhood of S. We need to check that $N(S) = S$ for every label S. This holds by the following case analysis:

- Each vertex labeled with S such that $k+1 \notin S$ existed before the first doubling and its neighborhood did not change (since it was doubled in the second doubling in case $k \in S$).
- Each vertex labeled with S such that $\{k, k+1\} \subseteq S$ was created in the first doubling, at which point we had $N(S) = S \setminus \{k+1\}$. Then, it was fixed in the second doubling and $k+1$ was added to its neighborhood.
- Each vertex labeled with S such that $k \notin S$ and $k+1 \in S$ was created in the second doubling with $N(S) = S$.

Therefore, we can construct $\mathfrak{S}_{k+1}$ from $\mathfrak{S}_k$ in 2 doublings and $\mathfrak{S}_{k+1}$ from $\mathfrak{S}_1$ in $2k$ doublings. □

Remark 3.62. A modification of this construction can be used to construct $\mathfrak{S}_{k,r}$ with $X := [k]$, $Y := \{S \subseteq [k] : |S| = r\}$ and $E := \{(x, S) : x \in S\}$. ◊

Now we turn to the proof of Theorem 3.59.

3.6.2 Decomposing last two steps

Definition 3.63. Let u, v be two vertices arising during a construction of a bipartite graph G. We write $u \sim v$ if u and v are adjacent. For two sets of vertices A, B, we write $E(A, B)$ for the set of edges between A and B. We also write $G(A)$ for the graph induced by vertices in A. ◊

Note that the operators $\sim$, $E(\cdot, \cdot)$ and $G(\cdot)$ do not depend on the stage of the construction: doubling and collapsing only add and remove vertices, without changing existing adjacencies.

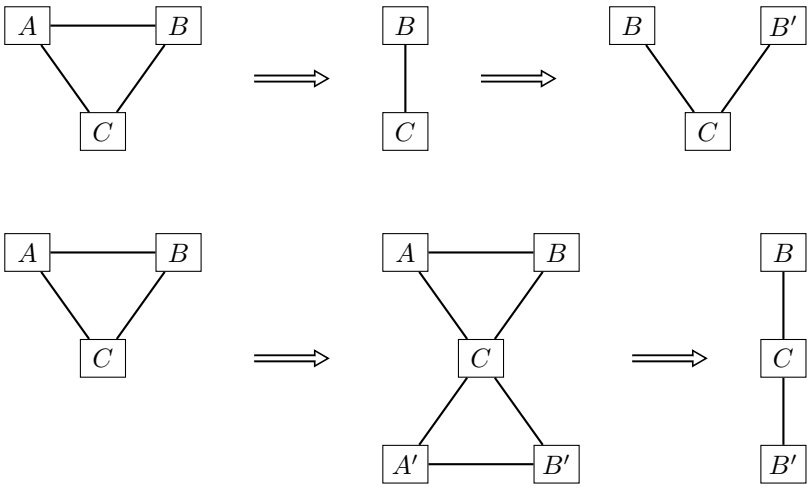
Lemma 3.64. *Let G be bipartite graph. If G is constructible, then it is constructible such that all the operations except for the last one are doublings.*

Proof. First, assume that in a construction of G there is a collapse operation immediately followed by a doubling operation. Assume that A is the set of the vertices collapsed in the first operation, B is the set of vertices that are fixed in the first operation and doubled onto B' in the second operation and C the set of vertices that are fixed throughout both operations (see Figure 3.8).

Then, those two operations can be exchanged as follows. First, double A onto A' and B onto B'. Then, collapse A onto $B \cup C$ and A' onto $B' \cup C$ (again see Figure 3.8). In both cases we end up with the same graph on vertices $B \cup B' \cup C$.

Finally, note that once all collapses are at the end of the sequence of the operations, they can be merged into a single collapse. □

Figure 3.8: Transposing a collapse and a doubling.



Definition 3.65. We say that a graph G is *collapsible* onto a graph H, if H can be constructed from G by a single collapse operation. ◊

Lemma 3.66. *Let H be a constructible graph with at least two edges. There exists a construction of H such that:*

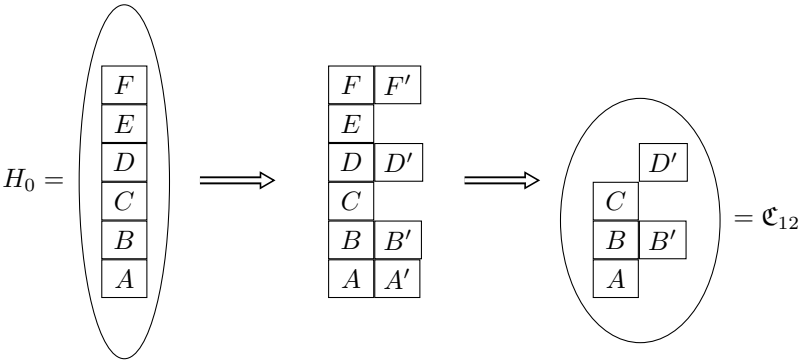
1. *The last operation is a collapse.*
2. *All other operations are doublings.*
3. *Leting H_0 be the graph before the last doubling, H_0 is not collapsible onto H.*

Proof. By Lemma 3.64, there exists a construction of H satisfying the first two conditions. Let us take such a construction with the smallest possible number of doublings. Since H is not a single edge, the number of doublings must be at least one.

If in this construction H_0 is collapsible onto H, the last doubling and the collapse can be replaced with a single collapse, which is a contradiction. □

Due to Lemma 3.64, we can assume that if the graph $\mathfrak{C}_{12}$ is constructible, the last two steps of its construction are, respectively, doubling and collapsing. Let us now divide the vertices of the construction depending on what happens to them in those last two steps (see Figure 3.9).

Figure 3.9: The last two steps in a construction of $\mathfrak{C}_{12}$.



The division is as follows: A are vertices that are doubled onto A' in the first step, with A fixed and A' collapsed in the second step. B are vertices doubled onto B' in the first step with both B and B' fixed in the second step. C are vertices that are fixed throughout both steps. D are vertices doubled onto D' in the first step with D collapsed and D' fixed in the second step. E are vertices fixed in the first step and collapsed in the second step. Finally, F are vertices that are doubled onto F' in the first step with both F and F' collapsed in the second step.

One checks that this division covers all possible events in the last two steps. The final graph $\mathfrak{C}_{12}$ consists of vertices $A \cup B \cup B' \cup C \cup D'$.

Our proof of Theorem 3.59 goes as follows: First, we show that if the last two steps of a construction of $\mathfrak{C}_{12}$ are as above, it must be $B = \emptyset$ and $E(A, D) = \emptyset$. Then, we prove that if $B = \emptyset$ and $E(A, D) = \emptyset$, then the initial graph H_0 must have been collapsible onto $\mathfrak{C}_{12}$ in the first place. Parts of the proof are computer-assisted, with the codes of C++ programs provided in Appendix B.

3.6.3 Non-collapsible graphs never produce $\mathfrak{C}_{12}$

Lemma 3.67. *Let $\mathfrak{C}_{12}$ be constructed in two steps from some bipartite H_0, as above. It cannot be that $E = F = \emptyset$, $E(A, D) = \emptyset$ and $B \neq \emptyset$.*

Proof. Computer-assisted (enumerate all partitions of $\mathfrak{C}_{12}$ into $A \cup B \cup B' \cup C \cup D'$ together with a bijection between B and B', since $E(A, D) = \emptyset$ such a partition implies a unique $H_0 = G(A \cup B \cup C \cup D)$), see the program `non_empty_b.cpp` in Listing 2. □

Lemma 3.68. *Let $\mathfrak{C}_{12}$ be constructed in two steps from some bipartite H_0, as above. It cannot be that $B = E = F = \emptyset$ and $|E(A, D)| = 1$.*

Proof. Computer-assisted (enumerate all partitions of $\mathfrak{C}_{12}$ into $A \cup C \cup D'$ and all edges between A and D, again this implies a unique $H_0 = G(A \cup C \cup D)$), see the program `non_empty_ad.cpp` in Listing 3. □

Lemma 3.69. *Let $\mathfrak{C}_{12}$ be constructed in two steps from some bipartite H_0, as above. Then, it must be that $B = \emptyset$ and $E(A, D) = \emptyset$.*

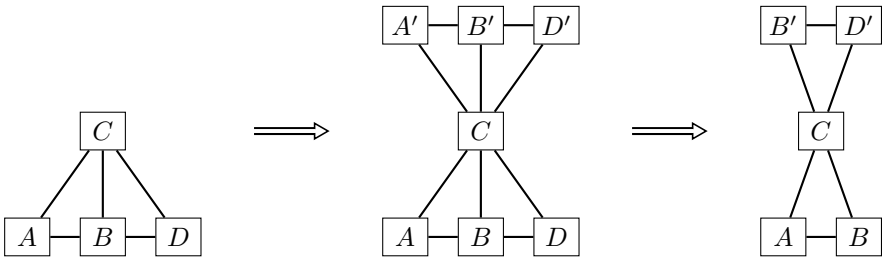
Proof. Assume by contradiction that there exists a construction of $\mathfrak{C}_{12}$ with $B \neq \emptyset$ or $E(A, D) \neq \emptyset$.

Firstly, note that the same construction but with the vertices from $E \cup F$ deleted from the initial graph H_0 is valid and also results in $\mathfrak{C}_{12}$. Therefore, we can assume w.l.o.g. that $E = F = \emptyset$.

We now proceed in two cases. If $B \neq \emptyset$, we can additionally assume that $E(A, D) = \emptyset$. This is again due to the fact that if we deleted $E(A, D)$ edges from H_0, we would still obtain a valid construction that results in $\mathfrak{C}_{12}$ (cf. Figure 3.10). But $B \neq \emptyset$ and $E(A, D) = \emptyset$ is impossible due to Lemma 3.67.

On the other hand, assume that $B = \emptyset$ and $E(A, D) \neq \emptyset$. Then, by the same argument as before, we can also assume that the size of $E(A, D)$ is as small as possible, namely $|E(A, D)| = 1$ (cf. Figure 3.11). But this also yields a contradiction by Lemma 3.68. □

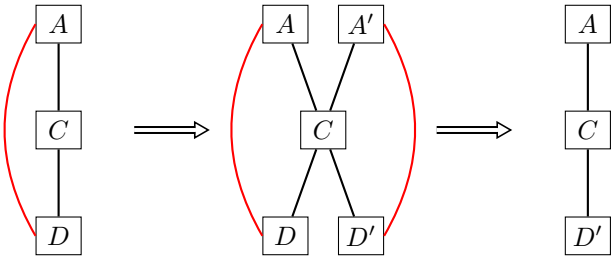
Figure 3.10: An illustration of Lemma 3.69, case $B \neq \emptyset$.



We need some additional concepts to deal with the remaining case $B = \emptyset$, $E(A, D) = \emptyset$.

Definition 3.70. Let A and B be disjoint sets of vertices that exist at some point during a construction of a bipartite graph H. Assume that a doubling

Figure 3.11: An illustration of Lemma 3.69, case $B = \emptyset$, $|E(A, D)| = 1$. The edge between A and D is marked red.



operation is performed and that all vertices from B (possibly together with some vertices from A and outside of $A \cup B$) are doubled.

Let B' be the set of copies of vertices from B. There is an obvious bijection between B and B' which we call the *natural bijection*. Similarly, we say that there is natural bijection between $A \cup B$ and $A \cup B'$. If this bijection is also an isomorphism between $G(A \cup B)$ and $G(A \cup B')$, we say that $G(A \cup B)$ and $G(A \cup B')$ are *naturally isomorphic*. ◊

Definition 3.71. Let $\mathfrak{C}_{12}$ be constructed from some H_0 in two steps, as above. We say that A' was *naturally collapsed* onto A, if in the collapse step each vertex of A' was collapsed onto A via the natural bijection. Analogously, we say that D was naturally collapsed onto D'. ◊

Lemma 3.72. *Let $\mathfrak{C}_{12}$ be constructed from some H_0 with one doubling and one collapse, as above. If $B = E = F = \emptyset$ and $E(A, D) = \emptyset$, then in the subsequent collapse either A' is naturally collapsed onto A or D is naturally collapsed onto D'.*

Proof. Computer-assisted (enumerate all partitions of $\mathfrak{C}_{12}$ into $A \cup C \cup D'$ and all possible collapses), see the program `natural_collapse.cpp` in Listing 4. □

Lemma 3.73. *Let $\mathfrak{C}_{12}$ be constructed from some H_0 with one doubling and one collapse, as above. If $B = \emptyset$ and $E(A, D) = \emptyset$, then in the subsequent collapse either A' is naturally collapsed onto A or D is naturally collapsed onto D'.*

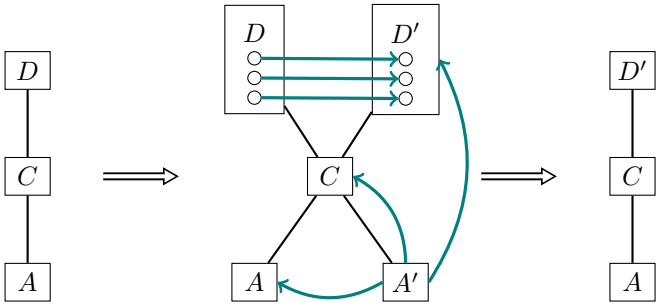
Proof. Assume there exists a construction of $\mathfrak{C}_{12}$ from some $H = G(A \cup C \cup D \cup E \cup F)$ such that:

1. $B = \emptyset$ and $E(A, D) = \emptyset$.

2. A' does not naturally collapse onto A and D does not naturally collapse onto D'.

Then, the same construction with vertices $E \cup F$ omitted from H_0 is also valid and satisfies both conditions. Therefore, we can assume w.l.o.g. that $E = F = \emptyset$. Then, the result follows from Lemma 3.72. □

Figure 3.12: An illustration of Lemma 3.73, case when $E = F = \emptyset$ and D collapses naturally onto D'. Blue arrows denote a collapse.



Lemma 3.74. *Let $\mathfrak{C}_{12}$ be constructed from some bipartite H_0 by one doubling and one collapse, as above. If $B = \emptyset$ and $E(A, D) = \emptyset$, then H_0 is collapsible onto $\mathfrak{C}_{12}$.*

Proof. Let the two steps in a construction of $\mathfrak{C}_{12}$ be such as in the statement. Recall that $G(S)$ denotes the induced graph on a vertex set S. Note that $H_0 = G(A \cup C \cup D \cup E \cup F)$ and that $\mathfrak{C}_{12} = G(A \cup C \cup D')$. For the following discussion cf. Figures 3.9 and 3.12.

Since $E(A, D) = \emptyset$, the graphs $G(A \cup C \cup D')$ and $G(A \cup C \cup D)$ are naturally isomorphic. Therefore, it is enough to show that it is possible to collapse $E \cup F$ onto $A \cup C \cup D$.

Let us write the collapse that produces $\mathfrak{C}_{12}$ as a homomorphism $f' : A' \cup D \cup E \cup F \cup F' \to A \cup C \cup D'$. By Lemma 3.73, either A' collapses naturally onto A or D collapses naturally onto D'.

Consider first that A' collapses naturally. We create a collapsing homomorphism $f : E \cup F \to A \cup C \cup D$ as follows:

- If $u \in E$ and $f'(u) \in A \cup C$, then $f(u) := f'(u)$. If $f'(u) = w' \in D'$, then $f(u) := w \in D$.
- For $u \in F$ with $u' \in F'$, if $f'(u') \in A \cup C$, then $f(u) := f'(u')$. If $f'(u') = w' \in D'$, then $f(u) := w \in D$.

We need to see that f is indeed a homomorphism, i.e., that all edges that touch $E \cup F$ are mapped onto edges of $G(A \cup C \cup D)$. To this end we make a case analysis:

- Since $G(E \cup F)$ is naturally isomorphic to $G(E \cup F')$ and $G(A \cup C \cup D)$ is naturally isomorphic to $G(A \cup C \cup D')$, the edges from $E(E \cup F, E \cup F)$ are preserved by f.
- Since $G(A \cup C \cup D \cup E)$ is naturally isomorphic to $G(A \cup C \cup D' \cup E)$, the edges from $E(E, A \cup C \cup D)$ are also preserved by f.
- Let $u \in F$, $v \in A$, $u \sim v$. Then $u' \sim v' \implies f'(u') \sim v \implies f(u) \sim v$, where we used that A' collapses naturally.
- Let $u \in F$, $v \in C$, $u \sim v$. Then $u' \sim v \implies f'(u') \sim v \implies f(u) \sim v$.
- Finally, let $u \in F$, $v \in D$, $u \sim v$. Then $u' \sim v' \implies f'(u') \sim v' \implies f(u) \sim v$.

Second, assume that D collapses naturally onto D'. In that case we give a collapsing homomorphism $f : \mathrm{E} \cup F \to A \cup C \cup D'$ as follows: if $f'(u) \in A \cup C$, then $f(u) := f'(u)$. If $f'(u) = w' \in D'$, then $f(u) := w \in D$. To see that f is a collapsing homomorphism, consider:

- Since $G(A \cup C \cup D)$ is naturally isomorphic to $G(A \cup C \cup D')$, f preserves the edges from $E(E \cup F, A \cup C \cup E \cup F)$.
- If $u \in E \cup F$, $v \in D$, $u \sim v$ consider the subcases (in all of them we use that D collapses naturally):
  - If $f'(u) \in A$, then $A \ni f'(u) \sim f'(v) = v' \in D'$, implying $E(A, D') \neq \emptyset$, a contradiction.
  - If $f'(u) \in C$, then $C \ni f(u) = f'(u) \sim f'(v) = v' \implies f(u) \sim v$.
  - If $f'(u) \in D'$, then $f'(u) \sim f'(v) = v' \implies f(u) \sim v$.

□

3.6.4 Putting things together

Proof of Theorem 3.59. By Lemma 3.66, if $\mathfrak{C}_{12}$ is constructible, there exists a construction of it by one doubling and one collapse starting from some H_0 that is not collapsible onto $\mathfrak{C}_{12}$ in the first place. But this is impossible by Lemmas 3.69 and 3.74. □

Remark 3.75. Our analysis, except for the computer-assisted part, does not depend on the number of vertices in $\mathfrak{C}_n$. Further program runs confirmed that also $\mathfrak{C}_{14}$ and $\mathfrak{C}_{16}$ are not constructible. On the other hand, one can see that $\mathfrak{C}_8$ and $\mathfrak{C}_{10}$ are constructible. $\Diamond$

Appendix A

Proof of Theorem 2.12

Our proof of Theorem 2.12 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 for convenience:

Theorem 2.12. *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 $\epsilon > 0$, there exists $\tau := \tau(\epsilon, \rho, \alpha, \ell) > 0$ such that if functions $f^{(1)}, \ldots, f^{(\ell)} : \underline{\Omega} \to [0,1]$ satisfy

$$\max_{i\in[n], j\in[\ell]} \mathrm{Inf}_i(f^{(j)}(\underline{X}^{(j)})) \leq \tau \,, \tag{11}$$

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

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \geq \left(\prod_{j=1}^{\ell} \mu^{(j)}\right)^{\ell/(1-\rho^2)} - \epsilon \,. \tag{12}$$

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

$$\tau := \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}}\right)^{C\frac{\ell \ln(\ell/\epsilon)\ln(1/\alpha)}{(1-\rho)\epsilon}} \,. \tag{13}$$

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

$$\int_\Omega f^2 \, \mathrm{d}\mathcal{P} < +\infty \,, \tag{67}$$

with inner product defined as

$$\langle f, g \rangle := \int_\Omega fg \, \mathrm{d}\mathcal{P} \,. \tag{68}$$

$\Diamond$

Remark A.2. As we will see shortly, if X is a random variable sampled from Ω according to $\mathcal{P}$, the equations (67) and (68) can be written as

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

$\Diamond$

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. $\Diamond$

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 $X : \Sigma \to \Sigma'$. $\Diamond$

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 $A \in \mathcal{F}$:

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

$\Diamond$

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

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

$\Diamond$

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". $\Diamond$

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

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

$\Diamond$

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 $(\mathcal{X}_0, \ldots, \mathcal{X}_p)$ of random variables *orthonormal* if they satisfy $\mathrm{E}[\mathcal{X}_k^2] = 1$ for every k and $\mathrm{E}[\mathcal{X}_j \mathcal{X}_k] = 0$ for every $j \neq k$. ◊

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

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

The notation $\mathcal{X}_{\star,k}$ 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 $\mathcal{X}_{\star,k}$ within an orthonormal ensemble with the orthonormal ensemble $\mathcal{X}_i$ itself. Whenever a random variable $\mathcal{X}_{i,k}$ is part of an ensemble $\mathcal{X}_i$, 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 $\mathrm{E}[\mathcal{X}_{\star,k}] = \mathrm{E}[\mathcal{X}_{\star,k}\mathcal{X}_{\star,0}] = 0$.

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

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

A *monomial compatible with* $\underline{\mathcal{X}}$ 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* $\underline{\mathcal{X}}$ 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 $\operatorname{supp}(\sigma) := \{i \in [n] : \sigma_i \neq 0\}$ and its *degree* as the size of its support: $|\sigma| := |\operatorname{supp}(\sigma)|$. Also, we will write the tuple $(0, \ldots, 0)$ as 0^n. ◊

Let a multilinear polynomial P compatible with $\underline{\mathcal{X}}$ be given. Then, $P(\underline{\mathcal{X}})$ 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 $\mathcal{X}_\sigma$ for the random variable corresponding to the evaluation of the monomial x_σ.

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

$$\mathrm{E}[\mathcal{X}_\sigma \mathcal{X}_\tau] = \begin{cases} 1 & \text{if } \sigma = \tau \\ 0 & \text{otherwise} \end{cases} \tag{69}$$

and

$$\mathrm{E}[\mathcal{X}_\sigma] = \begin{cases} 1 & \text{if } \sigma = 0^n \\ 0 & \text{otherwise.} \end{cases} \tag{70}$$

Proof. By independence of the coordinates we have $\mathrm{E}[\mathcal{X}_\sigma \mathcal{X}_\tau] = \prod_{i=1}^n \mathrm{E}[\mathcal{X}_{i,\sigma_i} \cdot \mathcal{X}_{i,\tau_i}]$ and now we can use the orthonomality of each ensemble $\mathcal{X}_i$. 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:

$$\deg(P) := \begin{cases} \max_{\sigma:\alpha_\sigma \neq 0} |\sigma| & \text{if } P \text{ is non-zero} \\ -\infty & \text{if } P \text{ is the zero polynomial} \end{cases} \tag{71}$$

$$\mathrm{E}[P] := \alpha(0^n) \tag{72}$$

$$\mathrm{E}[P^2] := \sum_\sigma \alpha(\sigma)^2 \tag{73}$$

$$\mathrm{Var}[P] := \mathrm{E}[P^2] - \mathrm{E}^2[P] \tag{74}$$

$$\mathrm{Inf}_i(P) := \sum_{\sigma:\sigma_i \neq 0} \alpha(\sigma)^2 \tag{75}$$

$$\mathrm{Inf}(P) := \sum_{i=1}^n \mathrm{Inf}_i(P) \tag{76}$$

◊

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 $\underline{\mathcal{X}}$ and a multilinear polynomial P compatible with it we have*

$$\mathrm{E}[P] = \mathrm{E}[P(\underline{\mathcal{X}})] \tag{77}$$
$$\mathrm{E}[P^2] = \mathrm{E}[(P(\underline{\mathcal{X}}))^2] \tag{78}$$
$$\mathrm{Var}[P] = \mathrm{Var}[P(\underline{\mathcal{X}})] \,. \tag{79}$$

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

$$\mathrm{Inf}_i(P) = \mathrm{E}\left[\mathrm{Var}\left[P(\underline{\mathcal{X}}) \mid \mathcal{X}_1, \ldots, \mathcal{X}_{i-1}, \mathcal{X}_{i+1}, \ldots, \mathcal{X}_n\right]\right] \,. \tag{80}$$

Proof. Linearity of expectation and (70) yield $\mathrm{E}[P(\underline{\mathcal{X}})] = \sum_\sigma \alpha(\sigma)\,\mathrm{E}[\mathcal{X}_\sigma] = \alpha(0^n)$, which is (77). Next, (69) gives $\mathrm{E}[P^2(\underline{\mathcal{X}})] = \sum_{\sigma,\tau} \alpha(\sigma)\alpha(\tau)\mathcal{X}_\sigma\mathcal{X}_\tau = \sum_\sigma \alpha(\sigma)^2$, i.e. (78), and hence (79) by the definition of the variance.

As for (80), fix an assignment $\underline{x}_{\setminus i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ to the ensemble sequence $\underline{\mathcal{X}}_{\setminus i} = (\mathcal{X}_1, \ldots, \mathcal{X}_{i-1}, \mathcal{X}_{i+1}, \ldots, \mathcal{X}_n)$.[1] We suppose that this tuple has a non-zero probability of occurence. Since $\mathcal{X}_i$ is an orthornormal ensemble,

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

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

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

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. □

[1] 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.

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,*

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

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

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

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,

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

□

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,

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

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

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

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

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

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

Proof.

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

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,

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

$\Diamond$

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. $\Diamond$

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 $(P^{(1)}(\underline{\mathcal{X}}^{(1)}), \ldots, P^{(\ell)}(\underline{\mathcal{X}}^{(\ell)}))$. $\Diamond$

A.3 Preliminaries — ensemble collections

In this section we recall the setting of Theorem 2.12 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. □

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

$$\begin{aligned} T_\rho f(\underline{x}) &= \mathrm{E}[f(\underline{Y}^{\rho,\underline{x}})] = \mathrm{E}[P(\underline{\mathcal{Y}}^{\rho,\underline{x}})] = \sum_\sigma \alpha(\sigma) \, \mathrm{E}[\mathcal{Y}_\sigma^{\rho,\underline{x}}] \\ &= \sum_\sigma \rho^{|\sigma|} \alpha(\sigma) \cdot \mathcal{X}_\sigma(\underline{x}) = T_\rho P(\underline{x}) \, . \end{aligned}$$

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

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 $\overline{\underline{X}}$ 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 $V(\mathcal{G}_\star)$ as

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

with the inner product of $A, B \in V(\mathcal{G}_\star)$ 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 $\overline{\mathcal{X}}_\star = (\mathcal{X}_\star^{(1)}, \ldots, \mathcal{X}_\star^{(\ell)})$ 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 $\overline{\mathcal{X}}_\star$ such that for all $j_1, j_2 \in [\ell]$, and all $k_1, k_2 \geq 0$ we have

$$\operatorname{Cov}\left[\mathcal{X}_{\star,k_1}^{(j_1)}, \mathcal{X}_{\star,k_2}^{(j_2)}\right] = \operatorname{Cov}\left[\mathcal{G}_{\star,k_1}^{(j_1)}, \mathcal{G}_{\star,k_2}^{(j_2)}\right] . \tag{81}$$

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 (81).

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 (81) 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 (81), their covariance matrix is identity. This finishes the proof, since joint Gaussians with the identity covariance matrix must be independent. □

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

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

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 $\mathcal{Z}_\star$. One easily checks that our construction gives

$$\begin{aligned}\mathcal{G}^{(1)}_{\star,0} = \mathcal{G}^{(2)}_{\star,0} &= \mathcal{H}_{\star,0}\\ \mathcal{G}^{(1)}_{\star,1} &= 1/2 \cdot \mathcal{H}_{\star,1} + \sqrt{3}/2 \cdot \mathcal{H}_{\star,2}\\ \mathcal{G}^{(2)}_{\star,1} &= 1/2 \cdot \mathcal{H}_{\star,1} - \sqrt{3}/2 \cdot \mathcal{H}_{\star,2}\ .\end{aligned}$$

$\Diamond$

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

Then, there exist Gaussian ensemble sequences $\underline{\overline{\mathcal{G}}} = (\underline{\mathcal{G}}^{(1)}, \dots, \underline{\mathcal{G}}^{(\ell)})$ compatible with $\underline{\overline{\mathcal{X}}}$ 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

$$\operatorname{Cov}\left[\mathcal{X}^{(j_1)}_{i_1,k_1}, \mathcal{X}^{(j_2)}_{i_2,k_2}\right] = \operatorname{Cov}\left[\mathcal{G}^{(j_1)}_{i_1,k_1}, \mathcal{G}^{(j_2)}_{i_2,k_2}\right]\ . \tag{82}$$

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

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

where

- $\underline{\overline{X}}$ is a random vector distributed according to $(\underline{\overline{\Omega}}, \underline{\mathcal{P}})$,
- $\underline{\mathcal{X}}^{(1)}, \dots, \underline{\mathcal{X}}^{(\ell)}$ are ensemble sequences constructed from $\underline{X}^{(1)}, \dots, \underline{X}^{(\ell)}$, respectively,
- and $\underline{\mathcal{G}}^{(1)}, \dots, \underline{\mathcal{G}}^{(\ell)}$ are obtained from Corollary A.33.

$\Diamond$

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

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

$\Diamond$

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

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 distribution (Ω, π) with $\min_{x\in\Omega} \pi(x) \geq \alpha \geq 0$.*

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 $1 \leq p \leq q < \infty$, $\rho \in [0,1]$ and let $\underline{\mathcal{X}} := (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ be an ensemble sequence such that for every $i \in [n]$, the ensemble $\mathcal{X}_i$ is (p,q,ρ)-hypercontractive. Then, the sequence $\underline{\mathcal{X}}$ 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 $\underline{\mathcal{X}} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ is $\underline{X}$-*Gaussian-mixed* if for each $i \in [n]$:

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

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

$\Diamond$

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 $\underline{\mathcal{X}}$ be an $\underline{X}$-Gaussian-mixed ensemble sequence.*

Then, $\underline{\mathcal{X}}$ 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 $\underline{\mathcal{X}}$ be an $\underline{X}$-Gaussian-mixed ensemble sequence. Let P be a multilinear polynomial compatible with $\underline{\mathcal{X}}$ of degree at most d. Then,*

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

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

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

□

Lemma A.43. *Let $\overline{\underline{X}}$ be a random vector distributed according to a (multistep) probability space with equal marginals $(\overline{\underline{\Omega}}, \underline{\mathcal{P}})$ with $\min_{x\in\Omega} \pi(x) \geq \alpha > 0$.*

Let $\underline{\mathcal{S}}^{(1)}, \ldots, \underline{\mathcal{S}}^{(\ell)}$ be ensemble sequences such that $\underline{\mathcal{S}}^{(j)}$ is $\underline{X}^{(j)}$-Gaussian-mixed. Let $P^{(1)}, \ldots, P^{(\ell)}$ be multilinear polynomials such that $P^{(j)}$ is compatible with $\underline{\mathcal{S}}^{(j)}$ and $\deg(P^{(j)}) \leq d$.

Then, for every triple $j_1, j_2, j_3 \in [\ell]$:

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

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

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

□

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 $\overline{x} = (x^{(1)}, \ldots, x^{(\ell)}) \in \mathbb{R}^\ell$ we have

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

◊

Theorem A.45 (Invariance Principle)**.** *Let $(\overline{X}, \overline{\mathcal{X}}, \overline{\mathcal{G}})$ be an ensemble collection for a probability space $(\overline{\Omega}, \overline{\mathcal{P}})$ 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 $\underline{\mathcal{X}}^{(j)}$.

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,

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

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 $\underline{\mathcal{U}}_{(i)}^{(j)}$ be defined as $\underline{\mathcal{U}}_{(i)}^{(j)} := (\mathcal{G}_1^{(j)}, \ldots, \mathcal{G}_i^{(j)}, \mathcal{X}_{i+1}^{(j)}, \ldots, \mathcal{X}_n^{(j)})$.

Claim A.47.

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

Proof. By the triangle inequality. □

Due to Claim A.47, we will estimate

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

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

$$P^{(j)}(\underline{\mathcal{T}}^{(j)}) = A^{(j)} + \sum_{k>0} \mathcal{X}^{(j)}_{i,k} \cdot B^{(j)}_k = A^{(j)} + P^{(j)}_i(\underline{\mathcal{T}}^{(j)}) \,, \tag{83}$$

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

$$P^{(j)}(\underline{\mathcal{U}}^{(j)}) = A^{(j)} + \sum_{k>0} \mathcal{G}^{(j)}_{i,k} \cdot B^{(j)}_k = A^{(j)} + P^{(j)}_i(\underline{\mathcal{U}}^{(j)}) \,.$$

We note for later use that the construction gives us

$$\deg(P^{(j)}_i) \le d \tag{84}$$

$$\mathrm{E}\left[\left(P^{(j)}_i\right)^2\right] = \mathrm{Inf}_i\left(P^{(j)}\right) \,. \tag{85}$$

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

$$\begin{aligned}
&\Psi(\overline{P}(\underline{\mathcal{T}})) - \Psi(\overline{P}(\underline{\mathcal{U}})) = \\
&\quad = \Psi\left(A^{(1)} + \sum_{k>0} \mathcal{X}^{(1)}_{i,k} B^{(1)}_k, \ldots, A^{(\ell)} + \sum_{k>0} \mathcal{X}^{(\ell)}_{i,k} B^{(\ell)}_k\right) \\
&\qquad - \Psi\left(A^{(1)} + \sum_{k>0} \mathcal{G}^{(1)}_{i,k} B^{(1)}_k, \ldots, A^{(\ell)} + \sum_{k>0} \mathcal{G}^{(\ell)}_{i,k} B^{(\ell)}_k\right)
\end{aligned}$$

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 $P^{(j)}_i$, 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 $\overline{x} = (x^{(1)}, \ldots, x^{(\ell)}), \overline{\epsilon} = (\epsilon^{(1)}, \ldots, \epsilon^{(\ell)}) \in \mathbb{R}^\ell$. Then,*

$$\left| \Psi\left(x^{(1)} + \epsilon^{(1)}, \ldots, x^{(\ell)} + \epsilon^{(\ell)}\right) - \left(\Psi(\overline{x}) + \sum_{j\in[\ell]} \epsilon^{(j)} \frac{\partial}{\partial x^{(j)}} \Psi(\overline{x}) + \frac{1}{2} \sum_{j_1,j_2\in[\ell]} \epsilon^{(j_1)}\epsilon^{(j_2)} \frac{\partial^2}{\partial x^{(j_1)}\partial x^{(j_2)}} \Psi(\overline{x})\right) \right| \leq \frac{B}{6} \sum_{j_1,j_2,j_3\in[\ell]} \left|\epsilon^{(j_1)}\epsilon^{(j_2)}\epsilon^{(j_3)}\right| .$$

We omit the proof of Theorem A.48.

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

$$\mathrm{E}\left[\Psi(\overline{P}(\underline{\mathcal{T}}))\right] = \mathrm{E}\left[\Psi(\overline{A}) + \frac{1}{2} \sum_{j_1,j_2\in[\ell]} \left(\sum_{k_1,k_2>0} \mathcal{X}^{(j_1)}_{i,k_1} \mathcal{X}^{(j_2)}_{i,k_2} B^{(j_1)}_{k_1} B^{(j_2)}_{k_2} \frac{\partial^2}{\partial A^{(j_1)} \partial A^{(j_2)}} \Psi(\overline{A}) \right) + R_{\underline{\mathcal{T}}} \right], \quad (86)$$

and

$$\mathrm{E}\left[\Psi(\overline{P}(\underline{\mathcal{U}}))\right] = \mathrm{E}\left[\Psi(\overline{A}) + \frac{1}{2} \sum_{j_1,j_2\in[\ell]} \left(\sum_{k_1,k_2>0} \mathcal{G}^{(j_1)}_{i,k_1} \mathcal{G}^{(j_2)}_{i,k_2} B^{(j_1)}_{k_1} B^{(j_2)}_{k_2} \frac{\partial^2}{\partial A^{(j_1)} \partial A^{(j_2)}} \Psi(\overline{A}) \right) + R_{\underline{\mathcal{U}}} \right], \quad (87)$$

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

$$\mathrm{E}\left[\left|R_{\underline{\mathcal{T}}}\right|\right], \mathrm{E}\left[\left|R_{\underline{\mathcal{U}}}\right|\right] \leq \frac{\ell^{3/2} B}{6} \left(\frac{8}{\sqrt{\alpha}}\right)^d \left(\sum_{j=1}^{\ell} \mathrm{Inf}_i(P^{(j)}) \right)^{3/2} . \quad (88)$$

Proof. We show only (86) and the bound on $\mathrm{E}[|R_{\underline{\mathcal{T}}}|]$, the proofs for the ensemble sequence $\underline{\mathcal{U}}$ 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 (83), by Theorem A.48

$$\Psi(\overline{P}(\overline{\underline{T}})) = \Psi(\overline{A}) + \sum_{j\in[\ell]} \left(\sum_{k>0} \mathcal{X}_{i,k}^{(j)} B_k^{(j)} \frac{\partial}{\partial A^{(j)}} \Psi(\overline{A}) \right) + \\ + \frac{1}{2} \sum_{j_1,j_2\in[\ell]} \left(\sum_{k_1,k_2>0} \mathcal{X}_{i,k_1}^{(j_1)} \mathcal{X}_{i,k_2}^{(j_2)} B_{k_1}^{(j_1)} B_{k_2}^{(j_2)} \frac{\partial^2}{\partial A^{(j_1)} \partial A^{(j_2)}} \Psi(\overline{A}) \right) + R_{\underline{\mathcal{T}}} \,, \quad (89)$$

where

$$\mathrm{E}[|R_{\underline{\mathcal{T}}}|] \le \frac{B}{6} \sum_{j_1,j_2,j_3\in[\ell]} \mathrm{E}\left[\left| \prod_{k=1}^{3} P_i^{(j_k)}(\underline{\mathcal{T}}^{(j_k)}) \right| \right] . \quad (90)$$

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

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

which together with (89) yields (86).

As for the bound on $\mathrm{E}[|R_{\underline{\mathcal{T}}}|]$, since $T^{(j)}$ is $\underline{X}^{(j)}$-Gaussian-mixed ensemble sequence, due to (90), Lemma A.43 (note that the degree is bounded due to (84)), and (85),

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

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

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,*

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

Proof. First, we need to show that the second-order terms in (86) and (87) 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

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

Therefore, by (86), (87) and (88),

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

as claimed. □

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

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

□

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 $(\underline{X}, \underline{\mathcal{X}}, \underline{\mathcal{G}})$ be an ensemble collection for a probability space $(\underline{\Omega}, \underline{\mathcal{P}})$ 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 $\underline{\mathcal{X}}^{(j)}$.

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

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

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 theorem from real calculus, 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 $(f*g)(x) := \int_{-\infty}^{\infty} f(x-t)g(t)\,\mathrm{d}t$. ◊

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* $[-1, 1]$.
- $\forall x : \psi(x) = \psi(-x)$.
- $\int_{-\infty}^{\infty} \psi(x)\,\mathrm{d}x = \int_{-1}^{1} \psi(x)\,\mathrm{d}x = 1$.

Proof. Consider

$$\Psi(x) := \begin{cases} \exp(-\frac{1}{(x+1)^2}) \cdot \exp(-\frac{1}{(x-1)^2}) & \text{if } x \in (-1,1) \\ 0 & \text{otherwise} \end{cases} \tag{91}$$

and set $\psi(x) := \Psi(x)/c$ where $c := \int_{-1}^{1} \Psi(x)\,\mathrm{d}x$. □

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 $z := y/\lambda$)

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

as claimed.

□

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

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

◊

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

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 $(\overline{\underline{X}}, \overline{\mathcal{X}}, \overline{\underline{\mathcal{G}}})$ 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 $\underline{\mathcal{X}}^{(j)}$.

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

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

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

$$\begin{aligned}\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathcal{X}})) - \chi(\overline{P}(\underline{\mathcal{G}}))\right]\right| \leq{}& \left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathcal{X}})) - \chi_\lambda(\overline{P}(\underline{\mathcal{X}}))\right]\right| \\ &+ \left|\mathrm{E}\left[\chi_\lambda(\overline{P}(\underline{\mathcal{X}})) - \chi_\lambda(\overline{P}(\underline{\mathcal{G}}))\right]\right| \\ &+ \left|\mathrm{E}\left[\chi_\lambda(\overline{P}(\underline{\mathcal{G}})) - \chi(\overline{P}(\underline{\mathcal{G}}))\right]\right| . \end{aligned} \tag{92}$$

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

$$\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathcal{X}})) - \chi_\lambda(\overline{P}(\underline{\mathcal{X}}))\right]\right| \leq \ell\lambda \leq O\left(\tfrac{\ell^{5/2}\tau^{1/8}}{\alpha^{4d}}\right) \tag{93}$$

$$\left|\mathrm{E}\left[\chi_\lambda(\overline{P}(\underline{\mathcal{G}})) - \chi(\overline{P}(\underline{\mathcal{G}}))\right]\right| \leq \ell\lambda \leq O\left(\tfrac{\ell^{5/2}\tau^{1/8}}{\alpha^{4d}}\right) . \tag{94}$$

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

$$\left|\mathrm{E}\left[\chi_\lambda(\overline{P}(\underline{\mathcal{X}})) - \chi_\lambda(\overline{P}(\underline{\mathcal{G}}))\right]\right| \leq O\left(\frac{\ell^{5/2} d 8^d \tau^{1/2}}{\lambda^3 \alpha^{d/2}}\right) . \tag{95}$$

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

$$\frac{\ell^{5/2} d 8^d \tau^{1/2}}{\lambda^3 \alpha^{d/2}} \leq O\left(\frac{\ell^{5/2} d 8^d \tau^{1/8}}{\alpha^{d/2}}\right) \leq O\left(\frac{\ell^{5/2}\tau^{1/8}}{\alpha^{4d}}\right) . \tag{96}$$

Inserting (93), (94), and the combination of (96) and (95) into (92) 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 $|\mathrm{E}[\chi(\overline{P}^{<d}(\underline{\mathcal{X}})) - \chi(\overline{P}(\underline{\mathcal{X}}))]|$ for appropriately chosen d.

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

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

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

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

$$\begin{aligned}&\left|\chi(x^{(1)} + \epsilon^{(1)}, \ldots, x^{(\ell)} + \epsilon^{(\ell)}) - \chi(x^{(1)}, \ldots, x^{(\ell)})\right| \\ &\quad\leq \sum_{j=1}^{\ell} \left|\chi(\overline{y}_{(j-1)}) - \chi(\overline{y}_{(j)})\right| \leq \sum_{j=1}^{\ell} \left|\epsilon^{(j)}\right| .\end{aligned}$$

□

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

$$\begin{aligned}\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathcal{X}})) - \chi(\overline{P}(\underline{\mathcal{G}}))\right]\right| \leq &\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathcal{X}})) - \chi(\overline{P}^{<d}(\underline{\mathcal{X}}))\right]\right| \\ &+ \left|\mathrm{E}\left[\chi(\overline{P}^{<d}(\underline{\mathcal{X}})) - \chi(\overline{P}^{<d}(\underline{\mathcal{G}}))\right]\right| \\ &+ \left|\mathrm{E}\left[\chi(\overline{P}^{<d}(\underline{\mathcal{G}})) - \chi(\overline{P}(\underline{\mathcal{G}}))\right]\right| . \end{aligned} \tag{97}$$

We proceed to demonstrate that all three terms on the right hand side of (97) are $O\left(\ell^4 \tau^{\Omega\left(\frac{\gamma}{\ln 1/\alpha}\right)}\right)$, which will finish the proof.

Lemma A.67.

$$\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathcal{X}})) - \chi(\overline{P}^{<d}(\underline{\mathcal{X}}))\right]\right| \leq \ell(1-\gamma)^{d/2} \leq O\left(\ell \tau^{\Omega\left(\frac{\gamma}{\ln 1/\alpha}\right)}\right) \tag{98}$$

and, similarly,

$$\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathcal{G}})) - \chi(\overline{P}^{<d}(\underline{\mathcal{G}}))\right]\right| \leq \ell(1-\gamma)^{d/2} \leq O\left(\ell \tau^{\Omega\left(\frac{\gamma}{\ln 1/\alpha}\right)}\right) \tag{99}$$

Proof. We prove only (98), the argument for (99) being the same. Using Lemma A.66, Cauchy-Schwarz, the fact that $\overline{P}$ is γ-decaying and the definition of d,

$$\begin{aligned}\left|\mathrm{E}\left[\chi(\overline{P}(\underline{\mathcal{X}})) - \chi(\overline{P}^{<d}(\underline{\mathcal{X}}))\right]\right| &\leq \sum_{j=1}^{\ell} \mathrm{E}\left[\left|\left(P^{(j)}\right)^{\geq d}(\underline{\mathcal{X}}^{(j)})\right|\right] \\ &\leq \sum_{j=1}^{\ell} \sqrt{\mathrm{E}\left[\left(P^{(j)}\right)^2\right]} \leq \ell(1-\gamma)^{d/2} \leq 2\ell\tau^{\frac{\gamma}{128 \ln 1/\alpha}} . \end{aligned}$$

□

Lemma A.68.

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

Proof. From Theorem A.65,

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

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. □

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 $\underline{\mathcal{Z}}$ be an ensemble sequence constructed from $\overline{X}$ and $\underline{\mathcal{X}}^{(1)}, \ldots, \underline{\mathcal{X}}^{(\ell)}$ be ensemble sequences constructed from $\underline{X}^{(1)}, \ldots, \underline{X}^{(\ell)}$, respectively.*

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

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

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

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 $Q^{(j)}$. □

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

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

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 $i \in [n]$, $k, k' \geq 0$. 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

$$|\mathrm{E}\left[\mathcal{X}_{i,k} \cdot \mathcal{Y}_{i,k'}\right]| = |\mathrm{Cov}\left[\mathcal{X}_{i,k}, \mathcal{Y}_{i,k'}\right]| \leq \rho. \tag{100}$$

Expanding the expectation and using (100) and Cauchy-Schwarz,

$$\begin{aligned}
|\mathrm{E}\left[P_S(\underline{\mathcal{X}}) Q_S(\underline{\mathcal{Y}})\right]| &= \left| \mathrm{E}\left[\left(\sum_{\sigma:\mathrm{supp}(\sigma)=S} \alpha(\sigma) \mathcal{X}_\sigma \right) \left(\sum_{\sigma':\mathrm{supp}(\sigma')=S} \beta(\sigma') \mathcal{Y}_{\sigma'} \right) \right] \right| \\
&\leq \sum_{\substack{\sigma,\sigma': \\ \mathrm{supp}(\sigma)=\mathrm{supp}(\sigma')=S}} \left| \alpha(\sigma)\beta(\sigma') \prod_{i \in S} \mathrm{E}\left[\mathcal{X}_{i,\sigma_i} \mathcal{Y}_{i,\sigma'_i} \right] \right| \\
&\leq \rho^{|S|} \sum_{\substack{\sigma,\sigma': \\ \mathrm{supp}(\sigma)=\mathrm{supp}(\sigma')=S}} |\alpha(\sigma)\beta(\sigma')| \\
&\leq \rho^{|S|} \sqrt{\mathrm{Var}[P_S] \, \mathrm{Var}[Q_S]} \,,
\end{aligned}$$

□

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

Proof. If $\rho \in \{0,1\}$ we are done, therefore assume that $\rho \in (0,1)$. If $k \ge \log_\rho \ell/\epsilon$, then $\rho^k \le \epsilon/\ell$.

If $0 \le k < \log_\rho \ell/\epsilon$, then by Bernoulli's inequality,

$$1-(1-\gamma)^k \le \gamma k \le \frac{1-\rho}{\ln(1/\rho)} \cdot \frac{\epsilon}{\ell} \le \frac{\epsilon}{\ell} .$$

□

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

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

Proof. As in the proof of Lemma A.71, we will write $P := P^{(j)}$, $Q := Q^{(j)}$, $\underline{\mathcal{X}} := \underline{\mathcal{X}}^{(j)}$ and $\underline{\mathcal{Y}} := \underline{\mathcal{Y}}^{(j)}$.

By definition of $T_{1-\gamma}$,

$$(\mathrm{Id} - T_{1-\gamma}) P_S(\underline{\mathcal{X}}) = (1-(1-\gamma)^{|S|}) P_S(\underline{\mathcal{X}}) . \tag{101}$$

From (101), Lemma A.71 and Lemma A.72,

$$\begin{aligned} &|\mathrm{E}\left[(\mathrm{Id} - T_{1-\gamma}) P_S(\underline{\mathcal{X}}) \cdot Q_S(\underline{\mathcal{Y}}) \right]| \\ &\quad \le \min\left(1-(1-\gamma)^{|S|}, \rho^{|S|}\right) \sqrt{\mathrm{Var}[P_S] \, \mathrm{Var}[Q_S]} \\ &\quad \le \frac{\epsilon}{\ell} \sqrt{\mathrm{Var}[P_S] \, \mathrm{Var}[Q_S]} . \end{aligned}$$

□

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

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

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 $\mathrm{Var}[P], \mathrm{Var}[Q] \le 1$.

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

$$\begin{aligned} |\mathrm{E}\left[(\mathrm{Id} - T_{1-\gamma}) P(\underline{\mathcal{X}}) \cdot Q(\underline{\mathcal{Y}}) \right]| &\le \sum_{S \subseteq [n]} |\mathrm{E}\left[(\mathrm{Id} - T_{1-\gamma}) P_S(\underline{\mathcal{X}}) \cdot Q_S(\underline{\mathcal{Y}}) \right]| \\ &\le \frac{\epsilon}{\ell} \sum_{S \neq \emptyset} \sqrt{\mathrm{Var}[P_S] \, \mathrm{Var}[Q_S]} \\ &\le \frac{\epsilon}{\ell} \sqrt{\mathrm{Var}[P] \, \mathrm{Var}[Q]} \le \epsilon/\ell . \end{aligned}$$

□

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

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

□

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

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

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

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 $\overline{G} = (\underline{G}^{(1)}, \ldots, \underline{G}^{(\ell)})$ 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* $\{\alpha_i^{(j)}\} \in \mathbb{R}$*:*

$$\operatorname{Var}\left[\sum_{i,j} \alpha_i^{(j)} \cdot G_i^{(j)}\right] \geq p \cdot \sum_{i,j} \left(\alpha_i^{(j)}\right)^2 . \tag{102}$$

Then, for all $f^{(1)}, \ldots, f^{(\ell)} \in L^2(\mathbb{R}^n, \gamma^n)$ *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 (102) is that the matrix $(T - p\,\mathrm{Id})$ is positive semidefinite, where T is the covariance matrix of $\overline{G}$. ◊

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* $(\overline{\Omega}, \underline{\mathcal{P}})$ *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 $\{\alpha_k^{(j)}\} \in \mathbb{R}$*:*

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

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

We compute

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

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

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

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

Therefore,

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

□

Lemma A.81. *Let $(\underline{X}, \overline{\mathcal{X}}, \overline{\mathcal{G}})$ be an ensemble collection for a probability space $(\overline{\Omega}, \underline{\mathcal{P}})$ with $\rho(\mathcal{P}) \le \rho < 1$.*

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

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

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

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

□

Lemma A.82. *Let $(\underline{X}, \overline{\mathcal{X}}, \overline{\mathcal{G}})$ be an ensemble collection for a probability space $(\overline{\Omega}, \underline{\mathcal{P}})$ with $\rho(\mathcal{P}) \le \rho < 1$.*

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

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

Proof. By Corollary A.31 and Lemma A.81. □

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

Since $\underline{\mathcal{G}}^{(j)}$ is a Gaussian ensemble sequence, $\underline{G}^{(j)}$ is distributed as pn independent $\mathcal{N}(0,1)$ Gaussians. Condition (102) for $p := \frac{1-\rho^2}{\ell}$ is fulfilled due to Lemma A.82. □

A.9 The main theorem

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

Theorem 2.12. *Let $\overline{X}$ be a random vector distributed according to $(\overline{\Omega}, \underline{\mathcal{P}})$ such that $\mathcal{P}$ has equal marginals, $\rho(\mathcal{P}) \le \rho < 1$ and $\min_{x\in\Omega} \pi(x) \ge \alpha > 0$.*

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

$$\max_{i\in[n], j\in[\ell]} \mathrm{Inf}_i(f^{(j)}(\underline{X}^{(j)})) \le \tau \,, \tag{11}$$

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

$$\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})\right] \ge \left(\prod_{j=1}^{\ell} \mu^{(j)}\right)^{\ell/(1-\rho^2)} - \epsilon \,. \tag{12}$$

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

$$\tau := \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}}\right)^{C\frac{\ell \ln(\ell/\epsilon)\ln(1/\alpha)}{(1-\rho)\epsilon}} \,. \tag{13}$$

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 $(\overline{\Omega}, \underline{\mathcal{P}})$.

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 $\mathcal{X}^{(j)}_i$ and $\mathcal{G}^{(j)}_i$, 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:

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

$$= \mathrm{E}\left[\prod_{j=1}^{\ell} P^{(j)}(\underline{\mathcal{X}}^{(j)}) - \prod_{j=1}^{\ell} Q^{(j)}(\underline{\mathcal{X}}^{(j)})\right] + \tag{103}$$

$$+ \mathrm{E}\left[\prod_{j=1}^{\ell} Q^{(j)}(\underline{\mathcal{X}}^{(j)}) - \prod_{j=1}^{\ell} R^{(j)}(\underline{\mathcal{G}}^{(j)})\right] + \tag{104}$$

$$+ \mathrm{E}\left[\prod_{j=1}^{\ell} R^{(j)}(\mathcal{G}^{(j)})\right] . \tag{105}$$

We use the theorems proved so far to bound each of the terms (103), (104) and (105) in turn. First, we apply Theorem A.69 to show that (103) has small absolute value. Then, we use the invariance principle (Theorem A.53) to argue that (104) has small absolute value. Finally, using Gaussian reverse hypercontractivity (Theorem A.76) we show that (105) 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 $\epsilon \le 1/2$ and $\alpha \le 1/2$.

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

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

Proof. By Theorem A.69. □

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

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

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,

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

□

Lemma A.85.

$$\mathrm{E}\left[\prod_{j=1}^{\ell} R^{(j)}(\underline{\mathcal{G}}^{(j)})\right] \geq \left(\prod_{j=1}^{\ell} \mu'^{(j)}\right)^{\ell/(1-\rho^2)} \ .$$

Proof. By Theorem A.76. □

Lastly, we need to show that the difference between the values $\prod_{j=1}^{\ell} \mu'^{(j)}$ and $\prod_{j=1}^{\ell} \mu^{(j)}$ is small.

Claim A.86. *Let $a \geq 0, \epsilon \geq 0, a + \epsilon \leq 1, \beta \geq 1$. Then, $(a+\epsilon)^\beta - a^\beta \leq \beta\epsilon$.*

Proof. The function $h_{\beta,\epsilon}(a) := (a+\epsilon)^\beta - a^\beta$ is non-decreasing (since $\frac{\mathrm{d}}{\mathrm{d}a} h_{\beta,\epsilon} = \beta((a+\epsilon)^{\beta-1} - a^{\beta-1}) \geq 0$). Hence,

$$(a+\epsilon)^\beta - a^\beta \leq 1 - (1-\epsilon)^\beta \leq \beta\epsilon \ ,$$

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,

$$\left| \left(\prod_{j=1}^{\ell} \mu^{(j)}\right)^{\ell/(1-\rho^2)} - \left(\prod_{j=1}^{\ell} \mu'^{(j)}\right)^{\ell/(1-\rho^2)} \right| \leq \frac{\ell}{1-\rho^2} \cdot \left| \prod_{j=1}^{\ell} \mu^{(j)} - \prod_{j=1}^{\ell} \mu'^{(j)} \right| \ . \tag{106}$$

Since $\mu^{(j)}, \mu'^{(j)} \in [0,1]$,

$$\left| \prod_{j=1}^{\ell} \mu^{(j)} - \prod_{j=1}^{\ell} \mu'^{(j)} \right| \leq \sum_{j=1}^{\ell} \left| \mu^{(j)} - \mu'^{(j)} \right| . \tag{107}$$

For a fixed $j \in [\ell]$, from the definition of χ and Theorem A.53 applied with $\ell = 1$,

$$\left| \mu^{(j)} - \mu'^{(j)} \right| = \left| \mathrm{E}\left[\chi\left(Q^{(j)}(\underline{\mathcal{X}}^{(j)}) \right) - \chi\left(Q^{(j)}(\underline{\mathcal{G}}^{(j)}) \right) \right] \right| \leq C \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} . \tag{108}$$

Inequalities (106), (107) and (108) together give the claim. □

Proof of Theorem 2.12. Following the decomposition of $\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)})$ into subexpressions (103), (104) and (105), from Lemma A.83, Lemma A.84, Lemma A.85 and Lemma A.87,

$$\begin{aligned}
&\mathrm{E}\left[\prod_{j=1}^{\ell} f^{(j)}(\underline{X}^{(j)}) \right] \\
&\quad \geq \left(\prod_{j=1}^{\ell} \mu^{(j)} \right)^{\ell/(1-\rho^2)} - \epsilon/2 - C\ell^{5/2} \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} - \frac{C\ell^2}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} \\
&\quad \geq \left(\prod_{j=1}^{\ell} \mu^{(j)} \right)^{\ell/(1-\rho^2)} - \epsilon/2 - \frac{2C\ell^{5/2}}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} .
\end{aligned}$$

By choosing $\tau(\epsilon, \rho, \alpha, \ell, \gamma)$ small enough we get

$$\frac{2C\ell^{5/2}}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} \leq \epsilon/2 , \tag{109}$$

which is the main part of the theorem (recall that $\gamma = \frac{(1-\rho)\epsilon}{2\ell \ln(2\ell/\epsilon)}$).

To see that we can choose τ as in (13), note that for $D > 0$ big enough we have

$$\begin{aligned}
\tau &:= \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}} \right)^{D \frac{\ell \ln(\ell/\epsilon) \ln(1/\alpha)}{(1-\rho)\epsilon}} \leq \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}} \right)^{D' \frac{2C\ell \ln(2\ell/\epsilon) \ln(1/\alpha)}{(1-\rho)\epsilon}} \\
&= \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}} \right)^{D' \frac{C \ln(1/\alpha)}{\gamma}}
\end{aligned}$$

for $D' > 0$ as needed. Hence, we obtain

$$\frac{2C\ell^{5/2}}{1-\rho^2} \cdot \tau^{\frac{\gamma}{C \ln 1/\alpha}} = 2C \cdot \frac{\ell^{5/2}}{1-\rho^2} \cdot \left(\frac{(1-\rho^2)\epsilon}{\ell^{5/2}}\right)^{D'} \leq 2C\epsilon^{D'} \leq \epsilon/2 \,,$$

which establishes (109) for this choice of τ. □

Appendix B

Listings of Computer-Assisted Proofs

Here we provide program codes for the computer-assisted proofs from Section 3.6. The programs are written in C++.

Listing 1: `construction.hpp` — The header file used by all programs.

```
#include <cassert>
#include <climits>
#include <cstdio>
#include <cstdlib>
#include <algorithm>
#include <vector>
using namespace std;

// Bitshifts have higher priority than comparisons.
// Comparisons have higher priority than bit
// operations.

// Mathematical modulo.
// Precondition: MOD > 0
inline int mod(int x, int MOD) {
  x %= MOD;
  return x + (x < 0 ? MOD : 0);
}

// Number of bits set to one in u.
struct PopCounter {
  int pcnt[1<<16];
  PopCounter() {
```

```
24      assert(CHAR_BIT == 8 && sizeof(unsigned) == 4);
25      for (int i = 1; i < 1<<16; ++i)
26        pcnt[i] = pcnt[i/2] + i%2;
27    }
28  } P;
29  inline int popcount (unsigned u) {
30    return P.pcnt[u & ((1<<16)-1)] + P.pcnt[u >> 16];
31  }
32
33  // Undirected graph with set of vertices S.
34  // Invariant: all edges inside S.
35  struct Graph {
36    unsigned S;
37    vector<unsigned> M;
38
39    Graph(const unsigned a_S,
40        const vector<unsigned>& a_M):
41      S(a_S), M(a_M) { }
42  };
43
44  // Doubles subset S of V(G).
45  // between[uprim] & (1<<u) indicates edge between
46  // u' in V(G') and u in V(G).
47  // Precondition: T is a subset of G.S
48  void double_graph(const unsigned T, const Graph& G,
49      Graph& Gprim, vector<unsigned>& between) {
50    const vector<unsigned>& M = G.M;
51    vector<unsigned>& Mprim = Gprim.M;
52    const int N = M.size();
53
54    Mprim.resize(N);
55    between.resize(N);
56    Gprim.S = T;
57
58    for (int u = 0; u < N; ++u)
59      if (1<<u & T) {
60        Mprim[u] = M[u] & T;
61        between[u] = M[u] & ~T;
62      } else Mprim[u] = between[u] = 0;
63  }
64
```

```
65  // Exchange vertices in T between G and G'.
66  // Precondition: T is a subset of G.S \cap Gprim.S
67  void exchange(const unsigned T, Graph& G, Graph& Gprim,
68       vector<unsigned>& between) {
69    vector<unsigned>& M = G.M;
70    vector<unsigned>& Mprim = Gprim.M;
71    const int N = M.size();
72
73    for (int u = 0; u < N; ++u) if (1<<u & T) {
74      unsigned old_M = M[u], old_Mprim = Mprim[u],
75        old_between = between[u];
76
77      M[u] = old_between & ~(1<<u);
78      for (int v = 0; v < N; ++v) {
79        M[v] &= ~(1<<u);
80        if (M[u] & 1<<v) M[v] |= 1<<u;
81      }
82
83      // It is important that 'between' has not been
84      // modified yet.
85      Mprim[u] = 0;
86      for (int vprim = 0; vprim < N; ++vprim)
87        if (u != vprim) {
88          Mprim[vprim] &= ~(1<<u);
89          if (between[vprim] & 1<<u) {
90            Mprim[u] |= 1<<vprim;
91            Mprim[vprim] |= 1<<u;
92          }
93        }
94
95      between[u] = old_M;
96      if (old_between & 1<<u) between[u] |= 1<<u;
97      for (int vprim = 0; vprim < N; ++vprim)
98        if (u != vprim) {
99          between[vprim] &= ~(1<<u);
100         if (old_Mprim & 1<<vprim)
101           between[vprim] |= 1<<u;
102       }
103   }
104 }
105
```

```
106 // Can Gprim be collapsed onto G?
107 // If yes, 'mapping' will contain a mapping
108 // from Gprim to G,  with mapping[uprim] == -1
109 // for uprim not in Gprim.S.
110 bool is_collapsible(const Graph& a_G,
111     const Graph& a_Gprim,
112     const vector<unsigned>& a_between,
113     vector<int>& a_mapping) {
114   struct RecursiveData {
115     const Graph& G;
116     const Graph& Gprim;
117     const vector<unsigned>& between;
118     vector<int>& mapping;
119     const vector<unsigned>& M;
120     const vector<unsigned>& Mprim;
121     const int N;
122
123     RecursiveData(const Graph& a_G,
124         const Graph& a_Gprim,
125         const vector<unsigned>& a_between,
126         vector<int>& a_mapping):
127           G(a_G), Gprim(a_Gprim), between(a_between),
128           mapping(a_mapping), M(G.M),  Mprim(Gprim.M),
129           N(M.size()) {
130       mapping.resize(N);
131       fill_n(mapping.begin(), N, -1);
132     }
133
134     bool is_collapsible_rec(int uprim) {
135       if (uprim == N) return true;
136       if (1<<uprim & ~Gprim.S)
137         return is_collapsible_rec(uprim+1);
138       // invariant: u' < N and u' in V(G')
139
140       for (int u = 0; u < N; ++u) if (1<<u & G.S) {
141         if ((M[u] & between[uprim]) != between[uprim])
142           continue;
143         // invariant: u' -> u preserves edges between
144         // u' and G
145
146         bool ok = true;
```

```
147           for (int vprim = 0; vprim < uprim && ok;
148               ++vprim) {
149             if (Mprim[uprim] & 1<<vprim &&
150                 !(M[u] & 1<<mapping[vprim])) {
151               ok = false;
152             }
153           }
154           if (!ok) continue;
155           // invariant: u' -> u preserves edges between
156           // u' and preceding vertices in G'
157
158           mapping[uprim] = u;
159           if (is_collapsible_rec(uprim+1)) return true;
160         }
161         return false;
162       }
163     } R(a_G, a_Gprim, a_between, a_mapping);
164
165     return R.is_collapsible_rec(0);
166   }
167
168   // Can G' be collapsed onto G such that both T and
169   // G'.S \setminus T do not collapse naturally?
170   // If yes, mapping will contain such mapping from
171   // G' to G, with mapping[u'] == -1 for
172   // u' not in G'.S.
173   // Precondition: T is a subset of G'.S which is
174   // a subset of G.S
175   bool is_unnaturally_collapsible(const unsigned a_T,
176       const Graph& a_G, const Graph& a_Gprim,
177       const vector<unsigned>& a_between,
178       vector<int>& a_mapping) {
179     struct RecursiveData {
180       const unsigned T;
181       const Graph& G;
182       const Graph& Gprim;
183       const vector<unsigned>& between;
184       vector<int>& mapping;
185       const vector<unsigned>& M;
186       const vector<unsigned>& Mprim;
187       const int N;
```

```
188
189     RecursiveData(const unsigned a_T, const Graph& a_G,
190         const Graph& a_Gprim,
191         const vector<unsigned>& a_between,
192         vector<int>& a_mapping):
193           T(a_T), G(a_G), Gprim(a_Gprim),
194           between(a_between), mapping(a_mapping),
195           M(G.M), Mprim(Gprim.M), N(M.size()) {
196       mapping.resize(N);
197       fill_n(mapping.begin(), N, -1);
198     }
199
200     bool is_collapsible_rec(int uprim) {
201       if (uprim == N) {
202         bool ok1 = false, ok2 = false;
203         for (int uprim = 0; uprim < N &&
204             (!ok1 || !ok2); ++uprim) {
205           if (1<<uprim & ~Gprim.S) continue;
206           if (1<<uprim & T && mapping[uprim] != uprim)
207             ok1 = true;
208           else if (1<<uprim & ~T &&
209               mapping[uprim] != uprim) {
210             ok2 = true;
211           }
212         }
213         return ok1 && ok2;
214       }
215
216       if (1<<uprim & ~Gprim.S)
217         return is_collapsible_rec(uprim+1);
218       // invariant: u' < N and u' in V(G')
219
220       for (int u = 0; u < N; ++u) if (1<<u & G.S) {
221         if ((M[u] & between[uprim]) != between[uprim])
222           continue;
223         // invariant: u' -> u preserves edges between
224         // u' and G
225
226         bool ok = true;
227         for (int vprim = 0; vprim < uprim && ok;
228             ++vprim) {
```

```
229             if (Mprim[uprim] & 1<<vprim &&
230                 !(M[u] & 1<<mapping[vprim])) {
231               ok = false;
232             }
233           }
234           if (!ok) continue;
235           // invariant: u' -> u preserves edges between
236           // u' and preceding vertices in G'.
237
238           mapping[uprim] = u;
239           if (is_collapsible_rec(uprim+1)) return true;
240         }
241         return false;
242       }
243     } R(a_T, a_G, a_Gprim, a_between, a_mapping);
244
245     return R.is_collapsible_rec(0);
246   }
247
248   // Precondition: T is a subset of G.S
249   inline unsigned neighbors (const unsigned T,
250       const Graph& G) {
251     const int N = G.M.size();
252     unsigned res = 0;
253     for (int u = 0; u < N; ++u) if (1<<u & T)
254       res |= G.M[u];
255     return res;
256   }
257
258   const int V = 12;
259   // Cycle with shortcuts C_V.
260   Graph original_G() {
261     Graph G((1<<V) - 1, vector<unsigned>(V));
262     for (int u = 0; u < V; ++u)
263       for (int s = -3; s <= 3; s += 2)
264         G.M[u] |= 1 << mod(u+s, V);
265     return G;
266   }
```

Listing 2: `non_empty_b.cpp` — Proof of Lemma 3.67.

```
1 #include "construction.hpp"
```

```
2
3  // Precondition: B, C disjoint, 0 in B
4  bool Bprim_filled(const unsigned a_C,
5      const unsigned a_B) {
6    struct RecData {
7      const Graph G;
8      const vector<unsigned>& M;
9      const unsigned C;
10     const unsigned B;
11     unsigned Bprim;
12     const int pB;
13     vector<int> B_list, Bprim_list;
14
15     RecData(const unsigned a_C, const unsigned a_B):
16         G(original_G()), M(G.M), C(a_C), B(a_B),
17         Bprim(0), pB(popcount(B)), B_list(pB),
18         Bprim_list(pB) {
19       for (int u = 0, ind = -1; u < V; ++u)
20         if (1<<u & B) {
21           ++ind;
22           B_list[ind] = u;
23         }
24     }
25
26     bool recursively_filled(int ind) {
27       if (ind == pB) {
28         // invariant: B, B', C (pairwise) disjoint
29         // invariant: edges of B and B' (inside and
30         // to C) isomorphic according to Bprim_list.
31         return is_rest_filled();
32       }
33
34       const int u = B_list[ind];
35       for (int uprim = 0; uprim < V; ++uprim) {
36         if (1<<uprim & (B|C|Bprim)) continue;
37         // invariant: uprim is "fresh"
38         if (M[uprim] & B) continue;
39         // invariant: no edges to B
40         if ((M[u]&C) != (M[uprim]&C)) continue;
41         // invariant: edges to C the same
42         bool ok = true;
```

```
43          for (int j = 0; j < ind && ok; ++j) {
44            const int v = B_list[j],
45              vprim = Bprim_list[j];
46            // a hack: '!' is used to convert to bool
47            if (!(M[v]&(1<<u)) != !(M[vprim]&(1<<uprim)))
48              ok = false;
49          }
50          if (!ok) continue;
51          // invariant: edges inside B and B' (so far)
52          // isomorphic
53          Bprim |= 1<<uprim;
54          Bprim_list[ind] = uprim;
55          if (recursively_filled(ind+1)) return true;
56          Bprim &= ~(1<<uprim);
57        }
58        return false;
59      }
60
61      // preconditions: B, Bprim, C disjoint
62      // B_list, Bprim_list, pB correctly filled
63      // B and B' isomorphic wrt each other and C
64      bool is_rest_filled() {
65        static Graph Gout(0, vector<unsigned>(V));
66        static vector<unsigned> between(V);
67        static vector<unsigned>& Mout = Gout.M;
68        static vector<int> mapping(V);
69
70        for (unsigned A = 0; A < 1<<V; A += 2) {
71          if (A & (B|C|Bprim)) continue;
72          // invariant: A, B, B', C disjoint
73          if (neighbors(A, G) & (Bprim)) continue;
74          // invariant: no edges between A and B'
75
76          const unsigned Dprim = ((1<<V)-1) &
77            ~(A|B|Bprim|C);
78          if (neighbors(Dprim, G) & (A|B)) continue;
79          // invariant: no edges between D' and A \cup B
80
81          Gout.S = A|Dprim;
82          for (int u = 0; u < V; ++u)
83            if (1<<u & A) {
```

```
84              Mout[u] = M[u] & A;
85              between[u] = M[u] & C;
86              for (int ind = 0; ind < pB; ++ind) {
87                const int v = B_list[ind],
88                  vprim = Bprim_list[ind];
89                if (M[u] & 1<<v) between[u] |= 1<<vprim;
90              }
91            } else if (1<<u & Dprim) {
92              Mout[u] = M[u] & Dprim;
93              between[u] = M[u] & C;
94              for (int ind = 0; ind < pB; ++ind) {
95                const int v = B_list[ind],
96                  vprim = Bprim_list[ind];
97                if (M[u] & 1<<vprim) between[u] |= 1<<v;
98              }
99            } else Mout[u] = between[u] = 0;
100
101         if (is_collapsible(G, Gout, between,
102             mapping)) {
103           printf("FAILURE\nA = ");
104           for (int u = 0; u < V; ++u) if (1<<u & A)
105             printf("%d ", u);
106           printf("\n(B,B') = ");
107           for (int ind = 0; ind < pB; ++ind)
108             printf("(%d, %d) ", B_list[ind],
109               Bprim_list[ind]);
110           printf("\nC = ");
111           for (int u = 0; u < V; ++u) if (1<<u & C)
112             printf("%d ", u);
113           printf("\nD' = ");
114           for (int u = 0; u < V; ++u) if (1<<u & Dprim)
115             printf("%d ", u);
116           printf("\nmapping = ");
117           for (int u = 0; u < V; ++u)
118             printf("(%d->%d) ", u, mapping[u]);
119           printf("\n");
120           exit(0);
121         }
122       }
123       return false;
124     }
```

```
125   } R(a_C, a_B);
126
127   return R.recursively_filled(0);
128 }
129
130 // Assume E(A, D) is empty.
131 // Try all partitions of C_12 into A, B, B', C, D'
132 // s.t. in the last doubling:
133 // A is doubled and then A is fixed and A' collapsed.
134 // (non-empty) B is doubled and fixed together with B'.
135 // C is fixed in both steps.
136 // D is doubled, with D collapsed and D' fixed.
137 // Objective: show that resulting A', D cannot be
138 // collapsed onto the rest.
139 int main() {
140   printf("non-empty B, V = %d\n", V);
141   // Assume w.l.o.g. that 0 is in B.
142   for (unsigned C = 0; C < 1<<V; C += 2)
143     for (unsigned B = 1; B < 1<<V; B += 2) {
144       // invariant: 0 in B
145       if (B&C || popcount(B)%2 == 1) continue;
146       // invariant: B, C disjoint
147       if (Bprim_filled(C, B)) {
148         // this should be never executed
149         printf("INTERNAL ERROR\n");
150         exit(1);
151       }
152     }
153   printf("SUCCESS\n");
154 }
```

Listing 3: `non_empty_ad.cpp` — Proof of Lemma 3.68.

```
1 #include "construction.hpp"
2
3 // Assume B is empty and |E(A, D)| = 1.
4 // Try all partitions of C_12 into A, C, D' s.t. in the
5 // last doubling:
6 // A is doubled and A' is later collapsed.
7 // C is not doubled.
8 // D is doubled and later collapsed and D' is kept.
9 // Then try all choices for the edge between A and D.
```

```
10 // Goal: Show that resulting A', D cannot be collapsed
11 // onto A, C, D'.
12 int main() {
13   printf("|E(A,D)| = 1, V = %d\n", V);
14   Graph G = original_G();
15   for (unsigned A = 0; A < 1<<V; ++A)
16     for (unsigned C = 0; C < 1<<V; ++C) {
17       if (A&C) continue;
18       // invariant: A, C disjoint
19       unsigned Dprim = ((1<<V)-1) & ~(A|C);
20       if (neighbors(Dprim, G) & A) continue;
21       // invariant: no edges between A and D'
22
23       Graph tmp_G = G, Gprim(0, vector<unsigned>());
24       vector<unsigned> between;
25       vector<int> mapping;
26
27       double_graph(A|Dprim, tmp_G, Gprim, between);
28       exchange(Dprim, tmp_G, Gprim, between);
29
30       for (int u = 0; u < V; ++u) if (1<<u & A)
31         for (int v = 0; v < V; ++v) if (1<<v & Dprim) {
32           if (u%2 == v%2) continue;
33           // invariant: u and v do not create odd cycle
34           between[u] |= 1<<v;
35           between[v] |= 1<<u;
36           if (is_collapsible(tmp_G, Gprim, between,
37               mapping)) {
38             printf("FAILURE\nA = ");
39             for (int w = 0; w < V; ++w) if (1<<w & A)
40               printf("%d ", w);
41             printf("\nC = ");
42             for (int w = 0; w < V; ++w) if (1<<w & C)
43               printf("%d ", w);
44             printf("\nDprim = ");
45             for (int w = 0; w < V; ++w)
46               if (1<<w & Dprim) printf("%d ", w);
47             printf("\nu = %d v = %d\nmapping = ",
48               u, v);
49             for (int w = 0; w < (int)mapping.size();
50                 ++w) {
```

```
51                  printf("(%d -> %d) ", w, mapping[w]);
52                }
53                printf("\n");
54                exit(0);
55              }
56              between[u] &= ~(1<<v);
57              between[v] &= ~(1<<u);
58            }
59        }
60    printf("SUCCESS\n");
61  }
```

Listing 4: `natural_collapse.cpp` — Proof of Lemma 3.72.

```
1  #include "construction.hpp"
2
3  // Assume E(A, D) is empty.
4  // Try partitioning vertices of C_12 into A, C, D' s.t.
5  // in the last doubling:
6  // A is doubled and A' is later collapsed.
7  // C is not doubled.
8  // D is doubled and later collapsed and D' is kept.
9  // Objective: Show that every time either A' or D must
10 // be naturally collapsed.
11 int main() {
12   printf("Natural collapse lemma, V = %d\n", V);
13   Graph G = original_G();
14   for (unsigned A = 0; A < 1<<V; ++A)
15     for (unsigned C = 0; C < 1<<V; ++C) {
16       if (A&C) continue;
17       // invariant: A, C disjoint
18       unsigned D = ((1<<V)-1) & ~(A|C);
19       if (neighbors(D, G) & A) continue;
20       // invariant: no edges between A and D
21
22       Graph tmp_G = G, Gprim(0, vector<unsigned>());
23       vector<unsigned> between;
24       vector<int> mapping;
25
26       double_graph(A|D, tmp_G, Gprim, between);
27       exchange(D, tmp_G, Gprim, between);
28       if (is_unnaturally_collapsible(A, tmp_G, Gprim,
```

```
29            between, mapping)) {
30          printf("FAILURE\nA = ");
31          for (int u = 0; u < V; ++u)
32            if (1<<u & A) printf("%d ", u);
33          printf("\nC = ");
34          for (int u = 0; u < V; ++u)
35            if (1<<u & C) printf("%d ", u);
36          printf("\nD = ");
37          for (int u = 0; u < V; ++u)
38            if (1<<u & D) printf("%d ", u);
39          printf("\nmapping = ");
40          for (int u = 0; u < (int)mapping.size(); ++u)
41            printf("(%d -> %d) ", u, mapping[u]);
42          printf("\n");
43          exit(0);
44        }
45      }
46    printf("SUCCESS\n");
47  }
```

Bibliography

[AM13] Per Austrin and Elchanan Mossel. Noise correlation bounds for uniform low degree functions. *Arkiv för Matematik*, 51(1):29–52, 2013.

[Arr50] Kenneth Arrow. A difficulty in the concept of social welfare. *The Journal of Political Economy*, 58(4):328–346, 1950.

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2):5:1–5:37, 2009.

[Ban65] John Banzhaf. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, 19:317–343, 1965.

[Bec75] William Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102(1):159–182, 1975.

[BFS14] Harry Buhrman, Serge Fehr, and Christian Schaffner. On the parallel repetition of multi-player games: The no-signaling case. In *9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2014)*, pages 24–35, 2014.

[BGKW88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, pages 113–131, 1988.

[BL85] Michael Ben-Or and Nathan Linial. Collective coin flipping, robust voting schemes and minima of Banzhaf values. In *Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 408–416, 1985.

[Bod07] Hans L. Bodlaender. Treewidth: Structure and algorithms. In *Structural Information and Communication Complexity: 14th International Colloquium, SIROCCO 2007, Castiglioncello, Italy, June 5-8, 2007. Proceedings*, pages 11–25, 2007.

[Bon70] Aline Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Annales de l'institut Fourier*, 20(2):335–402, 1970.

[Bor82] Christer Borell. Positivity improving operators and hypercontractivity. *Mathematische Zeitschrift*, 180(3):225–234, 1982.

[Bor85] Christer Borell. Geometric bounds on the Ornstein–Uhlenbeck velocity process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 70(1):1–13, 1985.

[Bra14] Johann Brault-Baron. Hypergraph acyclicity revisited. arXiv:1403.7076, 2014.

[BRR+09] Boaz Barak, Anup Rao, Ran Raz, Ricky Rosen, and Ronen Shaltiel. Strong parallel repetition theorem for free projection games. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 12th International Workshop, APPROX 2009, and 13th International Workshop, RANDOM 2009, Berkeley, CA, USA, August 21-23, 2009. Proceedings*, pages 352–365, 2009.

[BVY15] Mohammad Bavarian, Thomas Vidick, and Henry Yuen. Anchoring games for parallel repetition. arXiv:1509.07466, 2015.

[CCL92] Jin-yi Cai, Anne Condon, and Richard J. Lipton. On games of incomplete information. *Theoretical Computer Science*, 103(1):25–38, 1992.

[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.

[Col71] John Coleman. Control of collectivities and the power of a collectivity to act. In Bernhardt Lieberman, editor, *Social Choice*. Gordon and Breach, 1971.

[CWY15] Kai-Min Chung, Xiaodi Wu, and Henry Yuen. Parallel repetition for entangled k-player games via fast quantum search. In *30th Conference on Computational Complexity (CCC 2015)*, pages 512–536, 2015.

[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.

[Erd64] Paul Erdős. On extremal problems of graphs and generalized graphs. *Israel Journal of Mathematics*, 2(3):183–190, 1964.

[Fei91] Uriel Feige. On the success probability of the two provers in one-round proof systems. In *Structure in Complexity Theory Conference, 1991., Proceedings of the Sixth Annual*, pages 116–123, Jun 1991.

[Fei95] Uriel Feige. Error reduction by parallel repetition — the state of the art. Technical Report CS95-32, Weizmann Institute, 1995.

[FK91] Harry Furstenberg and Yitzhak Katznelson. A density version of the Hales-Jewett theorem. *Journal d'Analyse Mathématique*, 57(1):64–119, 1991.

[FRS88] Lance Fortnow, John Rompel, and Michael Sipser. On the power of multi-prover interactive protocols. In *Third Annual Structure in Complexity Theory Conference, 1988. Proceedings*, pages 156–161, 1988.

[FRS90] Lance Fortnow, John Rompel, and Michael Sipser. Errata for on the power of multi-prover interactive protocols. In *Fifth Annual Structure in Complexity Theory Conference, 1990. Proceedings*, pages 318–319, 1990.

[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.

[FV02] Uriel Feige and Oleg Verbitsky. Error reduction by parallel repetition — a negative result. *Combinatorica*, 22(4):461–478, 2002.

[GL15] Venkatesan Guruswami and Euiwoong Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 822–836, 2015.

[Gow01] W. T. Gowers. A new proof of Szemerédi's theorem. *Geometric & Functional Analysis GAFA*, 11(3):465–588, 2001.

[Gow07] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics*, 166(3):897–946, 2007.

[Gra79] Marc H. Graham. On the universal relation. Technical report, University of Toronto, 1979.

[Gro75] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995.

[Hås01] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, July 2001.

[HHC99] Chin-Wen Ho, Sun-Yuan Hsieh, and Gen-Huey Chen. Parallel decomposition of generalized series-parallel graphs. *Journal of Information Science and Engineering*, 15(3):407–417, 1999.

[HHM16] Jan Hązła, Thomas Holenstein, and Elchanan Mossel. Lower bounds on same-set inner product in correlated spaces. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*, pages 34:1–34:11, 2016.

[HHR16] Jan Hązła, Thomas Holenstein, and Anup Rao. On parallel repetition and density Hales-Jewett theorem. arXiv:1604.05757, 2016.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[Hol09] Thomas Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(8):141–172, 2009.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988.

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

[KV15] Subhash A. Khot and Nisheeth K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into ℓ_1. *J. ACM*, 62(1):8:1–8:39, 2015.

[KY15] Philip Klein and Neal E. Young. On the number of iterations for Dantzig-Wolfe optimization and packing-covering approximation algorithms. *SIAM Journal on Computing*, 44(4):1154–1172, 2015.

[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.

[LPW08] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.

[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.

[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.

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

[Mos12] Elchanan Mossel. A quantitative Arrow theorem. *Probability Theory and Related Fields*, 154(1):49–88, 2012.

[MOS13] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013.

[Nak35] Akira Nakashima. The theory of relay circuit composition. *The Journal of the Institute of Telegraph and Telephone Engineers of Japan*, 150:731–752, 1935.

[Nel73] Edward Nelson. The free Markoff field. *Journal of Functional Analysis*, 12(2):211–227, 1973.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[Ole03] Krzysztof Oleszkiewicz. On a nonsymmetric version of the Khinchine-Kahane inequality. In Evariste Giné, Christian Houdré, and David Nualart, editors, *Stochastic Inequalities and Applications*, volume 56 of *Progress in Probability*, pages 157–168. Birkhäuser Basel, 2003.

[Pal32] Raymond Paley. A remarkable series of orthogonal functions (I). *Proceedings of the London Mathematical Society*, 2(1):241–264, 1932.

[Pel95] David Peleg. On the maximum density of 0–1 matrices with no forbidden rectangles. *Discrete Mathematics*, 140(1–3):269–274, 1995.

[Pen46] Lionel Penrose. The elementary statistics of majority voting. *Journal of the Royal Statistical Society*, 109(1):53–57, 1946.

[Pol10] D. H. J. Polymath. Density Hales-Jewett and Moser numbers. In *An Irregular Mind: Szemerédi is 70*, volume 21 of *Bolyai Society Mathematical Studies*, pages 689–753. Springer Berlin Heidelberg, 2010.

[Pol12] D. H. J. Polymath. A new proof of the density Hales-Jewett theorem. *Annals of Mathematics*, 175(3):1283–1327, 2012.

[Rao11] Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM Journal on Computing*, 40(6):1871–1891, 2011.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[Raz10] Ran Raz. Parallel repetition of two prover games (invited survey). In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, June 9-12, 2010*, pages 3–6, 2010.

[Raz11] Ran Raz. A counterexample to strong parallel repetition. *SIAM Journal on Computing*, 40(3):771–777, 2011.

[Rot53] Klaus F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, s1-28(1):104–109, 1953.

[RS04] Vojtěch Rödl and Jozef Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures & Algorithms*, 25(1):1–42, 2004.

[RS06] Vojtěch Rödl and Jozef Skokan. Applications of the regularity lemma for uniform hypergraphs. *Random Structures & Algorithms*, 28(2):180–194, 2006.

[Rud87] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, Inc., 3rd edition, 1987.

[Sha37] Claude Shannon. A symbolic analysis of relay and switching circuits. Master's thesis, Massachusetts Institute of Technology, 1937.

[She38] Victor Shestakov. *Some Mathematical Methods for the Construction and Simplification of Two-Terminal Electrical Networks of Class A*. PhD thesis, Lomonosov State University, 1938.

[Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27(1):199–245, 1975.

[Tal94] Michel Talagrand. On Russo's approximate zero-one law. *The Annals of Probability*, 22(3):pp. 1576–1587, 1994.

[TV06] Terence Tao and Van H. Vu. *Additive Combinatorics*. Cambridge University Press, 2006.

[Ver95] Oleg Verbitsky. The parallel repetition conjecture for trees is true. *Electronic Colloquium on Computational Complexity (ECCC)*, 2(13), 1995.

[Ver96] Oleg Verbitsky. Towards the parallel repetition conjecture. *Theoretical Computer Science*, 157(2):277–282, 1996.

[Wal23] Joseph Walsh. A closed set of normal orthogonal functions. *American Journal of Mathematics*, 45(1):5–24, 1923.

[Wei13] Felix Weissenberger. Two-prover games for parallel repetition. Master's thesis, ETH Zurich, 2013.

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

[YÖ79] Clement T. Yu and Meral Z. Özsoyoğlu. Algorithm for tree-query membership of a distributed query. In *Proceedings — IEEE Computer Society's International Computer Software and Applications Conference*, pages 306–312, 1979.

Curriculum Vitae

Jan Hązła
Citizen of the Republic of Poland
Born on September 21, 1987 in Gniezno, Poland

Doctorate in Computer Science, *9/2011–6/2016*
ETH Zurich, Zurich, Switzerland.
Department of Computer Science,
Institute of Theoretical Computer Science.
Job title: research assistant.

Internship, *7/2010–10/2010*
Google, Zurich, Switzerland.
Job title: software engineer in test intern.

Master's Degree in Computer Science, *10/2006–7/2011*
Jagiellonian University, Kraków, Poland.
Faculty of Mathematics and Computer Science.
Studies in Mathematics and Natural Sciences.
Master thesis title: Effective implementation of Boltzmann samplers.

High School, *9/2003–6/2006*
Dąbrówka Secondary School (II LO), Gniezno, Poland.