GAUSSIAN BOUNDS FOR NOISE CORRELATION OF FUNCTIONS

ELCHANAN MOSSEL

Abstract. In this paper we derive tight bounds on the expected value of products of *low influence* functions defined on correlated probability spaces. The proofs are based on extending Fourier theory to an arbitrary number of correlated probability spaces, on a generalization of an invariance principle recently obtained with O’Donnell and Oleszkiewicz for multilinear polynomials with low influences and bounded degree and on properties of multi-dimensional Gaussian distributions.

We present two applications of the new bounds to the theory of social choice. We show that Majority is asymptotically the most predictable function among all low influence functions given a random sample of the voters. Moreover, we derive an almost tight bound in the context of Condorcet aggregation and low influence voting schemes on a large number of candidates. In particular, we show that for every low influence aggregation function, the probability that Condorcet voting on k candidates will result in a unique candidate that is preferable to all others is $k^{-1+o(1)}$. This matches the asymptotic behavior of the majority function for which the probability is $k^{-1-o(1)}$.

A number of applications in hardness of approximation in theoretical computer science were obtained using the results derived here in subsequent work by Raghuven- dra and by Austrin and Mossel.

A different type of applications involves hypergraphs and arithmetic relations in product spaces. For example, we show that if $A \subset \mathbb{Z}_m^n$ is of low influences, then the number of k tuples $(x_1, \dots, x_k) \in A^k$ satisfying $\sum_{i=1}^k x_i \in B^n \bmod m$ where $B \subset [m]$ satisfies $|B| \geq 2$ is $(1 \pm o(1))\mathbf{P}[A]^k(m^{k-1}|B|)^n$ which is the same as if A were a random set of probability $\mathbf{P}[A]$. Our results also show that, for a general set A without any restriction on the influences, there exists a set of coordinates $S \subset [n]$ with $|S| = O(1)$ such that if $C = \{x : \exists y \in A, y_{[n] \setminus S} = x_{[n] \setminus S}\}$ then the number of k -tuples $(x_1, \dots, x_k) \in C^k$ satisfying $\sum_{i=1}^k x_i \in B^n \bmod m$ is $(1 \pm o(1))\mathbf{P}[C]^k(m^{k-1}|B|)^n$.

Keywords and phrases: Invariance, discrete harmonic analysis, voting, hardness of approximation, Gaussian isoperimetric inequalities

2010 *Mathematics Subject Classification:* 60F05, 05A16

Supported by a Sloan fellowship in Mathematics, by BSF grant 2004105, NSF Career Award (DMS 054829) and by ONR award N00014-07-1-0506. Part of this work was carried out while the author was visiting IPAM, UCLA

1 Introduction

1.1 Harmonic analysis of boolean functions. This paper studies low influence functions $f : \Omega^n \rightarrow [0, 1]$, where (Ω^n, μ^n) is a product probability space and where the *influence of the i th coordinate on f* , denoted by $\text{Inf}_i(f)$ is defined by

$$\text{Inf}_i(f) = \mathbf{E}[\text{Var}[f(X_1, \dots, X_n) \mid X_j, 1 \leq j \leq n, j \neq i]], \quad (1)$$

where for any set $S \subset [n]$ the conditional variance $\text{Var}[f(X_1, \dots, X_n) \mid X_i, i \in S]$ is defined via

$$\begin{aligned} \text{Var}[f(X_1, \dots, X_n) \mid X_i, i \in S] \\ = \mathbf{E}[(f(X_1, \dots, X_n) - \mathbf{E}[f(X_1, \dots, X_n) \mid X_i, i \in S])^2 \mid X_i, i \in S]. \end{aligned}$$

The study of low influence functions is motivated by applications from the theory of social choice in mathematical economics, by applications in the theory of hardness of approximation in theoretical computer science and by problems in additive number theory. We refer the reader to some recent papers [KhKMO1,2], [MOO1,2], [DMR], [ST], [Gr] for motivation and general background. The main theorems established here provide tight bounds on the expected value of the product of functions defined on correlated probability spaces. These in turn imply some new results in the theory of social choice and in the theory of hypergraphs. Application to hardness of approximation in computer science were derived in subsequent work in [AM1] and [R].

In our main result we consider a probability measure $\mathbf{P}$ defined on a space $\prod_{i=1}^k \Omega^{(i)}$. Letting $f_i : (\Omega^{(i)})^n \rightarrow [0, 1]$, $1 \leq i \leq k$, be a collection of low influence functions we derive tight bounds on $\mathbf{E}[f_1 \dots f_k]$ in terms of $\mathbf{E}[f_1], \dots, \mathbf{E}[f_k]$ and a measure of correlation between the spaces $\Omega^{(1)}, \dots, \Omega^{(k)}$. The bounds are expressed in terms of extremal probabilities in Gaussian space, that can be calculated in the case $k = 2$. When $k \geq 2$ and $\mathbf{P}$ is a *pairwise independent distribution* our bounds show that $\mathbf{E}[f_1 \dots f_k]$ is close to $\prod_{i=1}^k \mathbf{E}[f_i]$. We also apply a simple recursive argument in order to obtain results for general functions not necessarily of low influences. The results show that the bounds for low influence functions hold for general functions after the functions have been “modified” in a bounded number of coordinates. The rest of the introduction is devoted to various applications followed by statements of the main technical results.

1.2 Prediction of low influence voting. Suppose n voters are to make a binary decision. Assume that the outcome of the vote is determined by a *social choice* function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, so that the outcome of the vote is $f(x_1, \dots, x_n)$ where $x_i \in \{-1, 1\}$ is the vote of voter i . We assume that the votes are independent, each ± 1 with probability $1/2$. It is natural to assume that the function f satisfies $f(-x) = -f(x)$, i.e. it does not discriminate between the two candidates. Note that this implies that $\mathbf{E}[f] = 0$ under the uniform distribution. A natural way to try and predict the outcome of the vote is to sample a subset of the voters, by sampling each voter independently with probability ρ . Conditioned on a vector X of votes the distribution of Y , the sampled votes, is i.i.d. where $Y_i = X_i$ with probability ρ and $Y_i = *$ (for unknown) otherwise.

Conditioned on $Y = y$, the vector of sampled votes, the optimal prediction of the outcome of the vote is given by $\text{sgn}((Tf)(y))$ where

$$(Tf)(y) = \mathbf{E}[f(X) \mid Y = y]. \quad (2)$$

This implies that the probability of correct prediction (also called predictability) is given by

$$\mathbf{P}[f = \text{sgn}(Tf)] = \frac{1}{2}(1 + \mathbf{E}[f \text{sgn}(Tf)]).$$

For example, when $f(x) = x_1$ is the dictator function, we have $\mathbf{E}[f \text{sgn}(Tf)] = \rho$ corresponding to the trivial fact that the outcome of the election is known when voter 1 is sampled and are ± 1 with probability $1/2$ otherwise. The notion of predictability is natural in statistical contexts. It was also studied in a more combinatorial context in [O].

In the first application presented here we show that

Theorem 1.1 (Majority is most predictable). *Let $0 \leq \rho \leq 1$ and $\epsilon > 0$ be given. Then there exists a $\tau > 0$ such that if $f : \{-1, 1\}^n \rightarrow [-1, 1]$ satisfies $\mathbf{E}[f] = 0$ and $\text{Inf}_i(f) \leq \tau$ for all i , then*

$$\mathbf{E}[f \text{sgn}(Tf)] \leq \frac{2}{\pi} \arcsin \sqrt{\rho} + \epsilon, \quad (3)$$

where T is defined in (2).

Moreover, it follows from the central limit theorem (see section 7.5; a version of this calculation also appears in [O]) that if $\text{Maj}_n(x_1, \dots, x_n) = \text{sgn}(\sum_{i=1}^n x_i)$, then

$$\lim_{n \rightarrow \infty} \mathbf{E}[\text{Maj}_n \text{sgn}(T \text{Maj}_n)] = \frac{2}{\pi} \arcsin \sqrt{\rho}.$$

REMARK 1.2. Note that Theorem 1.1 proves a weaker statement than showing that Majority is the most predictable function. The statement only asserts that if a function has low enough influences then its predictability cannot be more than ϵ larger than the asymptotic predictability value achieved by the majority function when the number of voters $n \rightarrow \infty$. This slightly inaccurate title of the theorem is in line with previous results such as the “Majority is stablest theorem” (see below). Similar language may be used later when informally discussing statements of various theorems.

REMARK 1.3. One may wonder if for a finite n , among *all* functions $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ with $\mathbf{E}[f] = 0$, majority is the most predictable function. Note that the predictability of the dictator function $f(x) = x_1$ is given by ρ , and $\frac{2}{\pi} \arcsin \sqrt{\rho} > \rho$ for $\rho \rightarrow 0$. Therefore when ρ is small and n is large the majority function is more predictable than the dictator function. However, note that when $\rho \rightarrow 1$ we have $\rho > \frac{2}{\pi} \arcsin \sqrt{\rho}$ and therefore for values of ρ close to 1 and large n the dictator function is more predictable than the majority function.

We note that the bound obtained in Theorem 1.1 is reminiscent of the “Majority is stablest theorem” [MOO1,2] as both involve the arcsin function. However, the two theorems are quite different. The “Majority is stablest theorem” asserts that under the same condition as in Theorem 1.1 it holds that

$$\mathbf{E}[f(X)f(Y)] \leq \frac{2}{\pi} \arcsin \rho + \epsilon,$$

where $(X_i, Y_i) \in \{-1, 1\}^2$ are i.i.d. with $\mathbf{E}[X_i] = \mathbf{E}[Y_i] = 0$ and $\mathbf{E}[X_i Y_i] = \rho$. Thus “Majority is stablest” considers two correlated voting vectors, while “Majority is most predictable” considers a sample of one voting vector. In fact, both results follow from the more general invariance principle presented here. We note a further difference between stability and predictability: It is well known that in the context of “Majority is stablest”, for all $0 < \rho < 1$, among all boolean functions with $\mathbf{E}[f] = 0$, the maximum of $\mathbf{E}[f(x)f(y)]$ is obtained for dictator functions of the form $f(x) = x_i$. As discussed above, for ρ close to 0 and large n , the dictator is less predictable than the majority function.

We also note that the “Ain’t over until it’s over” theorem [MOO1,2] provides a bound under the same conditions on

$$P[Tf > 1 - \delta],$$

for small δ . However, this bound is not tight and does not imply Theorem 1.1. Similarly, Theorem 1.1 does not imply the “Ain’t over until it’s over” theorem. The bounds in “Ain’t over until it’s over” were derived using invariance of Tf while the bound (3) requires the joint invariance of f and Tf .

1.3 Condorcet paradoxes. Suppose n voters rank k candidates. It is assumed that each voter i has a linear order $\sigma_i \in S(k)$ on the candidates. In *Condorcet voting*, the rankings are aggregated by deciding for each pair of candidates which one is superior among the n voters.

More formally, the aggregation results in a tournament G_k on the set $[k]$. Recall that G_k is a *tournament* on $[k]$ if it is a directed graph on the vertex set $[k]$ such that for all $a, b \in [k]$ either $(a > b) \in G_k$ or $(b > a) \in G_k$. Given individual rankings $(\sigma_i)_{i=1}^n$ the tournament G_k is defined as follows.

Let $x^{a>b}(i) = 1$, if $\sigma_i(a) > \sigma_i(b)$, and $x^{a>b}(i) = -1$ if $\sigma_i(a) < \sigma_i(b)$. Note that $x^{b>a} = -x^{a>b}$.

The binary decision between each pair of candidates is performed via a anti-symmetric function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ so that $f(-x) = -f(x)$ for all $x \in \{-1, 1\}^n$. The tournament $G_k = G_k(\sigma; f)$ is then defined by letting $(a > b) \in G_k$ if and only if $f(x^{a>b}) = 1$.

Note that there are $2^{\binom{k}{2}}$ tournaments while there are only $k! = 2^{\Theta(k \log k)}$ linear rankings. For the purposes of social choice, some tournaments make more sense than others.

DEFINITION 1.4. We say that a tournament G_k is linear if it is acyclic. We will write $\text{Acyc}(G_k)$ for the logical statement that G_k is acyclic. Non-linear tournaments are often referred to as non-rational in economics as they represent an order where there are 3 candidates a, b and c such that a is preferred to b , b is preferred to c and c is preferred to a .

We say that the tournament G_k is a unique max tournament if there is a candidate $a \in [k]$ such that for all $b \neq a$ it holds that $(a > b) \in G_k$. We write $\text{UniqMax}(G_k)$ for the logical statement that G_k has a unique max. Note that the unique max property is weaker than linearity. It corresponds to the fact that there is a candidate that dominates all other candidates.

Following [Ka1,2], we consider the probability distribution over n voters, where the voters have independent preferences and each one chooses a ranking uniformly at random among all $k!$ orderings. Note that the marginal distributions on vectors $x^{a>b}$ is the uniform distribution over $\{-1, 1\}^n$ and that if $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is anti-symmetric then $\mathbf{E}[f] = 0$.

The case that is now understood is $k = 3$. Note that in this case G_3 is a unique max if and only if it is linear. Kalai [Ka1] studied the *probability* of a rational outcome given that the n voters vote independently and at random from the six possible rational rankings. He showed that the probability of a rational outcome in this case may be expressed as $\frac{3}{4}(1 + \mathbf{E}[fTf])$ where T is the *Bonami–Beckner* operator with parameter $\rho = 1/3$. The Bonami–Beckner operator may be defined as follows. Let $(X_i, Y_i) \in \{-1, 1\}^2$ be i.i.d. with $\mathbf{E}[X_i] = \mathbf{E}[Y_i] = 0$ and $\mathbf{E}[X_i Y_i] = \rho$ for $1 \leq i \leq n$. For $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, and $x \in \{-1, 1\}^n$, the Bonami–Beckner operator T applied to f is defined via $(Tf)(x_1, \dots, x_n) = \mathbf{E}[f(Y_1, \dots, Y_n) \mid X_1 = x_1, \dots, X_n = x_n]$.

It is natural to ask which function f with small influences is most likely to produce a rational outcome. Instead of considering small influences, Kalai considered the essentially stronger assumption that f is monotone and “transitive-symmetric”; i.e. that for all $1 \leq i < j \leq n$ there exists a permutation σ on $[n]$ with $\sigma(i) = j$ such that $f(x_1, \dots, x_n) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)})$ for all $(x_1, \dots, x_n)$. Kalai conjectured that, as $n \rightarrow \infty$, the maximum of $\frac{3}{4}(1 + \mathbf{E}[fTf])$ among all transitive-symmetric functions approaches the same limit as $\lim_{n \rightarrow \infty} \frac{3}{4}(1 + \mathbf{E}[\text{Maj}_n T \text{Maj}_n])$. This was proven using the “Majority is stablest theorem”, [MOO1,2]. Here we obtain similar results for any value of k . Our result is not tight, but almost tight. More specifically we show

Theorem 1.5 (Majority is best for Condorcet). *Consider Condorcet voting on k candidates. Then for all $\epsilon > 0$ there exists $\tau = \tau(k, \epsilon) > 0$ such that if $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is anti-symmetric and $\text{Inf}_i(f) \leq \tau$ for all i , then*

$$\mathbf{P}[\text{UniqMax}(G_k(\sigma; f))] \leq k^{-1+o_k(1)} + \epsilon. \quad (4)$$

Moreover for $f = \text{Maj}_n$ we have $\text{Inf}_i(f) = O(n^{-1/2})$ and it holds that

$$\mathbf{P}[\text{UniqMax}(G_k(\sigma; f))] \geq k^{-1-o_k(1)} - o_n(1). \quad (5)$$

Interestingly, we are not able to derive similar results for Acyc. We do calculate the probability that Acyc holds for majority.

PROPOSITION 1.6. *We have*

$$\lim_{n \rightarrow \infty} \mathbf{P}[\text{Acyc}(G_k(\sigma; \text{Maj}_n))] = \exp(-\Theta(k^{5/3})). \quad (6)$$

We note that results in economics [Be] have shown that for a majority vote the probability that the outcome will contain a Hamiltonian cycle when the number of voters goes to infinity is $1 - o_k(1)$.

1.4 Hypergraph and additive applications. Here we discuss some applications concerning hypergraph problems. We let Ω be a finite set equipped with the uniform probability measure denoted $\mathbf{P}$. We let $R \subset \Omega^k$ denote a k -wise relation. For sets $A_1, \dots, A_k \subset \Omega^n$ we will be interested in the number of k -tuples

$x^1 \in A_1, \dots, x^k \in A_k$ satisfying the relation R in all coordinates, i.e. $(x_i^1, \dots, x_i^k) \in R$ for all i . Assume below that R satisfies the following two properties:

- For all $a \in \Omega$ and all $1 \leq j \leq k$ it holds that $\mathbf{P}[x^i = a | (x^1, \dots, x^k) \in R(x)] = |\Omega|^{-1}$. (This assumption is actually not needed for the general statement – we state it for simplicity only).
- The relation R is *connected*. This means that for all $x, y \in R$ there exists a path $x = y(0), y(1), \dots, y(r) = y$ in R such that $y(i)$ and $y(i+1)$ differ in one coordinate only.

We will say that the relation $R \subset \Omega^k$ is *pairwise smooth* if for all $i, j \in [k]$ and $a, b \in \Omega$ it holds that

$$\begin{aligned} \mathbf{P}[x^i = a, x^j = b | (x^1, \dots, x^k) \in R] \\ = \mathbf{P}[x^i = a | (x^1, \dots, x^k) \in R] \mathbf{P}[x^j = b | (x^1, \dots, x^k) \in R]. \end{aligned}$$

As a concrete example, consider the case where $\Omega = \mathbb{Z}_m$ and R consists of all k -tuples satisfying $\sum_{i=1}^k x_i \in B \bmod m$ where $B \subset \mathbb{Z}_m$. When $k \geq 2$ we have $\mathbf{P}[x_i = a | R] = m^{-1}$ for all i and a . When $k \geq 3$, we have pairwise smoothness. The connectivity condition holds whenever $|B| > 1$.

For a set $A \subset \mathbb{Z}_m^n$ and $S \subset [n]$ we define

$$\begin{aligned} \overline{A}_S &= \{y : \exists x \in A, x_{[n] \setminus S} = y_{[n] \setminus S}\}, \\ \underline{A}_S &= \{y : \forall x \text{ s.t. } x_{[n] \setminus S} = y_{[n] \setminus S}, \text{ it holds that } x \in A\}. \end{aligned}$$

Our main result in the context of hypergraphs is the following.

Theorem 1.7. *Let R be a connected relation on Ω^k . Then there exist two continuous functions $\underline{\Gamma} : (0, 1)^k \rightarrow (0, 1)$ and $\overline{\Gamma} : (0, 1)^k \rightarrow (0, 1)$ such that for every $\epsilon > 0$ there exists a $\tau > 0$ such that if $A_1, \dots, A_k \subset \Omega^n$ are sets with $\text{Inf}_i(A_j) \leq \tau$ for all i and j then*

$$\begin{aligned} (\underline{\Gamma}(\mathbf{P}[A_1], \dots, \mathbf{P}[A_k]) - \epsilon) \mathbf{P}[R^n] &\leq \mathbf{P}[R^n \cap (A_1, \dots, A_k)] \\ &\leq (\overline{\Gamma}(\mathbf{P}[A_1], \dots, \mathbf{P}[A_k]) + \epsilon) \mathbf{P}[R^n]. \end{aligned}$$

If R is pairwise smooth, then,

$$|\mathbf{P}[R^n \cap (A_1, \dots, A_k)] - \mathbf{P}[R^n] \mathbf{P}[A_1] \dots \mathbf{P}[A_k]| \leq \epsilon \mathbf{P}[R^n].$$

Moreover, one can take $\tau = \epsilon^{O(m^{k+1} \log(1/\epsilon)/\epsilon)}$.

For general sets $A_1, \dots, A_k$, not necessarily of low influences, there exists a set S of coordinates such that $|S| \leq O(1/\tau)$ and the statements above hold for $\underline{A}_1^S, \dots, \underline{A}_k^S$ and for $\overline{A}_1^S, \dots, \overline{A}_k^S$.

1.5 Correlated spaces. A central concept that is extensively studied and repeatedly used in the paper is that of correlated probability spaces. The notion of correlation between two probability spaces used here is the same as the “maximum correlation coefficient” introduced by Hirschfeld and Gebelein [G]. We will later show how to relate correlated spaces to noise operators.

DEFINITION 1.8. Given a probability measure $\mathbf{P}$ defined on $\prod_{i=1}^k \Omega^{(i)}$, we say that $\Omega^{(1)}, \dots, \Omega^{(k)}$ are correlated spaces. For $A \subset \Omega^{(i)}$ we let

$$\mathbf{P}[A] = \mathbf{P}\left[(\omega_1, \dots, \omega_k) \in \prod_{j=1}^k \Omega^{(j)} : \omega_i \in A\right],$$

and similarly $\mathbf{E}[f]$ for $f : \Omega^{(i)} \rightarrow \mathbb{R}$. We will abuse notation by writing $\mathbf{P}[A]$ for $\mathbf{P}^n[A]$, for $A \subset (\prod_{i=1}^k \Omega^{(i)})^n$ or $A \subset (\Omega^{(i)})^n$, and similarly for $\mathbf{E}$.

DEFINITION 1.9. Given two linear subspaces A and B of $L^2(\mathbf{P})$ we define the correlation between A and B by

$$\rho(A, B; \mathbf{P}) = \rho(A, B) = \sup \left\{ \mathbf{Cov}[f, g] : f \in A, g \in B, \mathbf{Var}[f] = \mathbf{Var}[g] = 1 \right\}. \quad (7)$$

Let $\Omega = (\Omega^{(1)} \times \Omega^{(2)}, \mathbf{P})$. We define the correlation $\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P})$ by letting

$$\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P}) = \rho(L^2(\Omega^{(1)}, \mathbf{P}), L^2(\Omega^{(2)}, \mathbf{P}); \mathbf{P}). \quad (8)$$

More generally, let $\Omega = (\prod_{i=1}^k \Omega^{(i)}, \mathbf{P})$ and for a subset $S \subset [k]$, write $\Omega^{(S)} = \prod_{i \in S} \Omega^{(i)}$. The correlation vector $\underline{\rho}(\Omega^{(1)}, \dots, \Omega^{(k)}; \mathbf{P})$ is a length $k-1$ vector whose i 'th coordinate is given by

$$\underline{\rho}(i) = \rho\left(\prod_{j=1}^i \Omega^{(j)}, \prod_{j=i+1}^k \Omega^{(j)}; \mathbf{P}\right),$$

for $1 \leq i \leq k-1$. The correlation $\rho(\Omega^{(1)}, \dots, \Omega^{(k)}; \mathbf{P})$ is defined by letting

$$\rho(\Omega^{(1)}, \dots, \Omega^{(k)}; \mathbf{P}) = \max_{1 \leq i \leq k} \rho\left(\prod_{j=1}^{i-1} \Omega^{(j)} \times \prod_{j=i+1}^k \Omega^{(j)}, \Omega^{(i)}; \mathbf{P}\right). \quad (9)$$

When the probability measure $\mathbf{P}$ will be clear from the context we will write $\rho(\Omega^{(1)}, \dots, \Omega^{(k)})$ for $\rho(\Omega^{(1)}, \dots, \Omega^{(k)}; \mathbf{P})$, etc.

REMARK 1.10. It is easy to see that $\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P})$ is the *second singular value* of the conditional expectation operator mapping $f \in L^2(\Omega^{(2)}, \mathbf{P})$ to $g(x) = \mathbf{E}[f(Y) \mid X = x] \in L^2(\Omega^{(1)}, \mathbf{P})$. Thus $\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P})$ is the second singular value of the matrix corresponding to the operator T with respect to orthonormal basis of $L^2(\Omega^{(1)}, \mathbf{P})$ and $L^2(\Omega^{(2)}, \mathbf{P})$.

DEFINITION 1.11. Given $(\prod_{i=1}^k \Omega^{(i)}, \mathbf{P})$, we say that $\Omega^{(1)}, \dots, \Omega^{(k)}$ are r -wise independent if for all $S \subset [k]$ with $|S| \leq r$ and for all $\prod_{i \in S} A_i \subset \prod_{i \in S} \Omega^{(i)}$ it holds that

$$\mathbf{P}\left[\prod_{i \in S} A_i\right] = \prod_{i \in S} \mathbf{P}[A_i].$$

The notion of r -wise independence is central in computer science and discrete mathematics, in particular in the context of randomized algorithms and computational complexity.

1.6 Gaussian stability. Our main result states bounds in terms of Gaussian stability measures which we discuss next. Let γ be the one-dimensional Gaussian measure.

DEFINITION 1.12. Given $\mu \in [0, 1]$, define $\chi_\mu : \mathbb{R} \rightarrow \{0, 1\}$ to be the indicator function of the interval $(-\infty, t]$, where t is chosen so that $\mathbf{E}_\gamma[\chi_\mu] = \mu$. Explicitly, $t = \Phi^{-1}(\mu)$, where Φ denotes the distribution function of a standard Gaussian. Furthermore, define

$$\begin{aligned}\bar{\Gamma}_\rho(\mu, \nu) &= \mathbf{P}[X \leq \Phi^{-1}(\mu), Y \leq \Phi^{-1}(\nu)], \\ \underline{\Gamma}_\rho(\mu, \nu) &= \mathbf{P}[X \leq \Phi^{-1}(\mu), Y \geq \Phi^{-1}(1 - \nu)],\end{aligned}$$

where (X, Y) is a two-dimensional Gaussian vector with covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

Given $(\rho_1, \dots, \rho_{k-1}) \in [0, 1]^{k-1}$ and $(\mu_1, \dots, \mu_k) \in [0, 1]^k$ for $k \geq 3$ we define by induction

$$\bar{\Gamma}_{\rho_1, \dots, \rho_{k-1}}(\mu_1, \dots, \mu_k) = \bar{\Gamma}_{\rho_1}(\mu_1, \bar{\Gamma}_{\rho_2, \dots, \rho_{k-1}}(\mu_2, \dots, \mu_k)),$$

and similarly $\underline{\Gamma}(\cdot)$.

1.7 Statements of main results. We now state our main results. We state the results both for low influence functions and for general functions. For the latter it is useful to define the following notions:

DEFINITION 1.13. Let $f : \Omega^n \rightarrow \mathbb{R}$ and $S \subset [n]$. We define

$$\bar{f}^S(x) = \sup(f(y) : y_{[n] \setminus S} = x_{[n] \setminus S}), \quad \underline{f}^S(x) = \inf(f(y) : y_{[n] \setminus S} = x_{[n] \setminus S}).$$

Theorem 1.14. Let $(\prod_{j=1}^k \Omega_i^{(j)}, \mathbf{P}_i)$, $1 \leq i \leq n$, be a sequence of finite probability spaces such that for all $1 \leq i \leq n$ the minimum probability of any atom in $\prod_{j=1}^k \Omega_i^{(j)}$ is at least α . Assume furthermore that there exists $\underline{\rho} \in [0, 1]^{k-1}$ and $0 \leq \rho < 1$ such that

$$\begin{aligned}\rho(\Omega_i^{(1)}, \dots, \Omega_i^{(k)}; \mathbf{P}_i) &\leq \rho, \\ \rho(\Omega_i^{\{1, \dots, j\}}, \Omega_i^{\{j+1, \dots, k\}}; \mathbf{P}_i) &\leq \underline{\rho}(j),\end{aligned}\tag{10}$$

for all i, j . Then for all $\epsilon > 0$ there exists $\tau > 0$ such that if

$$f_j : \prod_{i=1}^n \Omega_i^{(j)} \rightarrow [0, 1],$$

for $1 \leq j \leq k$, satisfy

$$\max_{i,j} (\text{Inf}_i(f_j)) \leq \tau,\tag{11}$$

then

$$\underline{\Gamma}_{\underline{\rho}}(\mathbf{E}[f_1], \dots, \mathbf{E}[f_k]) - \epsilon \leq \mathbf{E}\left[\prod_{j=1}^k f_j\right] \leq \bar{\Gamma}_{\underline{\rho}}(\mathbf{E}[f_1], \dots, \mathbf{E}[f_k]) + \epsilon.\tag{12}$$

If instead of (10) we assume that

$$\rho(\Omega_i^{(j)}, \Omega_i^{(j')}; \mathbf{P}_i) = 0,\tag{13}$$

for all $i, j \neq j'$, then

$$\prod_{j=1}^k \mathbf{E}[f_j] - \epsilon \leq \mathbf{E}\left[\prod_{j=1}^k f_j\right] \leq \prod_{j=1}^k \mathbf{E}[f_j] + \epsilon.\tag{14}$$

One may take

$$\tau = \epsilon^{O\left(\frac{\log(1/\epsilon) \log(1/\alpha)}{(1-\rho)\epsilon}\right)}.$$

A truncation argument allows one to relax the conditions on the influences.

PROPOSITION 1.15. *For statement (12) to hold in the case where $k = 2$ it suffices to require that*

$$\max_i (\min(\text{Inf}_i(f_1), \text{Inf}_i(f_2))) \leq \tau \quad (15)$$

instead of (11).

In the case where for each i the spaces $\Omega_i^{(1)}, \dots, \Omega_i^{(k)}$ are s -wise independent, for statement (14) to hold it suffices to require that for all i

$$|\{j : \text{Inf}_i(f_j) > \tau\}| \leq s. \quad (16)$$

An easy recursive argument allows one to conclude the following result that does not require low influences (11).

PROPOSITION 1.16. *Consider the setting of Theorem 1.14 without the assumptions on low influences (11).*

Assuming (10), there exists a set S of size $O(1/\tau)$ such that the functions $\bar{f}_j^S$ satisfy

$$\mathbf{E} \left[\prod_{j=1}^k \bar{f}_j^S \right] \geq \bar{\Gamma}_\rho(\mathbf{E}[\bar{f}_1^S], \dots, \mathbf{E}[\bar{f}_k^S]) - \epsilon \geq \bar{\Gamma}_\rho(\mathbf{E}[f_1], \dots, \mathbf{E}[f_k]) - \epsilon,$$

and the functions $\underline{f}_j^S$ satisfy

$$\mathbf{E} \left[\prod_{j=1}^k \underline{f}_j^S \right] \leq \underline{\Gamma}_\rho(\mathbf{E}[\underline{f}_1^S], \dots, \mathbf{E}[\underline{f}_k^S]) - \epsilon \leq \underline{\Gamma}_\rho(\mathbf{E}[f_1], \dots, \mathbf{E}[f_k]) - \epsilon.$$

Assuming (13), we have

$$\mathbf{E} \left[\prod_{j=1}^k \bar{f}_j^S \right] \geq \prod_{j=1}^k \mathbf{E}[\bar{f}_j^S] - \epsilon \geq \prod_{j=1}^k \mathbf{E}[f_j] - \epsilon,$$

and similarly for $\underline{f}$.

1.8 Road map. Let us review some of the main techniques we use in this paper.

- We develop a Fourier theory on correlated spaces in section 2. Previous work considered Fourier theory on one product space and reversible operators with respect to that space [DMR]. Our results here allow us to study non-reversible operators which in turn allows us to study products of k correlated spaces. An important fact we prove that is used repeatedly is that general noise operators respect “Efron–Stein” decomposition. This fact in particular allows us to “truncate” functions to their low-degree parts when considering the expected value of the product of functions on correlated spaces.
- In order to derive an invariance principle we need to extend the approach of [Ro], [MOO1,2] to prove the joint invariance of a number of multilinear polynomials. The proof of the extension appears in sections 3 and 4. The

proof follows the same main steps as in [Ro], [MOO1,2], i.e. the Lindeberg strategy for proving the CLT [L] where invariance is established by switching one variable at a time.

- In the Gaussian realm, we need to extend Borell's isoperimetric result [Bor] both in the case of two collections of Gaussians and in the case of $k > 2$ collections. This is done in section 5.
- The proof of the main result, Theorem 1.14 follows in section 6. The proof of the extensions given in Proposition 1.15 uses a truncation argument for which s -wise independence plays a crucial role. The proof of Proposition 1.16 is based on a simple recursive argument.
- In section 7 we apply the noise bounds in order to derive the social choice results. Some calculations with the majority function in the social choice setting, in particular showing the tightness of Theorems 1.1 and 1.5 are given in section 7.5. We conclude by discussing the applications to hypergraphs in section 8.

1.9 Subsequent work and applications in computer science. Subsequent to posting a draft of this paper on the arXiv, two applications of our results to hardness of approximation in computer science were established. Both results are in the context of the Unique Games conjecture in computational complexity [Kh]. Furthermore, both results consider an important problem in computer science, that is, the problem of solving *constraint satisfaction problems (CSP)*.

Given a predicate $P : [q]^k \rightarrow \{0, 1\}$, where $[q] = \{1, \dots, q\}$ for some integer q , we define $\text{MAX CSP}(P)$ to be the algorithmic problem where we are given a set of variables $x_1, \dots, x_n$ taking values in $[q]$ and a set of constraints of the form $P(l_1, \dots, l_k)$, where each $l_i = x_j + a$, where x_j is one of the variables and $a \in [q]$ is a constant (addition is mod q). More generally, in the problem of $\text{MAX } k\text{-CSP}_q$ we are given a set of constraints each involving k of the variables $x_1, \dots, x_n$. The most well-studied case is the case of $q = 2$ denoted $\text{MAX } k\text{-CSP}$.

The objective is to find an assignment to the variables satisfying as many of the constraints as possible. The problem of $\text{MAX } k\text{-CSP}_q$ is NP-hard for any $k \geq 2$, $q \geq 2$, and as a consequence, a large body of research is devoted to studying how well the problem can be approximated. We say that a (randomized) algorithm has *approximation ratio* α if, for all instances, the algorithm is guaranteed to find an assignment which (in expectation) satisfies at least $\alpha \cdot \text{Opt}$ of the constraints, where Opt is the maximum number of simultaneously satisfied constraints, over any assignment.

The results of [AM1] (see also [AM2]) show that assuming the Unique Games conjecture, for any predicate P for which there exists a pairwise independent distribution over $[q]^k$ with uniform marginals, whose support is contained in $P^{-1}(1)$, is approximation resilient. In other words, there is no polynomial time algorithm which achieves a better approximation factor than assigning the variables at random. This result implies in turn that for general $k \geq 3$ and $q \geq 2$, the $\text{MAX } k\text{-CSP}_q$ problem is UG-hard to approximate within $\mathcal{O}(kq^2)/q^k + \epsilon$. Moreover, for the special case of

$q = 2$, i.e. boolean variables, it gives hardness of $(k + \mathcal{O}(k^{0.525}))/2^k + \epsilon$, improving upon the best previous bound [ST] of $2k/2^k + \epsilon$ by essentially a factor 2. Finally, again for $q = 2$, assuming that the famous Hadamard conjecture is true, the results are further improved, and the $\mathcal{O}(k^{0.525})$ term can be replaced by the constant 4.

These results should be compared to prior work by Samorodnitsky and Trevisan [ST] who using the Gowers norm, proved that the MAX k -CSP problem has a hardness factor of $2^{\lceil \log_2 k+1 \rceil}/2^k$, which is $(k+1)/2^k$ for $k = 2^r - 1$, but can be as large as $2k/2^k$ for general k .

From the quantitative point of view [AM2] gives stronger hardness than [ST] for MAX k -CSP $_q$, even in the already thoroughly explored $q = 2$ case. These improvements may seem very small, being an improvement only by a multiplicative factor 2. However, it is well known that it is impossible to get non-approximability results which are better than $(k+1)/2^k$, and thus, in this respect, the hardness of $(k+4)/2^k$ assuming the Hadamard conjecture is in fact optimal to within a very small *additive* factor. Also, the results of [AM2] give approximation resistance of MAX CSP(P) for a much larger variety of predicates (any P containing a balanced pairwise independent distribution).

From a qualitative point of view, the analysis of [AM2] is very direct. Furthermore, it is general enough to accommodate any domain $[q]$ with virtually no extra effort. Also, their proof using the main result of the current paper, i.e. bounds on expectations of products under certain types of correlation, putting it in the same general framework as many other UGC-based hardness results, in particular those for 2-CSPs.

In a second beautiful result by Raghavendra [R] the results of the current paper were used to obtain very general hardness results for MAX CSP(P). In [R] it is shown that for every predicate P and for every approximation factor which is smaller than the UG-hardness of the problem, there exists a polynomial time algorithm which achieves this approximation ratio. Thus for every P the UG-hardness of MAX CSP(P) is sharp. The proof of the results uses the results obtained here in order to define and analyze the reduction from UG given the integrality gap of the corresponding convex optimization problem. We note that for most predicates the UG hardness of MAX CSP(P) is unknown and therefore the results of [R] complement those of [AM1].

1.10 Acknowledgments. I would like to thank Noam Nisan for suggesting that generalization of the invariance principle should be useful in the social choice context and Gil Kalai for pointing out some helpful references. I would like to thank Terence Tao for helpful discussions and references on additive number theory and Szemerédi regularity. Thanks to Per Austrin, Nick Crawford, Marcus Issacson and Ryan O'Donnell for a careful reading of a draft of this paper. Finally, many thanks to an anonymous referee for many helpful comments.

2 Correlated Spaces and Noise

In this section we define and study the notion of correlated spaces and noise operators in a general setting.

2.1 Correlated probability spaces and noise operators. We begin by defining noise operators and giving some basic examples.

DEFINITION 2.1. *Let $(\Omega^{(1)} \times \Omega^{(2)}, \mathbf{P})$ be two correlated spaces. The Markov operator associated with $(\Omega^{(1)}, \Omega^{(2)})$ is the operator mapping $f \in L^p(\Omega^{(2)}, \mathbf{P})$ to $Tf \in L^p(\Omega^{(1)}, \mathbf{P})$ by*

$$(Tf)(x) = \mathbf{E}[f(Y) \mid X = x],$$

for $x \in \Omega^{(1)}$ and where $(X, Y) \in \Omega^{(1)} \times \Omega^{(2)}$ is distributed according to $\mathbf{P}$.

EXAMPLE 2.2. In order to define the *Bonami–Beckner* operator $T = T_\rho$ on a space (Ω, μ) , consider the space $(\Omega \times \Omega, \nu)$ where $\nu(x, y) = (1 - \rho)\mu(x)\mu(y) + \rho\delta(x = y)\mu(x)$, where $\delta(x = y)$ is the function on $\Omega \times \Omega$ which takes the value 1 when $x = y$, and 0 otherwise. In this case, the operator T satisfies

$$(Tf)(x) = \mathbf{E}[f(Y) \mid X = x], \quad (17)$$

where the conditional distribution of Y given $X = x$ is $\rho\delta_x + (1 - \rho)\mu$, where δ_x is the delta measure on x .

REMARK 2.3. The construction above may be generalized as follows. Given any Markov chain on Ω that is reversible with respect to μ , we may look at the measure ν on $\Omega \times \Omega$ defined by the Markov chain. In this case T is the Markov operator determined by the chain. The same construction applies under the weaker condition that T has μ as its stationary distribution.

It is straightforward to verify that

PROPOSITION 2.4. *Suppose that for each $1 \leq i \leq n$, $(\Omega_i^{(1)} \times \Omega_i^{(2)}, \mu_i)$ are correlated spaces and T_i is the Markov operator associated with $\Omega_i^{(1)}$ and $\Omega_i^{(2)}$. Then $(\prod_{i=1}^n \Omega_i^{(1)}, \prod_{i=1}^n \Omega_i^{(2)}, \prod_{i=1}^n \mu_i)$ defines two correlated spaces and the Markov operator T associated with them is given by $T = \otimes_{i=1}^n T_i$.*

EXAMPLE 2.5. For product spaces $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mu_i)$, the Bonami–Beckner operator $T = T_\rho$ is defined by

$$T = \otimes_{i=1}^n T_\rho^i, \quad (18)$$

where T^i is the Bonami–Beckner operator on $(\Omega_i \times \Omega_i, \mu_i)$. This Markov operator is the one most commonly discussed in previous work, see e.g. [KKL], [KhKMO1], [MOO2]. In a more recent work [DMR] the case of $\Omega_i \times \Omega_i$ with T_i a reversible Markov operator with respect to a measure μ_i on Ω_i was studied.

EXAMPLE 2.6. In the context of the “Majority is most predictable theorem”, Theorem 1.1, the underlying space is $\Omega = \{\pm 1\} \times \{0, \pm 1\}$ where element $(x, y) \in \Omega$ corresponds to a voter with vote x and a sampled vote y where either $y = x$ if the vote is queried or $y = 0$ otherwise. The probability measure μ is given by

$$\mu(x, y) = \frac{1}{2}(\delta(x = 1) + \delta(x = -1))(\rho\delta(y = x) + (1 - \rho)\delta(y = 0)).$$

Note that the marginal distributions on $\Omega_S = \{0, \pm 1\}$ and $\Omega_V = \{\pm 1\}$ are given by

$$\mu = (1 - \rho)\delta_0 + \frac{\rho}{2}(\delta_{-1} + \delta_1), \quad \nu = \frac{1}{2}(\delta_{-1} + \delta_1),$$

and

$$\nu(\cdot \mid \pm 1) = \delta_{\pm 1}, \quad \nu(\cdot \mid 0) = \frac{1}{2}(\delta_1 + \delta_{-1}).$$

Given independent copies μ_i of μ and ν_i of ν , the measure $\mu = \otimes_{i=1}^n \mu_i$ corresponds to the distribution of a sample of voters where each voter is sampled independently with probability ρ and the distribution of the voters is given by $\nu = \otimes_{i=1}^n \nu_i$.

EXAMPLE 2.7. The second non-reversible example is natural in the context of Condorcet voting. For simplicity, we first discuss the case of 3 possible outcomes. The general case is discussed later.

Let τ denote the uniform measure on the set permutations on the set $[3]$ denoted S_3 . Note that each element $\sigma \in S_3$ defines an element $f \in \{-1, 1\}^{\binom{3}{2}}$ by letting $f(i, j) = \text{sgn}(\sigma(i) - \sigma(j))$. The measure so defined, defines 3 correlated probability spaces $(\{\pm 1\}^{\binom{3}{2}}, \mathbf{P})$.

Note that the projection of $\mathbf{P}$ to each coordinate is uniform and

$$\begin{aligned} \mathbf{P}(f(3, 1) = -1 \mid f(1, 2) = f(2, 3) = 1) &= 0, \\ \mathbf{P}(f(3, 1) = 1 \mid f(1, 2) = f(2, 3) = -1) &= 0, \end{aligned}$$

and

$$\mathbf{P}(f(3, 1) = \pm 1 \mid f(1, 2) \neq f(2, 3)) = 1/2.$$

2.2 Properties of correlated spaces and Markov operators. Here we derive properties of correlated spaces and Markov operators that will be repeatedly used below. We start with the following which was already known to Rényi [Ré].

LEMMA 2.8. *Let $(\Omega^{(1)} \times \Omega^{(2)}, \mathbf{P})$ be two correlated spaces. Let f be a $\Omega^{(2)}$ measurable function with $\mathbf{E}[f] = 0$, and $\mathbf{E}[f^2] = 1$. Then among all g that are $\Omega^{(1)}$ measurable satisfying $\mathbf{E}[g^2] = 1$, a maximizer of $|\mathbf{E}[fg]|$ is given by*

$$g = \frac{Tf}{\sqrt{\mathbf{E}[(Tf)^2]}}, \quad (19)$$

where T is the Markov operator associated with $(\Omega^{(1)}, \Omega^{(2)})$. Moreover,

$$|\mathbf{E}[gf]| = \frac{|\mathbf{E}[fTf]|}{\sqrt{\mathbf{E}[(Tf)^2]}} = \sqrt{\mathbf{E}[(Tf)^2]}. \quad (20)$$

Proof. To prove (19) let h be an $\Omega^{(1)}$ measurable function with $\|h\|_2 = 1$. Write $h = \alpha g + \beta h'$ where $\alpha^2 + \beta^2 = 1$ and $\|h'\|_2 = 1$ is orthogonal to g . From the properties of conditional expectation it follows that $\mathbf{E}[fh'] = 0$. Therefore we may choose an optimizer satisfying $\alpha \in \pm 1$. Equation (20) follows since Tf is a conditional expectation. The same reasoning shows that $\mathbf{E}[fg] = 0$ for every $\Omega^{(1)}$ measurable function g if Tf is identically 0. $\square$

The following lemma is useful in bounding $\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P})$ from Definition 1.9 in generic situations. Roughly speaking, it shows that connectivity of the support of $\mathbf{P}$ on correlated spaces $\Omega^{(1)} \times \Omega^{(2)}$ implies that $\rho < 1$.

LEMMA 2.9. *Let $(\Omega^{(1)} \times \Omega^{(2)}, \mathbf{P})$ be two correlated spaces such that the probability of the smallest atom in $\Omega^{(1)} \times \Omega^{(2)}$ is at least $\alpha > 0$. Define a bi-partite graph $G = (\Omega^{(1)}, \Omega^{(2)}, E)$ where $(a, b) \in \Omega^{(1)} \times \Omega^{(2)}$ satisfies $(a, b) \in E$ if $\mathbf{P}(a, b) > 0$. Then if G is connected then*

$$\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P}) \leq 1 - \alpha^2/2.$$

Proof. For the proof it would be useful to consider $G' = (\Omega^{(1)} \cup \Omega^{(2)}, E)$, a weighted directed graph where the weight $W(a, b)$ of the directed edge from a to b is $\mathbf{P}[b|a]$ and the weight of the directed edge from b to a is $W(b, a) = \mathbf{P}[a|b]$. Note that the minimal non-zero weight must be at least α and that $W(a \rightarrow b) > 0$ iff $W(b \rightarrow a) > 0$. This later fact implies that G' is strongly connected. Note furthermore that G' is bi-partite.

Let A be the transition probability matrix defined by the weighted graph G' . Since G' is connected and $W(a, b) \geq \alpha$ for all a and b such that $W(a \rightarrow b) > 0$, it follows by Cheeger's inequality that the spectral gap of A is at least $\alpha^2/2$. Since G' is connected and bi-partite, the multiplicities of the eigenvalues ± 1 are both 1. Corresponding eigenfunctions are the constant 1 functions and the function taking the value 1 on $\Omega^{(1)}$ and the value -1 on $\Omega^{(2)}$.

In order to bound ρ it suffices by Lemma 2.8 to bound $\|Af\|_2$ for a function f that is supported on $\Omega^{(2)}$ and satisfies $\mathbf{E}[f] = 0$. Note that such a function is orthogonal to the eigenvectors of A corresponding to the eigenvalues -1 and 1 . It therefore follows that $\|Af\|_2 \leq (1 - \alpha^2/2)\|f\|_2$ as needed. $\square$

One nice property of Markov operators that will be used below is that they respect the Efron–Stein decomposition. Given a vector x in an n -dimensional product space and $S \subset [n]$ we write x_S for the vector $(x_i : i \in S)$. Given probability spaces $\Omega_1, \dots, \Omega_n$, we use the convention of writing X_i for a random variable that is distributed according to the measure of Ω_i and x_i for an element of Ω_i . We will also write X_S for $(X_i : i \in S)$.

DEFINITION 2.10. *Let $(\Omega_1, \mu_1), \dots, (\Omega_n, \mu_n)$ be discrete probability spaces $(\Omega, \mu) = \prod_{i=1}^n (\Omega_i, \mu_i)$. The Efron–Stein decomposition of $f : \Omega \rightarrow \mathbb{R}$ is given by*

$$f(x) = \sum_{S \subseteq [n]} f_S(x_S), \quad (21)$$

where the functions f_S satisfy

- f_S depends only on x_S .
- For all $S \not\subseteq S'$ and all $x_{S'}$, it holds that

$$\mathbf{E}[f_S \mid X_{S'} = x_{S'}] = 0.$$

It is well known that the Efron–Stein decomposition exists and that it is unique [ES]. We quickly recall the proof of existence. The function f_S is given by

$$f_S(x) = \sum_{S' \subseteq S} (-1)^{|S \setminus S'|} \mathbf{E}[f(X) \mid X_{S'} = x_{S'}]$$

which implies

$$\sum_S f_S(x) = \sum_{S'} \mathbf{E}[f \mid X_{S'} = x] \sum_{S: S' \subseteq S} (-1)^{|S \setminus S'|} = \mathbf{E}[f \mid X_{[n]} = x_{[n]}] = f(x).$$

Moreover, for $S \not\subseteq S'$ we have $\mathbf{E}[f_S \mid X_{S'} = x_{S'}] = \mathbf{E}[f_S \mid X_{S' \cap S} = x_{S' \cap S}]$ and for S' that is a strict subset of S we have

$$\begin{aligned} \mathbf{E}[f_S \mid X_{S'} = x_{S'}] &= \sum_{S'' \subset S} (-1)^{|S \setminus S''|} \mathbf{E}[f(X) \mid X_{S'' \cap S'} = x_{S'' \cap S'}] \\ &= \sum_{S'' \subset S'} \mathbf{E}[f(X) \mid X_{S''} = x_{S''}] \sum_{S'' \subset \tilde{S} \subset S'' \cup (S \setminus S')} (-1)^{|S \setminus \tilde{S}|} = 0. \end{aligned}$$

We now prove that the Efron–Stein decomposition “commutes” with Markov operators.

PROPOSITION 2.11. *Let $(\Omega_i^{(1)} \times \Omega_i^{(2)}, \mathbf{P}_i)$ be correlated spaces and let T_i the Markov operator associated with $\Omega_i^{(1)}$ and $\Omega_i^{(2)}$ for $1 \leq i \leq n$. Let*

$$\Omega^{(1)} = \prod_{i=1}^n \Omega_i^{(1)}, \quad \Omega^{(2)} = \prod_{i=1}^n \Omega_i^{(2)}, \quad \mathbf{P} = \prod_{i=1}^n \mathbf{P}_i, \quad T = \otimes_{i=1}^n T_i.$$

Suppose $f \in L^2(\Omega^{(2)})$ has Efron–Stein decomposition (21). Then the Efron–Stein decomposition of Tf satisfies

$$(Tf)_S = T(f_S).$$

Proof. Clearly $T(f_S)$ is a function of x_S only. Moreover, for all $S \not\subseteq S'$ and $x_{S'}^0$ it holds that

$$\begin{aligned} \mathbf{E}[(T(f_S))(X) \mid X_{S'} = x_{S'}^0] &= \mathbf{E}[f_S(Y) \mid X_{S'} = x_{S'}^0] \\ &= \mathbf{E}[\mathbf{E}[f_S(Y) \mid Y_{S'}] \mathbf{P}[Y_{S'} \mid X_{S'} = x_{S'}^0]] = 0, \end{aligned}$$

where the second equality follows from the fact that Y is independent of $X_{S'}$ condition on $Y_{S'}$. $\square$

We next derive a useful bound showing that in the setting above if $\rho(\Omega_i^{(1)} \times \Omega_i^{(2)}; \mathbf{P}) < 1$ for all i then Tf depends on the “low degree expansion” of f .

PROPOSITION 2.12. *Assume the setting of Proposition 2.11 and that further for all i it holds that $\rho(\Omega_i^{(1)}, \Omega_i^{(2)}; \mathbf{P}_i) \leq \rho_i$. Then for all f it holds that*

$$\|T(f_S)\|_2 \leq \left(\prod_{i \in S} \rho_i \right) \|f_S\|_2.$$

Proof. Without loss of generality it suffices to prove the statement of Proposition 2.12 for $S = [n]$. Thus our goal is to know that

$$\|Tf\|_2 \leq \left(\prod_{i=1}^n \rho_i \right) \|f\|_2$$

(from now on S will denote a set different from $[n]$). For each $0 \leq r \leq n$, let $T^{(r)}$ denote the following operator. $T^{(r)}$ maps a function g of $z = (z_1, \dots, z_n) = (x_1, \dots, x_{r-1}, y_r, \dots, y_n)$ to a function $T^{(r)}g$ of $w = (w_1, \dots, w_n) = (x_1, \dots, x_r, y_{r+1}, \dots, y_n)$ defined as follows:

$$T^{(r)}g(w) = \mathbf{E}[g(Z) \mid W = w].$$

(Here $Z = (X_1, \dots, X_{r-1}, Y_r, \dots, Y_n)$ and similarly W .)

Let g be a function such that for any subset $S \subsetneq [n]$ and all z_S ,

$$\mathbf{E}[g(Z) \mid Z_S = z_S] = 0.$$

We claim that

$$\|T^{(r)}g\|_2 \leq \rho_r \|g\|_2 \quad (22)$$

and that for all subsets $S \subsetneq [n]$ it holds that

$$\mathbf{E}[(T^{(r)}g)(W) \mid W_S = w_S] = 0. \quad (23)$$

Note that (22) and (23) together imply the desired bound as $T = T^{(n)} \dots T^{(1)}$.

For (22) note that if $S = [n] \setminus \{r\}$ and $f = T^{(r)}g$ then by Lemma 2.8

$$\begin{aligned} \mathbf{E}[f^2(W) \mid Z_S = z_S] &= \mathbf{E}[g(Z)f(W) \mid Z_S = z_S] \\ &\leq \rho_r \sqrt{\mathbf{E}[f^2(W) \mid Z_S = z_S] \mathbf{E}[g^2(Z) \mid Z_S = z_S]}. \end{aligned}$$

So

$$\mathbf{E}[f^2(W) \mid Z_S = z_S] \leq \rho_r^2 \mathbf{E}[g^2(W) \mid Z_S = z_S],$$

which gives $\|f\|_2 \leq \rho_r \|g\|_2$ by integration.

For (23) we note that if $S \subsetneq [n]$ then

$$\mathbf{E}[f(W) \mid W_S = w_S] = \mathbf{E}[g(Z) \mid W_S = w_S] = \mathbf{E}[\mathbf{E}[g(Z) \mid Z_S] \mathbf{P}[Z_S \mid W_S = w_S]] = 0.$$

This concludes the proof. $\square$

PROPOSITION 2.13. *Assume the setting of Proposition 2.11. Then*

$$\rho\left(\prod_{i=1}^n \Omega_i^{(1)}, \prod_{i=1}^n \Omega_i^{(2)}; \prod_{i=1}^n \mathbf{P}_i\right) = \max \rho(\Omega_i^{(1)}, \Omega_i^{(2)}).$$

Proof. Let $f \in L^2(\prod_{i=1}^n \Omega_i^{(2)})$ with $\mathbf{E}[f] = 0$ and $\mathbf{Var}[f] = 1$. Expand f according to its Efron–Stein decomposition

$$f = \sum_{S \neq \emptyset} f_S,$$

($f_\emptyset = 0$ since $\mathbf{E}[f] = 0$). Then by Propositions 2.11 and 2.12

$$\begin{aligned} \mathbf{E}[(Tf)^2] &= \mathbf{E}\left[\left(T\left(\sum_S f_S\right)\right)^2\right] = \mathbf{E}\left[\left(\sum_S T f_S\right)^2\right] = \sum_S \mathbf{E}[(T f_S)^2] \leq \sum_{S \neq \emptyset} \prod_{i \in S} \rho_i \|f_S\|_2^2 \\ &\leq \max_i \rho_i^2 \sum_{S \neq \emptyset} \|f_S\|_2^2 = \max_i \rho_i^2. \end{aligned}$$

The other inequality is trivial. $\square$

3 Background: Influences and Hypercontractivity

In this section we recall and generalize some definitions and results from [MOO2]. In particular, the generalizations allow us the study non-reversible Markov operators and correlated ensembles. For the reader who is familiar with [MOO2] it suffices to look at subsections 3.3 and 3.5.

3.1 Influences and noise stability in product spaces. Let $(\Omega_1, \mu_1), \dots, (\Omega_n, \mu_n)$ be probability spaces and let (Ω, μ) denote the product probability space. Let

$$f = (f_1, \dots, f_k) : \Omega_1 \times \dots \times \Omega_n \rightarrow \mathbb{R}^k.$$

DEFINITION 3.1. *The influence of the i th coordinate on f is*

$$\text{Inf}_i(f) = \sum_{1 \leq j \leq k} \mathbf{E}[\text{Var}[f_j \mid x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]].$$

3.2 Multilinear polynomials. In this subsection we recall and slightly generalize the setup and notation used in [MOO2]. Recall that we are interested in functions on product of finite probability spaces, $f : \Omega_1 \times \dots \times \Omega_n \rightarrow \mathbb{R}$. For each i , the space of all functions $\Omega_i \rightarrow \mathbb{R}$ can be expressed as the span of a finite set of orthonormal random variables, $X_{i,0} = 1, X_{i,1}, X_{i,2}, X_{i,3}, \dots$; then f can be written as a multilinear polynomial in the $X_{i,j}$'s. In fact, it will be convenient for us to mostly disregard the Ω_i 's and work directly with sets of orthonormal random variables; in this case, we can even drop the restriction of finiteness. We thus begin with the following definition:

DEFINITION 3.2 [MOO2]. *We call a collection of finitely many orthonormal real random variables, one of which is the constant 1, an orthonormal ensemble. We will write a typical sequence of n orthonormal ensembles as $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_n)$, where $\mathcal{X}_i = \{X_{i,0} = 1, X_{i,1}, \dots, X_{i,m_i}\}$. We call a sequence of orthonormal ensembles $\mathcal{X}$ independent if the ensembles are independent families of random variables.*

We will henceforth be concerned only with independent sequences of orthonormal ensembles, and we will call these sequences of ensembles, for brevity. Similarly, when writing an ensemble we will always mean an orthogonal ensemble.

REMARK 3.3 [MOO2]. Given a sequence of independent random variables $X_1, \dots, X_n$ with $\mathbf{E}[X_i] = 0$ and $\mathbf{E}[X_i^2] = 1$, we can view them as a sequence of ensembles $\mathcal{X}$ by renaming $X_i = X_{i,1}$ and setting $X_{i,0} = 1$ as required.

DEFINITION 3.4 [MOO2]. *We denote by $\mathcal{G}$ the Gaussian sequence of ensembles, in which $\mathcal{G}_i = \{G_{i,0} = 1, G_{i,1}, G_{i,2}, \dots, G_{i,m_i}\}$ and all $G_{i,j}$'s with $j \geq 1$ are independent standard Gaussians.*

The Gaussian ensembles discussed in this paper will often have m_i chosen to match the m_i of a given ensemble.

As mentioned, we will be interested in *multilinear polynomials* over sequences of ensembles. By this we mean sums of products of the random variables, where each product is obtained by multiplying one random variable from each ensemble.

DEFINITION 3.5 [MOO2]. A multi-index σ is a sequence $(\sigma_1, \dots, \sigma_n)$ in $\mathbb{N}^n$. The degree of σ , denoted $|\sigma|$, is $|\{i \in [n] : \sigma_i > 0\}|$. Given a doubly-indexed set of indeterminates $\{x_{i,j}\}_{i \in [n], j \in \mathbb{N}}$, we write x_σ for the monomial $\prod_{i=1}^n x_{i,\sigma_i}$. We now define a multilinear polynomial over such a set of indeterminates to be any expression

$$Q(x) = \sum_{\sigma} c_{\sigma} x_{\sigma} \tag{24}$$

where the c_σ 's are real constants, all but finitely many of which are zero. The degree of $Q(x)$ is $\max\{|\sigma| : c_\sigma \neq 0\}$, at most n . We also use the notation

$$Q^{\leq d}(x) = \sum_{|\sigma| \leq d} c_\sigma x_\sigma$$

and, analogously, $Q^{=d}(x)$ and $Q^{>d}(x)$.

Naturally, we will consider applying multilinear polynomials Q to sequences of ensembles $\mathcal{X}$; the distribution of these random variables $Q(\mathcal{X})$ is the subject of our invariance principle. Since $Q(\mathcal{X})$ can be thought of as a function on a product space $\Omega_1 \times \cdots \times \Omega_n$ as described at the beginning of this section, there is a consistent way to define the notions of influences, T_ρ , and noise stability from section 3.1. For example, the “influence of the i th ensemble on Q ” is

$$\text{Inf}_i(Q(\mathcal{X})) = \mathbf{E}[\text{Var}[Q(\mathcal{X}) \mid \mathcal{X}_1, \dots, \mathcal{X}_{i-1}, \mathcal{X}_{i+1}, \dots, \mathcal{X}_n]].$$

Using independence and orthonormality, it is easy to show the following formulas:

PROPOSITION 3.6. *Let $\mathcal{X}$ be a sequence of ensembles and Q a multilinear polynomial as in (24). Then*

$$\mathbf{E}[Q(\mathcal{X})] = c_0; \quad \mathbf{E}[Q(\mathcal{X})^2] = \sum_{\sigma} c_\sigma^2; \quad \text{Var}[Q(\mathcal{X})] = \sum_{|\sigma| > 0} c_\sigma^2;$$

and

$$\text{Inf}_i(Q(\mathcal{X})) = \sum_{\sigma: \sigma_i > 0} c_\sigma^2.$$

For $\rho \in [0, 1]$ we define the operator T_ρ as acting formally on multilinear polynomials $Q(x)$ as in (24) by

$$(T_\rho Q)(x) = \sum_{\sigma} \rho^{|\sigma|} c_\sigma x_\sigma. \quad (25)$$

We note the definition in (17) and (18) are consistent with the definition in (25) in the sense that for any ensemble $\mathcal{X}$ the two definitions result in the same function $(T_\rho Q)(\mathcal{X})$.

We finally recall the notion of “low-degree influences”, a notion that has proven crucial in the analysis of PCPs in hardness of approximation in computer science (see, e.g. [KhKMO1]).

DEFINITION 3.7 [MOO2]. *The d -low-degree influence of the i th ensemble on $Q(\mathcal{X})$ is*

$$\text{Inf}_i^{\leq d}(Q(\mathcal{X})) = \text{Inf}_i^{\leq d}(Q) = \sum_{\sigma: |\sigma| \leq d, \sigma_i > 0} c_\sigma^2.$$

Note that this gives a way to define low-degree influences $\text{Inf}_i^{\leq d}(f)$ for functions $f : \Omega_1 \times \cdots \times \Omega_n \rightarrow \mathbb{R}$ on finite product spaces.

There isn't an especially natural interpretation of $\text{Inf}_i^{\leq d}(f)$. However, the notion is important for PCPs due to the fact that a function with variance 1 cannot have too many coordinates with substantial low-degree influence; this is reflected in the following easy proposition:

PROPOSITION 3.8 [MOO2]. Suppose Q is multilinear polynomial as in (24). Then

$$\sum_i \text{Inf}_i^{\leq d}(Q) \leq d \cdot \text{Var}[Q].$$

The proof follows since

$$\sum_i \text{Inf}_i^{\leq d}(Q) = \sum_i \sum_{\sigma: |\sigma| \leq d, \sigma_i > 0} c_\sigma^2 = \sum_{\sigma: 0 < |\sigma| \leq d} |\sigma| c_\sigma^2 \leq d \sum_{\sigma: 0 < |\sigma|} c_\sigma^2 = d \text{Var}[Q].$$

3.3 Vector-valued multilinear polynomials. For the invariance principle discussed here we will need to consider vector-valued multilinear polynomials.

DEFINITION 3.9. A k -dimensional multilinear polynomial over a set of indeterminates is given by

$$Q = (Q_1, \dots, Q_k) \quad (26)$$

where each Q_j is a multilinear polynomial as in (24). The degree of Q is the maximal degree of the Q_j 's.

DEFINITION 3.10. We adopt the standard notation and write $\|Q\|_q$ for $\mathbf{E}^{1/q}[\sum_{j=1}^k |Q_j|^q]$; we write $\text{Var}[Q]$ for $\mathbf{E}[\|Q - \mathbf{E}[Q]\|_2^2]$ and

$$\text{Inf}_i(Q(\mathcal{X})) = \sum_{j=1}^k \text{Inf}_i(Q_j(\mathcal{X})).$$

Using these definitions, it is easy to see

PROPOSITION 3.11. Let $\mathcal{X}$ be a sequence of ensembles and $Q = (Q_1, \dots, Q_k)$ a k -dimensional multilinear polynomial where Q_j is defined as in (24) with c_σ^j as its coefficients. Then,

$$\mathbf{E}[Q(\mathcal{X})] = (c_0^1, \dots, c_0^k); \quad \|Q(\mathcal{X})\|_2^2 = \sum_{j, \sigma} c_\sigma^{j2}; \quad \text{Var}[Q(\mathcal{X})] = \sum_{j, |\sigma^j| > 0} c_\sigma^{j2}.$$

Finally, we recall the standard multi-index notation associated with k -dimensional multilinear polynomials. A *multi-index* $\mathbf{i}$ of dimension k is a vector $(i_1, \dots, i_k)$, where each i_j is an integer. We write $|\mathbf{i}|$ for $i_1 + \dots + i_k$ and $\mathbf{i}!$ for $i_1! i_2! \dots i_k!$. Given a function ψ of k variables, we will write $\psi^{(\mathbf{i})}$ for the partial derivative of f taken i_1 times with respect to the first variable, i_2 with respect to the second etc. (We will only consider functions ψ that are smooth enough that the order of derivatives does not matter.) We will also write $Q^{\mathbf{i}}$ for the product $Q_1^{i_1} \dots Q_k^{i_k}$.

3.4 Hypercontractivity. As in [MOO2] the invariance principle requires that the ensembles involved are hypercontractive. Recall that Y is $(2, q, \eta)$ -hypercontractive with some $\eta \in (0, 1)$ if and only if $\mathbf{E}[Y] = 0$ and $\mathbf{E}[|Y|^q] < \infty$. Also, if Y is $(2, q, \eta)$ -hypercontractive then $\eta \leq (q - 1)^{-1/2}$.

DEFINITION 3.12. Let $\mathcal{X}$ be a sequence of ensembles. For $1 \leq p \leq q < \infty$ and $0 < \eta < 1$ we say that $\mathcal{X}$ is (p, q, η) -hypercontractive if

$$\|(T_\eta Q)(\mathcal{X})\|_q \leq \|Q(\mathcal{X})\|_p$$

for every multilinear polynomial Q over $\mathcal{X}$.

Since T_η is a contractive semi-group, we have

REMARK 3.13. If $\mathcal{X}$ is (p, q, η) -hypercontractive then it is (p, q, η') -hypercontractive for any $0 < \eta' \leq \eta$.

There is a related notion of hypercontractivity for *sets* of random variables which considers all polynomials in the variables, not just multilinear polynomials; see, e.g. Janson [J]. We summarize some of the basic properties below, see [MOO2] for details.

PROPOSITION 3.14 [MOO2]. Suppose $\mathcal{X}$ is a sequence of n_1 ensembles and $\mathcal{Y}$ is an independent sequence of n_2 ensembles. Assume both are (p, q, η) -hypercontractive. Then the sequence of ensembles $\mathcal{X} \cup \mathcal{Y} = (\mathcal{X}_1, \dots, \mathcal{X}_{n_1}, \mathcal{Y}_1, \dots, \mathcal{Y}_{n_2})$ is also (p, q, η) -hypercontractive.

PROPOSITION 3.15 [MOO2]. Let $\mathcal{X}$ be a $(2, q, \eta)$ -hypercontractive sequence of ensembles and Q a multilinear polynomial over $\mathcal{X}$ of degree d . Then

$$\|Q(\mathcal{X})\|_q \leq \eta^{-d} \|Q(\mathcal{X})\|_2.$$

We end this section by recording the some hypercontractive estimates to be used later. The result for ± 1 Rademacher variables is well known and due originally to Bonami [Bo] and independently Beckner [B]; the same result for Gaussian and uniform random variables is also well known and in fact follows easily from the Rademacher case. The optimal hypercontractivity constants for general finite spaces was recently determined by Wolff [W] (see also [Ol]):

Theorem 3.16. Let X denote either a uniformly random ± 1 bit, a standard one-dimensional Gaussian, or a random variable uniform on $[-\sqrt{3}, \sqrt{3}]$. Then X is $(2, q, (q-1)^{-1/2})$ -hypercontractive.

Theorem 3.17 (Wolff [W]). Let X be any mean-zero random variable on a finite probability space in which the minimum nonzero probability of any atom is $\alpha \leq 1/2$. Then X is $(2, q, \eta_q(\alpha))$ -hypercontractive, where

$$\eta_q(\alpha) = \left(\frac{A^{1/q'} - A^{-1/q'}}{A^{1/q} - A^{-1/q}} \right)^{-1/2}$$

with $A = \frac{1-\alpha}{\alpha}, \quad \frac{1}{q} + \frac{1}{q'} = 1.$

Note the following special case:

PROPOSITION 3.18 [W].

$$\eta_3(\alpha) = (A^{1/3} + A^{-1/3})^{-1/2} \alpha \stackrel{\alpha \rightarrow 0}{\sim} \alpha^{1/6},$$

and also

$$\frac{1}{2} \alpha^{1/6} \leq \eta_3(\alpha) \leq 2^{-1/2},$$

for all $\alpha \in [0, 1/2]$.

3.5 Vector hypercontraction. For our purposes we will also need to obtain hypercontraction results in cases where Q is a k -dimensional multilinear polynomial. We will need to consider vector-valued multilinear polynomials.

PROPOSITION 3.19. *Let $\mathcal{X}$ be a $(2, q, \eta)$ -hypercontractive sequence of ensembles and Q a multilinear polynomial over $\mathcal{X}$ of degree d and dimension k . Assume q is integer and let $\mathbf{i}$ be a multi-index with $|\mathbf{i}| = q$. Then*

$$\mathbf{E}[|Q^{\mathbf{i}}(\mathcal{X})|] \leq \eta^{-dq} \prod_{j=1}^k \mathbf{E}[Q_j^2]^{i_j/2}.$$

Proof.

$$\mathbf{E}[|Q^{\mathbf{i}}(\mathcal{X})|] \leq \prod_{j=1}^k \mathbf{E}[|Q_j|^q]^{i_j/q} \leq \eta^{-dq} \prod_{j=1}^k \mathbf{E}[Q_j^2]^{i_j/2},$$

where the first inequality is Hölder and the second follows by hypercontractivity. $\square$

4 Multidimensional Invariance Principle

In this section we generalize the invariance principle from [MOO2] to the multi-dimensional setting. We omit some easy steps that are either identical or easy adaptation of the proofs of [MOO2].

4.1 Hypotheses for invariance theorems. Below we will prove a generalization of the invariance principle [MOO2]. The invariance principle proven there concerns a multilinear polynomial Q over two hypercontractive sequences of ensembles, $\mathcal{X}$ and $\mathcal{Y}$; furthermore, $\mathcal{X}$ and $\mathcal{Y}$ are assumed to satisfy a “matching moments” condition, described below.

It is possible to generalize the invariance principle to vector-valued multilinear polynomials under each of the hypercontractivity assumptions **H1**, **H2**, **H3** and **H4** of [MOO2]. However, since the proof of all generalizations is essentially the same and since for the applications studied here it suffices to consider the hypothesis **H3**, this is the only hypothesis that will be discussed in the paper. It is defined as follows:

H3 Let $\mathcal{X}$ be a sequence of n ensembles in which the random variables in each ensemble $\mathcal{X}_i$ form a basis for the real-valued functions on some finite probability space Ω_i . Further assume that the least nonzero probability of any atom in any Ω_i is $\alpha \leq \frac{1}{2}$, and let $\eta = \frac{1}{2}\alpha^{1/6}$. Let $\mathcal{Y}$ be any $(2, 3, \eta)$ -hypercontractive sequence of ensembles such that all the random variables in $\mathcal{Y}$ are independent. Finally, let Q be a k -dimensional multilinear polynomial as in (26).

4.2 Functional setting. The essence of our invariance principle is that if Q is of bounded degree and has low influences then the random variables $Q(\mathcal{X})$ and $Q(\mathcal{Y})$ are close in distribution. The simplest way to formulate this conclusion is to say that if $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$ is a sufficiently nice “test function” then $\Psi(Q(\mathcal{X}))$ and $\Psi(Q(\mathcal{Y}))$ are close in expectation.

Theorem 4.1. *Assume hypothesis **H3**. Further assume that Q is a k -dimensional multilinear polynomial, that $\mathbf{Var}[Q] \leq 1$, $\deg(Q) \leq d$, and $\mathbf{Inf}_i(Q) \leq \tau$ for all i . Let $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$ be a C^3 function with $|\Psi^{(\mathbf{i})}| \leq B$ uniformly for every vector $\mathbf{i}$ with*

$|\mathbf{i}| = \mathbf{3}$. Then

$$|\mathbf{E}[\Psi(Q(\mathcal{X}))] - \mathbf{E}[\Psi(Q(\mathcal{Y}))]| \leq \epsilon := 2dB k^3 (8\alpha^{-1/2})^d \tau^{1/2}.$$

Proof. Note that by Proposition 3.18, the random variables satisfy $(2, 3, \eta)$ hypercontractivity with $\eta = \frac{1}{2}\alpha^{1/6}$.

We begin by defining intermediate sequences between $\mathcal{X}$ and $\mathcal{Y}$. For $i = 0, 1, \dots, n$, let $\mathcal{Z}^{(i)}$ denote the sequence of n ensembles $(\mathcal{Y}_1, \dots, \mathcal{Y}_i, \mathcal{X}_{i+1}, \dots, \mathcal{X}_n)$ and let $Q^{(i)} = Q(\mathcal{Z}^{(i)})$. Our goal will be to show

$$|\mathbf{E}[\Psi(Q^{(i-1)})] - \mathbf{E}[\Psi(Q^{(i)})]| \leq 2Bk^3 \eta^{-3d} \text{Inf}_i(Q)^{3/2} \quad (27)$$

for each $i \in [n]$. Summing this over i will complete the proof since $\mathcal{Z}^{(0)} = \mathcal{X}$, $\mathcal{Z}^{(n)} = \mathcal{Y}$, and

$$\sum_{i=1}^n \text{Inf}_i(Q)^{3/2} \leq \tau^{1/2} \cdot \sum_{i=1}^n \text{Inf}_i(Q) = \tau^{1/2} \cdot \sum_{i=1}^n \text{Inf}_i^{\leq d}(Q) \leq d\tau^{1/2},$$

where we used Proposition 3.8 and $\sum_j \mathbf{Var}[Q_j] \leq 1$.

Let us fix a particular $i \in [n]$ and proceed to prove (27). Given a multi-index σ , write $\sigma \setminus i$ for the same multi-index except with $\sigma_i = 0$. Now write

$$\begin{aligned} \tilde{Q} &= \sum_{\sigma: \sigma_i=0} c_\sigma \mathcal{Z}_\sigma^{(i)}, \\ R &= \sum_{\sigma: \sigma_i>0} c_\sigma X_{i, \sigma_i} \cdot \mathcal{Z}_{\sigma \setminus i}^{(i)}, \\ S &= \sum_{\sigma: \sigma_i>0} c_\sigma Y_{i, \sigma_i} \cdot \mathcal{Z}_{\sigma \setminus i}^{(i)}. \end{aligned}$$

Note that $\tilde{Q}$ and the variables $\mathcal{Z}_{\sigma \setminus i}^{(i)}$ are independent of the variables in $\mathcal{X}_i$ and $\mathcal{Y}_i$ and that $Q^{(i-1)} = \tilde{Q} + R$ and $Q^{(i)} = \tilde{Q} + S$.

To bound the left side of (27), i.e. $|\mathbf{E}[\Psi(\tilde{Q} + R) - \Psi(\tilde{Q} + S)]|$, we use Taylor's theorem: for all $x, y \in \mathbb{R}$,

$$\left| \Psi(x+y) - \sum_{|\mathbf{k}|<3} \frac{\Psi^{(\mathbf{k})}(x) y^{\mathbf{k}}}{\mathbf{k}!} \right| \leq \sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!} |y|^{\mathbf{k}}.$$

In particular,

$$\left| \mathbf{E}[\Psi(\tilde{Q} + R)] - \sum_{|\mathbf{k}|<3} \mathbf{E} \left[\frac{\Psi^{(\mathbf{k})}(\tilde{Q}) R^{\mathbf{k}}}{\mathbf{k}!} \right] \right| \leq \sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!} \mathbf{E}[|R|^{\mathbf{k}}] \quad (28)$$

and similarly,

$$\left| \mathbf{E}[\Psi(\tilde{Q} + S)] - \sum_{|\mathbf{k}|<3} \mathbf{E} \left[\frac{\Psi^{(\mathbf{k})}(\tilde{Q}) S^{\mathbf{k}}}{\mathbf{k}!} \right] \right| \leq \sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!} \mathbf{E}[|S|^{\mathbf{k}}]. \quad (29)$$

We will see below that R and S have finite 3'rd moments. Moreover, for $0 \leq \mathbf{k} \leq \mathbf{r}$ with $|\mathbf{r}| = 3$ it holds that $|\Psi^{(\mathbf{k})}(\tilde{Q}) R^{\mathbf{k}}| \leq |\mathbf{k}! B \tilde{Q}^{\mathbf{r}-\mathbf{k}} R^{\mathbf{k}}|$ (and similarly for S). Thus all moments above are finite. We now claim that for all $0 \leq |\mathbf{k}| < 3$ it holds that

$$\mathbf{E}[\Psi^{(\mathbf{k})}(\tilde{Q}) R^{\mathbf{k}}] = \mathbf{E}[\Psi^{(\mathbf{k})}(\tilde{Q}) S^{\mathbf{k}}]. \quad (30)$$

This follows from the fact that the expressions in the expected values when viewed as multilinear polynomials in the variables in $\mathcal{X}_i$ and $\mathcal{Y}_i$ respectively are of degree ≤ 2 and each monomial term in $\mathcal{X}_i$ has the same coefficient as the corresponding monomial in $\mathcal{Y}_i$.

From (28), (29) and (30) it follows that

$$|\mathbf{E}[\Psi(\tilde{Q} + R) - \Psi(\tilde{Q} + S)]| \leq \sum_{r=3} \frac{B}{r!} (\mathbf{E}[|R|^r] + \mathbf{E}[|S|^r]). \quad (31)$$

We now use hypercontractivity. By Proposition 3.14 each $\mathcal{Z}^{(i)}$ is $(2, r, \eta)$ -hypercontractive. Thus by Proposition 3.19,

$$\mathbf{E}[|R|^r] \leq \eta^{-3d} \prod_{j=1}^k \mathbf{E}[R_j^2]^{r_j/2}, \quad \mathbf{E}[|S|^r] \leq \eta^{-3d} \prod_{j=1}^k \mathbf{E}[S_j^2]^{r_j/2}. \quad (32)$$

However,

$$\mathbf{E}[S_j^2] = \mathbf{E}[R_j^2] = \sum_{\sigma^j: \sigma_i^j > 0} c_{\sigma^j}^2 = \text{Inf}_i(Q_j) \leq \text{Inf}_i(Q). \quad (33)$$

Combining (31), (32) and (33) it follows that

$$|\mathbf{E}[\Psi(\tilde{Q} + R) - \Psi(\tilde{Q} + S)]| \leq 2Bk^3 \eta^{-3d} \cdot \text{Inf}_i(Q)^{3/2}$$

confirming (27) and completing the proof. $\square$

4.3 Invariance principle – other functionals, and smoothed version. The basic invariance principle shows that $\mathbf{E}[\Psi(Q(\mathcal{X}))]$ and $\mathbf{E}[\Psi(Q(\mathcal{Y}))]$ are close if Ψ is a $\mathcal{C}^3$ functional with bounded 3rd derivative. To show that the distributions of $Q(\mathcal{X})$ and $Q(\mathcal{Y})$ are close in other senses we need the invariance principle for less smooth functionals. This we can obtain using straightforward approximation arguments, see for example [MOO2]. For applications involving bounded functions, it will be important to bound the following functionals. We let $f_{[0,1]} : \mathbb{R} \rightarrow \mathbb{R}$ be defined by

$$f_{[0,1]}(x) = \max(\min(x, 1), 0) = \min(\max(x, 0), 1),$$

and $\zeta : \mathbb{R}^k \rightarrow \mathbb{R}$ be defined by

$$\zeta(x) = \sum_{i=1}^k (x_i - f_{[0,1]}(x_i))^2. \quad (34)$$

Similarly, we define

$$\chi(x) = \prod_{i=1}^k f_{[0,1]}(x_i). \quad (35)$$

Repeating the proofs of [MOO2] one obtains

Theorem 4.2. *Assume hypothesis **H3**. Further assume that $Q = (Q_1, \dots, Q_k)$ is a k -dimensional multilinear polynomial with $\mathbf{Var}[Q] \leq 1$ and $\text{Inf}_i^{\leq \log(1/\tau)/K}(Q) \leq \tau$ for all i , where*

$$K = \log(1/\alpha).$$

Suppose further that for all d it holds that $\mathbf{Var}[Q^{>d}] \leq (1 - \gamma)^{2d}$ where $0 < \gamma < 1$.

Write $R = Q(\mathcal{X})$ and $S = Q(\mathcal{Y})$. Then

$$|\mathbf{E}[\zeta(R)] - \mathbf{E}[\zeta(S)]| \leq \tau^{\Omega(\gamma/K)},$$

where $\Omega(\cdot)$ hides a constant depending only on k . Similarly,

$$|\mathbf{E}[\chi(R)] - \mathbf{E}[\chi(S)]| \leq \tau^{\Omega(\gamma/K)}.$$

Proof. The proof for ζ uses the fact that the function ζ admits approximations ζ_λ such that

•

$$\|\zeta - \zeta_\lambda\|_\infty = O(\lambda^2),$$

•

$$\|\zeta_\lambda^{\mathbf{r}}\|_\infty = O(\lambda^{-1}),$$

for all $\mathbf{r}$ with $|\mathbf{r}| = 3$.

This implies that for all k -dimensional degree d polynomials we have

$$|\mathbf{E}[\zeta(R)] - \mathbf{E}[\zeta(S)]| \leq O(\alpha^{-d/3} \tau^{-1/3}). \quad (36)$$

See [MOO2] for details. In order to obtain the result for polynomials with decaying tails we use the fact that

$$|\zeta(a_1 + b_1, \dots, a_k + b_k) - \zeta(a_1, \dots, a_k)| \leq \sum_{i=1}^k (|a_i b_i| + b_i^2).$$

This implies that evaluating ζ at polynomials truncated at level d results in an error of at most $O(\exp(-d\gamma))$ which together with the bound (36) implies the desired bound for ζ .

The proof for χ is similar as the function ζ admits approximations χ_λ such that

•

$$\|\chi - \chi_\lambda\|_\infty = O(\lambda),$$

•

$$\|\chi_\lambda^{\mathbf{r}}\|_\infty = O(\lambda^{-2}),$$

for all $\mathbf{r}$ with $|\mathbf{r}| = 3$. □

The following corollary is of particular interest to us.

COROLLARY 4.3. *Assume hypothesis **H3**. For each $1 \leq i \leq n$ and $1 \leq j \leq k$, let $\mathcal{X}_i^j$ denote an ensemble all of whose elements are linear combinations of elements in $\mathcal{X}_i$. Similarly for each $1 \leq i \leq n$ and $1 \leq j \leq k$, let $\mathcal{Y}_i^j$ denote an ensemble all of whose elements are linear combinations of elements in $\mathcal{Y}_i$.*

Assume further that for all i and j it holds that $|\mathcal{Y}_i^j| = |\mathcal{X}_i^j|$ and that there is a one-to-one correspondence between $Y \in \mathcal{Y}_i^j$ given by

$$Y = \sum_{r=0}^{|\mathcal{Y}_i|-1} a(i, j, k) Y_k,$$

and $X \in \mathcal{X}_i^j$ given by

$$X = \sum_{r=0}^{|\mathcal{X}_i|-1} a(i, j, k) X_k,$$

where $\mathcal{Y}_i = \{Y_0, \dots, Y_{|\mathcal{Y}_i|-1}\}$ and $\mathcal{X}_i = \{X_0, \dots, X_{|\mathcal{X}_i|-1}\}$.

Let $Q = (Q_1, \dots, Q_k)$ a be multilinear polynomial with $\mathbf{Var}[Q] \leq 1$ and $\inf_i^{\leq \log(1/\tau)/K} (Q) \leq \tau$ for all i where

$$K = \log(1/\alpha).$$

Suppose further that for all d it holds that $\mathbf{Var}[Q_j^{>d}] \leq (1 - \gamma)^{2d}$ where $0 < \gamma < 1$. Write $R = (Q_1(\mathcal{X}^1), \dots, Q_k(\mathcal{X}^k))$ and $S = (Q_1(\mathcal{Y}^1), \dots, Q_k(\mathcal{Y}^k))$. Then

$$|\mathbf{E}[\zeta(R)] - \mathbf{E}[\zeta(S)]| \leq \tau^{\Omega(\gamma/K)},$$

and

$$|\mathbf{E}[\chi(R)] - \mathbf{E}[\chi(S)]| \leq \tau^{\Omega(\gamma/K)},$$

where the $\Omega(\cdot)$ hides a constant depending only on k .

Proof. The proof follows immediately from the previous theorem noting that $\inf_i$ and $\mathbf{Var}[Q^{>d}]$ are basis independent. $\square$

5 Noise in Gaussian Space

In this section we derive the Gaussian correlation bounds needed for our applications. The first bound derived in subsection 5.1 is an easy extension of [Bor]. The second one gives a quantitative estimate on iterations of the first bound that will be needed for some of the applications.

5.1 Noise stability in Gaussian space. We begin by recalling some definitions and results relevant for “Gaussian noise stability”. Throughout this section we consider $\mathbb{R}^n$ to have the standard n -dimensional Gaussian distribution, and our probabilities and expectations are over this distribution. Recall Definition 1.12. We denote by U_ρ the Ornstein–Uhlenbeck operator acting on $L^2(\mathbb{R}^n)$ by

$$(U_\rho f)(x) = \mathbf{E}[f(\rho x + \sqrt{1 - \rho^2} y)],$$

where y is a standard n -dimensional Gaussian.

It is immediate to see that if $f, g \in L^2(\mathbb{R}^n)$ then

$$\mathbf{E}[f U_\rho g] = \mathbf{E}[g U_\rho f] = \mathbf{E}[f(X_1, \dots, X_n)g(Y_1, \dots, Y_n)]$$

where (X_i, Y_i) are independent two-dimensional Gaussian vectors with covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The results of Borell [Bor] imply the following (see [MOO2] for more details):

Theorem 5.1. *Let $f, g : \mathbb{R}^n \rightarrow [0, 1]$ be two measurable functions on Gaussian space with $\mathbf{E}[f] = \mu$ and $\mathbf{E}[g] = \nu$. Then for all $0 \leq \rho \leq 1$ we have*

$$\overline{\Gamma}_\rho(\mu, \nu) \leq \mathbf{E}[f U_\rho g] \leq \overline{\Gamma}_\rho(\mu, \nu).$$

We will need the following corollary.

COROLLARY 5.2. *Let $X_1, \dots, X_n, Y_1, \dots, Y_m$ be jointly Gaussian such that $X_1, \dots, X_n$ are independent, $Y_1, \dots, Y_m$ are independent and*

$$\sup_{\|\alpha\|_2 = \|\beta\|_2 = 1} \left| \mathbf{Cov} \left[\sum_i \alpha_i X_i, \sum_i \beta_i Y_i \right] \right| \leq \rho.$$

Let $f_1 : \mathbb{R}^n \rightarrow [0, 1]$, $f_2 : \mathbb{R}^m \rightarrow [0, 1]$ with $\mathbf{E}[f_j] = \mu_j$. Then

$$\underline{\Gamma}_\rho(\mu_1, \mu_2) \leq \mathbf{E}[f_1 f_2] \leq \overline{\Gamma}_\rho(\mu_1, \mu_2).$$

Proof. Note that if $m > n$ then we may define $X_{n+1}, \dots, X_m$ to be Gaussian and independent of all the other variables. This implies that without loss of generality we may assume that $m = n$.

We claim that without loss of generality the covariance matrix between X and Y is given by

$$\mathbf{Cov}[X_i, Y_j] = 0,$$

for $i \neq j$ and $|\mathbf{Cov}[X_i, Y_i]| \leq \rho$ for all i . To see this, take $\|\alpha\|_2 = \|\beta\|_2 = 1$ which maximizes $\mathbf{Cov}[\sum \alpha_i X_i, \sum \beta_i Y_i]$ and define $\tilde{X}_1 = \sum \alpha_i X_i$, $\tilde{Y}_1 = \sum \beta_i Y_i$ and $\tilde{X}_2, \dots, \tilde{X}_n$ to be orthonormal basis of the projection of the span of $\{X_2, \dots, X_n\}$ to the orthogonal complement of the space spanned by $\tilde{X}_1$ and similarly $\tilde{Y}_2, \dots, \tilde{Y}_n$. It is easy to check that $\mathbf{Cov}[\tilde{X}_1, \tilde{Y}_i] = \mathbf{Cov}[\tilde{Y}_1, \tilde{X}_i] = 0$ for $i > 1$. Repeating this process for $\tilde{X}_2, \dots, \tilde{X}_n$ and $\tilde{Y}_2, \dots, \tilde{Y}_n$, etc., we obtain the desired covariance matrix.

Write $\mathbf{Cov}[X, Y] = \rho J$, where J is diagonal with all entries in $[-1, 1]$. Then clearly

$$\mathbf{E}[f_1(X_1, \dots, X_n) f_2(Y_1, \dots, Y_n)] = \mathbf{E}[f_1 U_\rho(U_J f_2)],$$

where U_ρ is the Ornstein–Uhlenbeck operator and U_J is the operator defined by

$$(U_J f)(x_1, \dots, x_n) = \mathbf{E}\left[f\left(j_1 x_1 + \sqrt{1 - J_{1,1}^2} y_1, \dots, j_n x_n + \sqrt{1 - J_{n,n}^2} y_n\right)\right],$$

where y is distributed according to the Gaussian measure. Since U_J is a Markov operator, we have $\mathbf{E}[U_J f_2] = \mathbf{E}[f_2]$ and $0 \leq U_J f_2 \leq 1$. Now applying Borell's result we obtain the desired conclusion. $\square$

We note that in general there is no closed form for $\Gamma_\rho(\mu, \nu)$; however, for balanced functions we have Sheppard's formula [Sh]: $\Gamma_\rho(1/2, 1/2) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho$. Finally we record a fact to be used later.

PROPOSITION 5.3. *Let $((X_{i,j}, Y_{i,j})_j)_{i=1}^n$ be jointly Gaussian, each distributed $N(0, 1)$. Suppose further that, for each i ,*

$$\sup_{\|\alpha\|_2 = \|\beta\|_2 = 1} \left| \mathbf{Cov}\left[\sum_j \alpha_j X_{i,j}, \sum_j \beta_j Y_{i,j}\right] \right| \leq \rho,$$

and that the n collections $((X_{i,j}, Y_{i,j})_j)_{i=1}^n$ are independent. Then we have

$$\sup_{\|\alpha\|_2 = \|\beta\|_2 = 1} \left| \mathbf{Cov}\left[\sum_{i,j} \alpha_{i,j} X_{i,j}, \sum_{i,j} \beta_{i,j} Y_{i,j}\right] \right| \leq \rho,$$

Proof. Using the fact that a linear combination of Gaussians is a Gaussian it suffices to show that if (X_i, Y_i) are independent Gaussian vectors, each satisfying $|\mathbf{Cov}(X_i, Y_i)| \leq \rho$, $X_i, Y_i \sim N(0, 1)$ and $\|\alpha\|_2 = \|\beta\|_2 = 1$ then

$$\left| \mathbf{Cov}\left[\sum_i \alpha_i X_i, \sum_i \beta_i Y_i\right] \right| \leq \rho.$$

This follows immediately from Cauchy–Schwarz:

$$\left| \mathbf{Cov}\left[\sum_i \alpha_i X_i, \sum_i \beta_i Y_i\right] \right| = \left| \sum_i \alpha_i \beta_i \mathbf{Cov}[X_i, Y_i] \right| \leq \rho \sum_i |\alpha_i \beta_i| \leq \rho. \quad \square$$

5.2 Asymptotics of $\bar{\Gamma}(\cdot)$. In some of the applications below we will need to estimate $\bar{\Gamma}_{\rho_1, \dots, \rho_{k-1}}(\mu_1, \dots, \mu_k)$. In particular, we will need the following estimate:

LEMMA 5.4. *Let $0 < \mu < 1$ and $0 < \rho < 1$. Define*

$$B_k(\rho, \mu) = \bar{\Gamma}_{\rho_1, \dots, \rho_{k-1}}(\mu_1, \dots, \mu_k),$$

where $\rho_1 = \dots = \rho_{k-1} = \rho$ and $\mu_1 = \dots = \mu_k = \mu$. Then as $k \rightarrow \infty$ we have

$$B_k(\rho, \mu) \leq k^{(\rho^2-1)/\rho^2+o(1)}. \quad (37)$$

Proof. Clearly, we have

$$B_{i+1}(\rho, \mu) = \bar{\Gamma}_\rho(\mu, B_i). \quad (38)$$

The proof proceeds by deriving bounds on recursion (38). This is a straightforward (but not very elegant) calculation with Gaussians. Writing B_i for $B_i(\rho, \mu)$, the main two steps in verifying (37) are to show that

- The sequence B_j converges to 0 as $j \rightarrow \infty$. This follows from the fact that the functions $B \rightarrow \bar{\Gamma}_\rho(\mu, B)$ are easily seen to be strictly decreasing and have no fixed points other than $B = 0$ and $B = 1$ when $0 < \rho < 1$.
- Using Gaussian estimates sketched below, we see that for B_{j-1} sufficiently small it holds that

$$B_{j-1} - B_j \geq \frac{1}{2} B_{j-1}^{1+\frac{\rho^2}{1-\rho^2}+o(1)}. \quad (39)$$

This corresponds to B_j of the form

$$B_j \leq C j^{-\frac{1-\rho^2}{\rho^2}+o(1)}.$$

More formally, it is easy to see that if B_{j-1} is sufficiently small and satisfies

$$B_{j-1} \leq C(j-1)^{-\alpha}$$

and

$$B_j \leq B_{j-1} \left(1 - \frac{1}{2} B_{j-1}^{1/\alpha}\right)$$

for $\alpha > 0$, then

$$B_j \leq C j^{-\alpha}.$$

This follows using the fact that for small values of δ the maximum of the function $x(1 - x^{1/\alpha}/2)$ in the interval $[0, \delta]$ is obtained at δ and, therefore,

$$\begin{aligned} B_j &\leq B_{j-1} \left(1 - \frac{1}{2} B_{j-1}^{1/\alpha}\right) \leq C(j-1)^{-\alpha} \left(1 - \frac{1}{2} (C(j-1)^{-\alpha})^{1/\alpha}\right) \\ &= C(j-1)^{-\alpha} \left(1 - \frac{1}{2} C^{1/\alpha} (j-1)^{-1}\right). \end{aligned}$$

In order for

$$B_j \leq C j^{-\alpha},$$

we need that

$$\frac{1}{1 - \frac{1}{2} C^{1/\alpha} (j-1)^{-1}} \geq \left(\frac{j}{j-1}\right)^\alpha$$

which holds for large enough value of C .

In order to obtain (39) for small values of B_{j-1} , one uses the lemma stated below together with the approximation

$$\Phi(-t) = \exp\left(-\left(1 + o_t(1)\right)\frac{t^2}{2}\right).$$

which implies that, for every fixed $\epsilon > 0$,

$$\Phi(-t)\Phi\left(-\frac{\rho(t+\epsilon)}{\sqrt{1-\rho^2}}\right) = \Phi(-t)^{(1+o_t(1))(1+\frac{\rho^2}{1-\rho^2})}. \quad \square$$

LEMMA 5.5. *Let (X, Y) be a bi-variate normal with $\mathbf{Var}[X] = \mathbf{Var}[Y] = 1$ and $\mathbf{Cov}[X, Y] = \rho$. Then for all $\epsilon > 0$ and $t > 0$ it holds that*

$$\begin{aligned} \Phi(-t) - \bar{\Gamma}_\rho(1/2, \Phi(-t)) &= \mathbf{P}[X \leq -t] - \mathbf{P}[X \leq -t, Y \leq 0] = \mathbf{P}[X \leq -t, Y \geq 0] \\ &\geq \Phi(-t)(1 - \exp(-t\epsilon - \epsilon^2/2))\Phi\left(-\frac{\rho(t+\epsilon)}{\sqrt{1-\rho^2}}\right). \end{aligned}$$

Proof. The equalities follow by the definitions. For the inequality, we write

$$\begin{aligned} \mathbf{P}[X \leq -t, Y \geq 0] &\geq \mathbf{P}[X \leq -t]\mathbf{P}[Y \geq 0, X \geq -t-\epsilon \mid X \leq -t] \\ &= \mathbf{P}[X \leq -t]\mathbf{P}[X \geq -t-\epsilon \mid X \leq -t]\mathbf{P}[Y \geq 0 \mid -t \geq X \geq -t-\epsilon]. \end{aligned}$$

The bound in the lemma follows by bounding each of the three terms starting with $\mathbf{P}[X \leq -t] = \Phi(-t)$. Then note that

$$\frac{\mathbf{P}[X \leq -t-\epsilon]}{\mathbf{P}[X \leq -t]} \leq \exp(-t\epsilon - \epsilon^2/2),$$

and therefore

$$\mathbf{P}[X \geq -t-\epsilon \mid X \leq -t] \geq (1 - \exp(-t\epsilon - \epsilon^2/2)).$$

Finally, writing Z for a $N(0, 1)$ variable that is independent of X we obtain

$$\begin{aligned} \mathbf{P}[Y \geq 0 \mid -t \geq X \geq -t-\epsilon] &\geq \mathbf{P}[Y \geq 0 \mid X = -t-\epsilon] = \mathbf{P}[\rho(-t-\epsilon) + \sqrt{1-\rho^2}Z \geq 0] \\ &= \mathbf{P}\left[Z \geq \frac{\rho(t+\epsilon)}{\sqrt{1-\rho^2}}\right] = \Phi\left(-\frac{\rho(t+\epsilon)}{\sqrt{1-\rho^2}}\right), \end{aligned}$$

as needed. $\square$

6 Gaussian Bounds on Non-reversible Noise forms

In this section we prove the main results of the paper: Theorem 1.14 and its relaxations Propositions 1.15 and 1.16. As in previous work [MOO1,2], [DMR], the proof idea is to use an invariance principle, in this case Theorem 4.2, together with the Gaussian bounds of section 5.

Since the invariance principle requires working either with low-degree polynomials, or polynomials that have exponentially decaying weight, an important step of the proof is the reduction to this case. This reduction is proved in subsection 6.1. It is based on the fact that $\rho < 1$ and on the properties of correlated spaces and Efron–Stein decompositions derived in section 2.

The reduction, Theorem 4.2 and truncation arguments allow us to prove Theorem 1.14 for $k = 2$ in subsection 6.2 and for $k > 2$ in subsection 6.3.

The relaxed conditions on the influences for $k = 2$ and for r -wise independent distributions are derived in subsection 6.4 using a “two-threshold” technique. A related technique has been used before in [DS], [DMR]. However, the variant presented here is more elegant, gives more explicit dependency on the influences and allows us to exploit s -wise independence.

Finally, using a recursive argument we derive Proposition 1.15 in subsection 6.5. We don’t know of any previous application of this idea in the context of the study of influences.

6.1 Noise forms are determined by low-degree expansion. In order to use the invariance principle, it is crucial to apply it to multilinear polynomials that are either of low degree or well approximated by their low-degree part. Here we show that noise stability quantities do not change by much if one replaces a function by a slight smoothing of it. For the following statement recall Definition 1.9 for the definition of ρ and (17), (18) and (25) for the definition of the Bonami–Beckner operator $T_{1-\gamma}$.

LEMMA 6.1. *Let $\Omega_1, \dots, \Omega_n, \Lambda_1, \dots, \Lambda_n$ be a collection of finite probability spaces. Let $\mathcal{X}, \mathcal{Y}$ be two ensembles of collections of random variables such that $(\mathcal{X}_i, \mathcal{Y}_i)$ are independent and $\mathcal{X}_i$ is a basis for the functions in Ω_i , $\mathcal{Y}_i$ is a basis for the functions in Λ_i . Suppose further that for all i it holds that $\rho(\Omega_i, \Lambda_i) \leq \rho$.*

Let P and Q be two multilinear polynomials. Let $\epsilon > 0$ and γ be chosen sufficiently close to 0 so that

$$\gamma \leq 1 - (1 - \epsilon)^{\log \rho / (\log \epsilon + \log \rho)}.$$

Then,

$$|\mathbf{E}[P(\mathcal{X})Q(\mathcal{Y})] - \mathbf{E}[T_{1-\gamma}(\mathcal{X})T_{1-\gamma}Q(\mathcal{Y})]| \leq 2\epsilon \mathbf{Var}[P] \mathbf{Var}[Q].$$

In particular, there exists an absolute constant C such that it suffices to take

$$\gamma = C \frac{(1 - \rho)\epsilon}{\log(1/\epsilon)}.$$

Proof. Without loss of generality it suffices to assume that $\mathbf{Var}[P] = \mathbf{Var}[Q] = 1$ and show that

$$|\mathbf{E}[P(\mathcal{X})Q(\mathcal{Y})] - \mathbf{E}[P(\mathcal{X})(T_{1-\gamma}Q)(\mathcal{Y})]| = |\mathbf{E}[P(\mathcal{X})((I - T_{1-\gamma})Q)(\mathcal{Y})]| \leq \epsilon.$$

Let T be the Markov operator defined by $Tg(x) = \mathbf{E}[g(Y)|X = x]$, where (X, Y) are distributed according to $(\mathcal{X}, \mathcal{Y})$. In order to prove the lemma it suffices to show that

$$|\mathbf{E}[P(\mathcal{X})(T(I - T_{1-\gamma})Q)(\mathcal{X})]| \leq \epsilon. \quad (40)$$

Write P and Q in terms of their Efron–Stein decomposition, that is,

$$P = \sum_{S \subset [n]} P_S, \quad Q = \sum_{S \subset [n]} P_Q.$$

It is easy to see that

$$(I - T_{1-\gamma})Q_S = (1 - (1 - \gamma)^{|S|})Q_S,$$

and Propositions 2.11 and 2.12 imply that

$$\|TQ_S\|_2 \leq \rho^{|S|} \|Q_S\|_2,$$

and that TQ_S is orthogonal to $P_{S'}$ for $S' \neq S$. Writing $T' = T(I - T_{1-\gamma})$ we conclude that

$$\|T'Q_S\|_2 \leq \min(\rho^{|S|}, 1 - (1 - \gamma)^{|S|}) \|Q_S\|_2 \leq \epsilon \|Q_S\|_2,$$

and that $T'Q_S$ is orthogonal to P'_S when $S' \neq S$.

By Cauchy–Schwarz we get that

$$\begin{aligned} |\mathbf{E}[P(\mathcal{X})(T'Q)(\mathcal{X})]| &= \left| \sum_{S \neq \emptyset} \mathbf{E}[P_S T'Q_S] \right| \leq \sqrt{\mathbf{Var}[P]} \sqrt{\sum_{S \neq \emptyset} \|T'Q_S\|_2^2} \\ &\leq \epsilon \sqrt{\mathbf{Var}[P]} \sqrt{\mathbf{Var}[Q]}, \end{aligned}$$

as needed. $\square$

Similarly, we have

LEMMA 6.2. *Let $(\Omega_1^{(j)}, \dots, \Omega_n^{(j)})_{j=1}^k$ be k collections of finite probability spaces. Let $((\mathcal{X}_i^j)_{i=1}^n : j = 1, \dots, k)$ be k ensembles of collections of random variables such that $((\mathcal{X}_i^j)_{j=1}^k)_{i=1}^n$ are independent and $\mathcal{X}_i^j$ is a basis for the functions in $\Omega_i^{(j)}$. Suppose further that for all i it holds that $\rho(\Omega_i^{(j)} : 1 \leq j \leq k) \leq \rho$.*

Let $P_1, \dots, P_k$ be k multilinear polynomials. Let γ be chosen sufficiently close to 0 so that

$$\gamma \leq 1 - (1 - \epsilon)^{\log \rho / (\log \epsilon + \log \rho)}.$$

Then,

$$\left| \mathbf{E} \left[\prod_{j=1}^k P_j(\mathcal{X}^j) \right] - \mathbf{E} \left[\prod_{j=1}^k T_{1-\gamma} P_j(\mathcal{X}^j) \right] \right| \leq \epsilon \sum_{j=1}^k \sqrt{\mathbf{Var}[P_j]} \sqrt{\mathbf{Var} \left[\prod_{j' < j} T_{1-\gamma} P_{j'} \prod_{j' > j} P_{j'} \right]}.$$

In particular, there exists an absolute constant C such that it suffices to take

$$\gamma = C \frac{(1 - \rho)\epsilon}{\log(1/\epsilon)}.$$

Proof. The proof follows the proof of the previous lemma. $\square$

6.2 Bilinear Gaussian bounds. In this section we prove the bilinear stability bound. We repeat the statement of Theorem 1.14 with more explicit dependency on the influences.

Theorem 6.3. *Let $(\Omega_i^{(1)} \times \Omega_i^{(2)}, \mathbf{P}_i)$ be a sequence of correlated spaces such that for each i the minimum $\mathbf{P}_i$ probability of any atom in Ω_i is at least $\alpha \leq 1/2$ and such that $\rho(\Omega_i^{(1)}, \Omega_i^{(2)}; \mathbf{P}_i) \leq \rho$ for all i . Then for every $\epsilon > 0$ there exists a $\tau = \tau(\epsilon) < 1/2$ that if $f : \prod_{i=1}^n \Omega_i^{(1)} \rightarrow [0, 1]$ and $g : \prod_{i=1}^n \Omega_i^{(2)} \rightarrow [0, 1]$ satisfy*

$$\max(\text{Inf}_i^{\leq \log(1/\tau) \log(1/\alpha)}(f), \text{Inf}_i^{\leq \log(1/\tau) \log(1/\alpha)}(g)) \leq \tau \quad (41)$$

for all i (see Definition 3.7 for the definition of low-degree influence), then

$$\underline{\Gamma}_\rho(\mathbf{E}[f], \mathbf{E}[g]) - \epsilon \leq \mathbf{E}[fg] \leq \underline{\Gamma}_\rho(\mathbf{E}[f], \mathbf{E}[g]) + \epsilon. \quad (42)$$

Moreover, there exists an absolute constant C such that one may take

$$\tau = \epsilon^{\frac{C \log(1/\alpha) \log(1/\epsilon)}{\epsilon(1-\rho)}}. \quad (43)$$

Proof. Write $\mu = \mathbf{E}[f]$ and $\nu = \mathbf{E}[g]$ and $K = \log(1/\alpha)$. As discussed in section 3.2, let $\mathcal{X}$ be the sequence of ensembles such that $\mathcal{X}_i$ spans the functions on $\Omega_i = \Omega_i^{(1)} \times \Omega_i^{(2)}$, $\mathcal{X}_i$ spans functions on $\Omega_i^{(1)}$ and $\mathcal{Y}_i$ spans the functions on $\Omega_i^{(2)}$. We now express f and g as multilinear polynomials P and Q of $\mathcal{X}$ and $\mathcal{Y}$. Let $\gamma > 0$ be chosen so that

$$|\mathbf{E}[P(\mathcal{X})Q(\mathcal{Y})] - \mathbf{E}[T_{1-\gamma}(\mathcal{X})T_{1-\gamma}Q(\mathcal{Y})]| \leq \epsilon/2.$$

Note that by Lemma 6.1 it follows that we may take $\gamma = \Theta(\epsilon(1-\rho)/\log(1/\epsilon))$. Thus it suffices to prove the bound stated in the theorem for $T_{1-\gamma}P(\mathcal{X})$ and $T_{1-\gamma}Q(\mathcal{Y})$.

We use the invariance principle under hypothesis **H3**. Let $\mathcal{G}$ and $\mathcal{H}$ be two Gaussian ensembles such that for all i the covariance matrix of $\mathcal{H}_i$ and $\mathcal{G}_i$ is identical to the covariance matrix of $\mathcal{X}_i$ and $\mathcal{Y}_i$ and such that $(\mathcal{G}_i, \mathcal{H}_i)$ are independent. Clearly,

$$\mathbf{E}[T_{1-\gamma}P(\mathcal{X})T_{1-\gamma}Q(\mathcal{Y})] = \mathbf{E}[T_{1-\gamma}P(\mathcal{G})T_{1-\gamma}Q(\mathcal{H})].$$

Since $(P(\mathcal{X}), Q(\mathcal{Y}))$ takes values in $[0, 1]^2$ the same is true for

$$(\tilde{P}, \tilde{Q}) = ((T_{1-\gamma}P)(\mathcal{X}), T_{1-\gamma}Q(\mathcal{Y})).$$

In other words, $\mathbf{E}[\zeta(\tilde{P}, \tilde{Q})] = 0$, where ζ is the function in (34). Writing

$$(\bar{P}, \bar{Q}) = ((T_{1-\gamma}P)(\mathcal{G}), T_{1-\gamma}Q(\mathcal{H})),$$

we conclude from Theorem 4.2 that $\mathbf{E}[\zeta(\bar{P}, \bar{Q})] \leq \tau^{\Omega(\gamma/K)}$. That is, $\|(\bar{P}, \bar{Q}) - (P', Q')\|_2^2 \leq \tau^{\Omega(\gamma/K)}$, where P' is the function of $\bar{P}$ defined by

$$P' = \begin{cases} 0 & \text{if } \bar{P} \leq 0, \\ \bar{P} & \text{if } \bar{P} \in [0, 1], \\ 1 & \text{if } \bar{P} \geq 1. \end{cases}$$

and Q' is defined similarly. Now using Cauchy–Schwarz it is easy to see that

$$|\mathbf{E}[\bar{P}\bar{Q}] - \mathbf{E}[P'Q']| \leq \tau^{\Omega(\gamma/K)}.$$

Write $\mu' = \mathbf{E}[P']$ and $\nu' = \mathbf{E}[Q']$. Then using the Gaussian Corollary 5.1 we obtain that

$$\mathbf{E}[P'Q'] \leq \Gamma_\rho(\mu', \nu').$$

From Cauchy–Schwarz it follows that $|\mu - \mu'| \leq \tau^{\Omega(\gamma/K)}$ and similarly for ν, ν' . It is immediate to check that

$$|\Gamma_\rho(\mu, \nu) - \Gamma_{\rho'}(\mu', \nu')| \leq |\mu - \mu'| + |\nu - \nu'| \leq \tau^{\Omega(\gamma/K)}.$$

Thus we have

$$\mathbf{E}[PQ] \leq \Gamma_\rho(\mu, \nu) + \tau^{\Omega(\gamma/K)} + \epsilon/2.$$

Taking τ as in (43) yields

$$\tau^{\Omega(\gamma/K)} \leq \epsilon/2,$$

and thus we obtain the upper bound in (42). The proof of the lower bound is identical. $\square$

The following proposition completes the proof of (12).

PROPOSITION 6.4. *Let $(\prod_{j=1}^k \Omega_i^{(j)}, \mathbf{P}_i)$, $1 \leq i \leq n$, be a sequence of correlated spaces such that for each i the minimum $\mathbf{P}_i$ probability of any atom in Ω_i is at least $\alpha \leq 1/2$ and such that $\rho(\Omega_i^{(1)}, \dots, \Omega_i^{(k)}) \leq \rho < 1$ for all i .*

Then for every $\epsilon > 0$ there exists a $\tau > 0$ such that if $f_j : \prod_{i=1}^n \Omega_i^{(j)} \rightarrow [0, 1]$ for $1 \leq j \leq k$ satisfy

$$\max (\text{Inf}_i^{\leq \log(1/\tau)/\log(1/\alpha)}(f_j)) \leq \tau$$

for all i and j (see Definition 3.7 for the definition of low-degree influence), then it holds that

$$\underline{\Gamma}_\rho(\mathbf{E}[f_1], \dots, \mathbf{E}[f_k]) - \epsilon \leq \mathbf{E}\left[\prod_{i=1}^k f_i\right] \leq \overline{\Gamma}_\rho(\mathbf{E}[f_1], \dots, \mathbf{E}[f_k]) + \epsilon.$$

There exists an absolute constant C such that one may take

$$\tau = \epsilon^{C \frac{\log(1/\alpha) \log(1/\epsilon)}{\epsilon(1-\rho)}}. \quad (44)$$

The proof uses the following lemma, see e.g. [ST].

LEMMA 6.5. *Let $f_1, \dots, f_k : \Omega^n \rightarrow [0, 1]$. Then, for all j ,*

$$\text{Inf}_j\left(\prod_{i=1}^k f_i\right) \leq k \sum_{i=1}^k \text{Inf}_j(f_i).$$

Proof. The proof is based on the fact that

$$\text{Var}\left[\prod_{i=1}^k f_i\right] = \frac{1}{2} \mathbf{E}\left[\left(\prod_{i=1}^k f_i(X) - \prod_{i=1}^k f_i(Y)\right)^2\right],$$

where X and Y are independent. Now

$$\begin{aligned} \mathbf{E}\left[\left(\prod_{i=1}^k f_i(X) - \prod_{i=1}^k f_i(Y)\right)^2\right] &\leq \mathbf{E}\left[\left(\sum_{i=1}^k |f_i(X) - f_i(Y)|\right)^2\right] \\ &\leq k \sum_{i=1}^k \mathbf{E}[(f_i(X) - f_i(Y))^2] = 2k \sum_{i=1}^k \text{Inf}_i(f) \end{aligned}$$

which gives the desired result. $\square$

Proof of Proposition 6.4. The proof follows by applying Theorem 6.3 iteratively for the functions $f_1, f_1 f_2, f_1 f_2 f_3, \dots$ and using the previous lemma and the fact that $\overline{\Gamma}_\rho(\cdot)$ and $\underline{\Gamma}_\rho(\cdot)$ have Lipschitz constant 1 in each coordinate. $\square$

6.3 Multilinear bounds. Next we prove (14).

Theorem 6.6. *Let $(\prod_{j=1}^k \Omega_j^{(i)}, \mathbf{P}_i)$, $1 \leq i \leq n$, be a sequence of correlated spaces such that for each i the minimum $\mathbf{P}_i$ probability of any atom in Ω_i is at least $\alpha \leq 1/2$ and such that $\rho(\Omega_i^{(1)}, \dots, \Omega_i^{(k)}) \leq \rho < 1$ for all i . Suppose furthermore that for all i and $j \neq j'$, it holds that $\rho(\Omega_i^{(j)}, \Omega_i^{(j')}) = 0$, i.e. $(\prod_{j=1}^k \Omega_i^{(j)}, \mathbf{P}_i)$ is pairwise independent.*

Then for every $\epsilon > 0$ there exists a $\tau = \tau(\epsilon) < 1/2$ such that if $f_j : \prod_{i=1}^n \Omega_i^{(j)} \rightarrow [0, 1]$ for $1 \leq j \leq k$ satisfy

$$\max \left(\inf_i^{\leq \log(1/\tau)/\log(1/\alpha)} (f_j) \right) \leq \tau, \quad (45)$$

then,

$$\left| \mathbf{E} \left[\prod_{i=1}^k f_i \right] - \prod_{i=1}^k \mathbf{E}[f_i] \right| \leq \epsilon. \quad (46)$$

There exists an absolute constant C such that one may take

$$\tau = \epsilon^{\frac{C \log(1/\epsilon)}{\epsilon(1-\rho)}}. \quad (47)$$

Proof. Note that for all i and all γ we have that the functions f_i and $T_{1-\gamma}f_i$ are $[0, 1]$ valued functions. Therefore, as in the previous proof we obtain by Lemma 6.2 that

$$\left| \mathbf{E} \left[\prod_{i=1}^k f_i \right] - \mathbf{E} \left[\prod_{i=1}^k T_{1-\gamma}f_i \right] \right| \leq \epsilon/4.$$

for $\gamma = \Omega(\epsilon(1-\rho)/\log(1/\epsilon))$. Thus it suffices to prove the bound stated in the theorem for the functions $T_{1-\gamma}f_i$.

We now use the invariance principle. Recall that the f_i may be written as a multilinear polynomial P_i of an ensemble $\mathcal{X}_i$. Let $\mathcal{G}_i$, $1 \leq i \leq k$, denote Gaussian ensembles with the same covariances as $\mathcal{X}_i$, $1 \leq i \leq k$, and let $(g_1, \dots, g_k)$ be the multilinear polynomials $(P_1, \dots, P_k)$ applied to $(\mathcal{X}_1, \dots, \mathcal{X}_k)$. Let $h_i = f_{[0,1]}(T_{1-\gamma}g_i)$. By the invariance principle Corollary 4.3, we have

$$\left| \mathbf{E} \left[\prod_{i=1}^k T_{1-\gamma}f_i \right] - \mathbf{E} \left[\prod_{i=1}^k h_i \right] \right| \leq \epsilon/4.$$

Note that h_i are functions of ensembles of Gaussian random variables such that each pair of ensembles is independent. Therefore, the h_i 's are independent which implies in turn that

$$\mathbf{E} \left[\prod_{i=1}^k h_i \right] = \prod_{i=1}^k \mathbf{E}[h_i].$$

Corollary 4.3 also implies that

$$|\mathbf{E}[h_i] - \mathbf{E}[f_i]| = |\mathbf{E}[h_i] - \mathbf{E}[T_{1-\gamma}f_i]| \leq \epsilon/4k,$$

so

$$\left| \prod_{i=1}^k \mathbf{E}[h_i] - \prod_{i=1}^k \mathbf{E}[f_i] \right| \leq \epsilon/4,$$

which concludes the proof. $\square$

6.4 Relaxed influence conditions. In this subsection we will relax the conditions imposed on the influence, i.e. Proposition 1.15. In particular we will show that in Theorem 1.14 and $k = 2$ it suffices to assume that for each coordinate at least one of the functions has low influence. Similarly for $k > 2$ and s -wise independent

distributions it suffices to have that in each coordinate at most s of the functions have large influence.

LEMMA 6.7. Assume μ is an r -wise independent distribution on Ω^k . Let $f_1, \dots, f_k : \Omega^n \rightarrow [0, 1]$. Let $S \subset [n]$ be a set of coordinates such that for each $i \in S$ at most r of the functions f_j have $\text{Inf}_i(f_j) > \epsilon$. Define

$$g_i(x) = \mathbf{E}[f_i(Y) \mid Y_{[n] \setminus S} = x_{[n] \setminus S}].$$

Then the functions g_i do not depend on the coordinates in S and

$$\left| \mathbf{E} \left[\prod_{i=1}^k f_i \right] - \mathbf{E} \left[\prod_{i=1}^k g_i \right] \right| \leq k|S|\sqrt{\epsilon}.$$

Proof. Recall that averaging over a subset of the variables preserves expected value. It also maintains the property of taking values in $[0, 1]$ and decreases influences. Thus it suffices to prove the claim for the case where $|S| = 1$. The general case then follows by induction.

So assume without loss of generality that $S = \{1\}$ consists of the first coordinate only and that $f_{r+1}, \dots, f_k$ all have $\text{Inf}_1(f_j) \leq \epsilon$, so that $\mathbf{E}[(f_j - g_j)^2] \leq \epsilon$ for $j > r$. Then by Cauchy–Schwarz we have $\mathbf{E}[|f_j - g_j|] \leq \sqrt{\epsilon}$ for $j > r$ and using the fact that the functions are bounded in $[0, 1]$ we obtain

$$\left| \mathbf{E} \left[\prod_{i=1}^k f_i \right] - \mathbf{E} \left[\prod_{i=1}^r f_i \prod_{i=r+1}^k g_i \right] \right| \leq \mathbf{E} \left[\left| \prod_{i=r+1}^k f_i - \prod_{i=r+1}^k g_i \right| \right] \leq \sum_{i=r+1}^k \mathbf{E}[|f_i - g_i|] \leq k\sqrt{\epsilon}. \quad (48)$$

Let us write $\mathbf{E}_1$ for the expected value with respect to the first variable. Recalling that the g_i do not depend on the first variable and that the f_i are r -wise independent we obtain that

$$\mathbf{E}_1 \left[\prod_{i=1}^r f_i \prod_{i=r+1}^k g_i \right] = \prod_{i=r+1}^k g_i \mathbf{E}_1 \left[\prod_{i=1}^r f_i \right] = \prod_{i=r+1}^k g_i \prod_{i=1}^r \mathbf{E}_1[f_i] = \prod_{i=1}^r g_i \prod_{i=r+1}^k g_i = \prod_{i=1}^k g_i.$$

This implies that

$$\mathbf{E} \left[\prod_{i=1}^r f_i \prod_{i=r+1}^k g_i \right] = \mathbf{E} \left[\prod_{i=1}^k g_i \right], \quad (49)$$

and the proof follows from (48) and (49). $\square$

We now prove that condition (15) suffices instead of (11).

LEMMA 6.8. Theorem 6.3 holds with the condition

$$\max_i \left(\min \left(\text{Inf}_i^{\leq \log^2(1/\tau)/\log(1/\alpha)}(f_1), \text{Inf}_i^{\leq \log^2(1/\tau)/\log(1/\alpha)}(f_2) \right) \right) \leq \tau \quad (50)$$

instead of (41).

Proof. Let

$$\gamma = C_1 \frac{(1 - \rho)\epsilon}{\log(1/\epsilon)},$$

$$\tau = \epsilon^{C_2 \frac{\log(1/\alpha)\log(1/\epsilon)}{(1-\rho)\epsilon}},$$

and

$$R = \frac{\log(1/\tau)}{\log(1/\alpha)}.$$

From the proof of Theorem 6.3 it follows that the constant C_1 and C_2 may be chosen so that

$$(1 - \gamma)^R \leq \epsilon/4, \\ |\mathbf{E}[fg] - \mathbf{E}[T_{1-\gamma}fT_{1-\gamma}g]| \leq \epsilon/4.$$

Moreover the conclusion of Theorem 6.3 holds with error at most $\epsilon/4$ for any pair of functions $\tilde{f}$ and $\tilde{g}$ if *all* influences of $\tilde{f}$ and $\tilde{g}$ satisfy $\text{Inf}_i^{\leq R} \leq \tau$. Let

$$R' = \frac{\log^2(1/\tau)}{\log(1/\alpha)},$$

and choose C_3 large enough so that

$$\tau' = \epsilon^{C_3 \frac{\log(1/\alpha) \log(1/\epsilon)}{(1-\rho)\epsilon}},$$

satisfy

$$\frac{2R}{\tau} \sqrt{\tau' + (1 - \gamma)^{2R'}} \leq \frac{\epsilon}{4}.$$

Assume that f and g satisfy

$$\max_i (\min(\text{Inf}_i^{\leq R'}(f), \text{Inf}_i^{\leq R'}(g))) \leq \tau'.$$

We will show that the statement of the theorem holds for f and g . For this let

$$S_f = \{i : \text{Inf}_i^{\leq R}(f) > \tau\}, \quad S_g = \{i : \text{Inf}_i^{\leq R}(g) > \tau\}.$$

Let $S = S_f \cup S_g$. Since $R' \geq R$ and $\tau' \leq \tau$, the sets S_f and S_g are disjoint. Moreover, both S_f and S_g are of size at most R/τ . Also, if $i \in S$ and $\text{Inf}_i^{\leq R}(f) > \tau$ then $\text{Inf}_i^{\leq R'}(g) < \tau'$ and therefore $\text{Inf}_i(g) \leq \tau' + (1 - \gamma)^{2R'}$. In other words, for all $i \in S$, we have $\min(\text{Inf}_i(g), \text{Inf}_i(f)) \leq \tau' + (1 - \gamma)^{2R'}$.

Letting $S = S_f \cup S_g$ and applying Lemma 6.7 with

$$g'(y') = \mathbf{E}[g(Y) \mid Y_{[n] \setminus S} = y'_{[n] \setminus S}], \quad f'(x') = \mathbf{E}[f(X) \mid X_{[n] \setminus S} = x'_{[n] \setminus S}],$$

we obtain that

$$|\mathbf{E}[f(x)g(y)] - \mathbf{E}[f'(x')g'(y')]| \leq \frac{2R}{\tau} \sqrt{\tau' + (1 - \gamma)^{2R'}} \leq \frac{\epsilon}{4}. \quad (51)$$

Note that the functions f' and g' satisfy that $\max(\text{Inf}_i^{\leq R}(f'), \text{Inf}_i^{\leq R}(g')) \leq \tau$ for all i . This implies that the results of Theorem 6.3 hold for f' and g' . This together with (51) implies the desired result. $\square$

The proof of the relaxed condition on influences in Theorem 6.6 is similar.

LEMMA 6.9. Assume the setup of Theorem 6.6 where $(\prod_{j=1}^k \Omega_i^{(j)}, \mathbf{P}_i)$ is s -wise independent for all i . Then the conclusion of the theorem holds when the following condition:

$$\forall i, |\{j : \text{Inf}_i^{\leq \log^2(1/\tau)/\log(1/\alpha)}(f_j) > \tau\}| \leq s \quad (52)$$

replaces condition (45).

Proof. Again we start by looking at $T_{1-\gamma}f_i$ where $\gamma = \Omega(\epsilon(1-\rho)/\log(1/\epsilon))$. We let τ' and R' be chosen so that

$$\frac{kR}{\tau} \sqrt{\tau' + (1-\gamma)^{2R'}} \leq \frac{\epsilon}{4}.$$

The set S will consist of all coordinates j where at least one of the functions f_i has $\text{Inf}_j^{\leq R}(f_i) > \tau$. The rest of the proof is similar. $\square$

6.5 A recursive argument. Here we show how Proposition 1.16 follows from Theorem 1.14. The proof uses the following lemma.

LEMMA 6.10. *Let $(\Omega_1, \mu_1), \dots, (\Omega_n, \mu_n)$ be finite probability spaces such that for all i the minimum probability of any atom in Ω_i is at least $\alpha \leq 1/2$. Let $f : \prod_{i=1}^n \Omega_i \rightarrow [0, 1]$ and suppose that $\text{Inf}_i(f) \geq \tau$. Let $S = \{i\}$. Then*

$$\mathbf{E}[\bar{f}^S] \geq \mathbf{E}[f] + \alpha\tau, \quad \mathbf{E}[\underline{f}^S] \leq \mathbf{E}[f] - \alpha\tau.$$

Proof. Note that if $g : \Omega_i \rightarrow [0, 1]$ satisfies $\mathbf{Var}[g] \geq a$ then

$$\left(\max_x g(x) - \mathbf{E}[g]\right)^2 \geq a\alpha, \quad \left(\min_x g(x) - \mathbf{E}[g]\right)^2 \geq a\alpha.$$

Therefore,

$$\mathbf{E}[\bar{f}^S - f] \geq \mathbf{E}[(\bar{f}^S - f)^2] \geq \alpha \text{Inf}_i(f) = \alpha\tau,$$

which implies the first inequality. The second inequality is proved similarly. $\square$

We now prove Proposition 1.16.

Proof. The set T is defined recursively as follows. We let $T_o = \emptyset$. $\underline{a}_t = \bar{a}_t = 0$. Then we repeat the following: If all of the functions

$$\bar{f}_1^{T_t}, \dots, \bar{f}_k^{T_t}, \underline{f}_1^{T_t}, \dots, \underline{f}_k^{T_t}$$

have influences lower than τ , then we halt and let $T = T_t$. Otherwise, there exists at least one i and one j such that either $\text{Inf}_i(\bar{f}_j^{T_t}) \geq \tau$ or $\text{Inf}_i(\underline{f}_j^{T_t}) \geq \tau$. We then let $T_{t+1} = T_t \cup \{i\}$. In the first case we let $\bar{a}_{t+1} = \bar{a}_t + 1$, $\underline{a}_{t+1} = \underline{a}_t$. In the second case we let $\bar{a}_{t+1} = \bar{a}_t$, $\underline{a}_{t+1} = \underline{a}_t + 1$.

Note that by Lemma 6.10, this process must terminate within

$$\frac{2k}{\alpha\tau}$$

steps since

$$k \geq \mathbf{E}\left[\sum_{j=1}^k \bar{f}_j^{T_t}\right] \geq \alpha\tau \bar{a}_t, \quad 0 \leq \mathbf{E}\left[\sum_{j=1}^k \underline{f}_j^{T_t}\right] \leq k - \alpha\tau \underline{a}_t.$$

and $\bar{a}_t + \underline{a}_t \geq t$. $\square$

7 Applications to Social Choice

In this section we apply Theorem 6.2 to the two social choice models.

7.1 ρ for samples of votes. In the first social choice example we consider Example 2.6. The correlated probability spaces are the ones given by

$$\Omega_V = \{\{x = 1\}, \{x = -1\}\},$$

representing the intended vote and

$$\Omega_S = \{\{(x = 1, y = 1)\}, \{(x = -1, y = 1)\}, \{y = 0\}\}$$

representing the sampled status.

In order to calculate $\rho(\Omega_1, \Omega_2)$ it suffices by Lemma 2.8 to calculate $\sqrt{\mathbf{E}[(Tf)^2]}$ where $f(x, y) = x$ is the (only) Ω_V measurable with $\mathbf{E}[f] = 0$ and $\mathbf{E}[f^2] = 1$. We see that $Tf(x, y) = 0$ if $y = 0$ and $Tf(x, y) = x$ when $y \neq 0$. Therefore

$$\sqrt{\mathbf{E}[(Tf)^2]} = \rho^{1/2}.$$

LEMMA 7.1.

$$\rho(\Omega_V, \Omega_S) = \rho^{1/2}.$$

7.2 Predictability of binary vote. Here we prove Theorem 1.1.

Proof. The proof follows directly from Proposition 1.15 and Lemma 7.1 as

$$\Gamma_{\sqrt{\rho}}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \sqrt{\rho}. \quad \square$$

7.3 ρ in Condorcet voting. In the context of Condorcet voting, Ω is given by S_k , the group of permutations on k elements and μ is the uniform measure. We write $\binom{[k]}{2}$ for the collection of subsets of $[k]$ of size 2.

DEFINITION 7.2. Let $Q \subset \binom{[k]}{2}$ and define $R_Q : \Omega \rightarrow \{0, 1\}^Q$ be letting $(R_Q(\sigma))_{i < j} = 1$ if $\sigma(i) < \sigma(j)$ and $(R_Q(\sigma))_{i < j} = 0$ if $\sigma(i) > \sigma(j)$. R_Q summarizes the pairwise relations in the permutation σ for pairs in Q .

Given a subset $Q \subset \binom{[k]}{2}$ we define

$$\Omega_Q = \{\{\sigma : R(\sigma) = x\} : x \in \{0, 1\}^Q\}.$$

Thus Ω_Q is the coarsening of Ω summarizing the information about pairwise relations in Q .

We will mostly be interested in $\rho(\Omega_Q, \Omega_{i < j})$ where $(i < j) \notin Q$.

LEMMA 7.3. Suppose $Q = \{(1 > 2), (1 > 3), \dots, (1 > r)\}$ then

$$\rho(\Omega_Q, \Omega_{1 > (1+r)}) = \sqrt{\frac{r-1}{3(r+1)}} \leq \frac{1}{\sqrt{3}}.$$

Proof. We use Lemma 2.8 again. The space $\Omega_{1 > (r+1)}$ has a unique function with $\mathbf{E}[f] = 0$ and $\mathbf{E}[f^2] = 1$. This is the function that satisfies $f(\sigma) = 1$ when $\sigma(1) > \sigma(r+1)$ and $f(\sigma) = -1$ if $\sigma(1) < \sigma(r+1)$.

The conditional probability that $\sigma(1) > \sigma(r+1)$ given that s of the inequalities $\sigma(1) > \sigma(2), \dots, \sigma(1) > \sigma(r)$ hold is

$$\frac{s+1}{r+1}.$$

Therefore the conditional expectation of f under this conditioning is

$$\frac{(s+1) - (r-s)}{r+1} = \frac{2s+1-r}{r+1}.$$

Noting that the number of inequalities satisfied is uniform in the range $\{0, \dots, r-1\}$ we see that

$$\begin{aligned} \mathbf{E}[(Tf)^2] &= \frac{1}{r(r+1)^2} \sum_{s=0}^{r-1} (2s+1-r)^2 = \frac{1}{r(r+1)^2} \left(4 \sum_{s=0}^{r-1} s^2 - 4(r-1) \sum_{s=0}^{r-1} s + r(r-1)^2 \right) \\ &= \frac{1}{r(r+1)^2} \left(4 \frac{2(r-1)^3 + 3(r-1)^2 + (r-1)}{6} \right. \\ &\quad \left. - 4(r-1) \frac{(r-1)^2 + (r-1)}{2} + r(r-1)^2 \right) \\ &= \frac{r-1}{3(r+1)}. \end{aligned}$$

Therefore,

$$\rho = \sqrt{\frac{r-1}{3(r+1)}}. \quad \square$$

7.4 Condorcet paradox. We now prove Theorem 1.5 dealing with Condorcet paradoxes.

Proof. We wish to bound the asymptotic probability in terms of the number of candidates k for the probability that there is a unique maximum in Condorcet aggregation. Clearly this probability equals k times the probability that candidate 1 is the maximum.

Recall that the votes are aggregated as follows. Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be an anti-symmetric function with $\mathbf{E}[f] = 0$ and low influences. Let $\sigma_1, \dots, \sigma_n \in S_k$ denote the n rankings of the individual voters. Recall that we denote $x^{a>b}(i) = 1$ if $\sigma_i(a) > \sigma_i(b)$ and $x^{a<b}(i) = -1$ if $\sigma_i(a) < \sigma_i(b)$. Note that $x^{b>a} = -x^{a>b}$. We recall further that the binary decision between each pair of coordinates is performed via a anti-symmetric function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ so that $f(-x) = -f(x)$ for all $x \in \{-1, 1\}^n$. Finally, the tournament $G_K = G_K(\sigma; f)$ is defined by having $(a > b) \in G_K$ if and only if $f(x^{a>b}) = 1$.

In order to obtain an upper bound on the probability that 1 is the unique maximum, define $f^{a,b}(\sigma_1, \dots, \sigma_n) = (1 + f(x^{a>b}))/2$. Then the probability that 1 is the maximum is given by

$$\mathbf{E} \left[\prod_{a=2}^k f^{1,a} \right].$$

Using (12) we obtain that

$$\mathbf{E} \left[\prod_{a=2}^k f^{1,a} \right] \leq \bar{\Gamma}_{\left(\frac{1}{\sqrt{3}}, \dots, \frac{1}{\sqrt{3}}\right)} \left(\frac{1}{2}, \dots, \frac{1}{2} \right) + \epsilon,$$

where $\epsilon \rightarrow 0$ as $\tau \rightarrow 0$; there are $k-1$ expected values all given by $1/2$ as $\mathbf{E}[f^{1,a}] = 1/2$ for all a ; the $k-2$ values of ρ all bounded by $1/\sqrt{3}$ by Lemma 7.3.

By Lemma 5.4 we now obtain

$$\mathbf{E}\left[\prod_{a=2}^k f^{1,a}\right] \leq k^{-2+o(1)}.$$

Taking the union bound on the k possible maximal values we obtain that in Condorcet voting the probability that there is a unique max is at most $k^{-1+o(1)}$ as needed. $\square$

7.5 Majority and tightness. The majority function shows the tightness of Theorem 1.1 and Theorem 1.5.

7.5.1 Tightness for prediction. For the tightness of Theorem 1.1, write

$$X = \frac{\sum_{i=1}^n x_i}{\sqrt{n}}, \quad Y = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\rho n}},$$

where x_i is the intended vote of voter i and $y_i = 1$ if voter i was queried, $y_i = 0$ otherwise. X is the total bias of the actual vote and Y the bias of the sample.

Note that (X, Y) is asymptotically a normal vector with covariance matrix

$$\begin{pmatrix} 1 & \sqrt{\rho} \\ \sqrt{\rho} & 1 \end{pmatrix}.$$

Therefore asymptotically we obtain

$$\mathbf{E}[\operatorname{sgn}(X)\operatorname{sgn}(Y)] = 2\mathbf{P}[X > 0, Y > 0] - 1 = 2\Gamma_\rho\left(\frac{1}{2}, \frac{1}{2}\right) - 1 = \frac{2}{\pi} \arcsin \sqrt{\rho}.$$

7.5.2 Tightness in Condorcet voting. The tightness in Theorem 1.5 follows from [RiR] and [NW]. We briefly sketch the main steps of the proof.

For each a and b let

$$X^{a>b} = \frac{\sum_{i=1}^n x_i^{a>b}}{\sqrt{n}},$$

be the bias preference in a majority vote towards a . By the CLT, all the random variables $X^{a>b}$ are asymptotically $N(0, 1)$.

Consider the random variables $X^{1>2}, \dots, X^{1>k}$. Note that this set of variables is exchangeable (they are identically distributed under any permutation of their order). Moreover,

$$\mathbf{E}[X^{1>2}X^{1>3}] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[x_i^{1>2}x_i^{1>3}] = \frac{1}{3}.$$

By the CLT the limiting value (as n to ∞) of

$$\lim_{n \rightarrow \infty} \mathbf{P}[X^{1>2} > 0, \dots, X^{1>k} > 0] = \mathbf{P}[N^{1>2} > 0, \dots, N^{1>k} > 0]$$

where $(N^{1>a})$ is an exchangeable collection of normal $N(0, 1)$ random variables, the correlation between each pair of which is $1/3$. The results of [RiR] imply that, as $k \rightarrow \infty$,

$$\mathbf{P}[N^{1>2} > 0, \dots, N^{1>k} > 0] \sim 2\Gamma(2)\sqrt{2\pi}\frac{\sqrt{\ln k}}{k^2},$$

This in turn implies that the probability of a unique max for majority voting for

large k as $n \rightarrow \infty$ is given by

$$(2 + o(1))\Gamma(2)\sqrt{2\pi}\frac{\sqrt{\ln k}}{k},$$

showing the tightness of the result up to sub-polynomial terms.

7.5.3 The probability that majority will result in linear order. Here we prove Proposition 1.6 and show that the probability that majority will result in a linear order is $\exp(-\Theta(k^{5/3}))$. We find this asymptotic behavior quite surprising. Indeed, given the previous results that the probability that there is a unique max is $k^{-1+o(1)}$, one may expect that the probability that the order is linear would be

$$k^{-1+o(1)}(k-1)^{-1+o(1)} \dots = (k!)^{-1+o(1)}.$$

However, it turns out that there is a strong negative correlation between the event that there is a unique maximum among the k candidates and that among the other candidates there is a unique max.

Proof. We use the multi-dimensional CLT. Let

$$X_{a>b} = \frac{1}{\sqrt{n}}(|\{\sigma : \sigma(a) > \sigma(b)\}| - |\{\sigma : \sigma(b) > \sigma(a)\}|).$$

By the CLT at the limit the collection of variables $(X_{a>b})_{a \neq b}$ converges to a joint Gaussian vector $(N_{a>b})_{a \neq b}$ satisfying, for all distinct a, b, c, d ,

$$N_{a>b} = -N_{b>a}, \quad \mathbf{Cov}[N_{a>b}, N_{a>c}] = \frac{1}{3}, \quad \mathbf{Cov}[N_{a>b}, N_{c>d}] = 0.$$

and $N_{a>b} \sim N(0, 1)$ for all a and b .

We are interested in providing bounds on

$$\mathbf{P}[\forall a > b : N_{a>b} > 0]$$

as the probability that the resulting tournament is an order is obtained by multiplying by a $k! = \exp(\Theta(k \log k))$ factor.

We claim that there exist independent $N(0, 1)$ random variables X_a for $1 \leq a \leq k$ and $Z_{a>b}$ for $1 \leq a \neq b \leq k$ such that

$$N_{a>b} = \frac{1}{\sqrt{3}}(X_a - X_b + Z_{a>b})$$

(where $Z_{a>b} = -Z_{b>a}$). This follows from the fact that the joint distribution of Gaussian random variables is determined by the covariance matrix (this is noted in the literature in [NW]).

We now prove the upper bound. Let α be a constant to be chosen later. Note that for all α and large enough k it holds that

$$\mathbf{P}[|X_a| > k^\alpha] \leq \exp(-\Omega(k^{2\alpha})).$$

Therefore the probability that for at least half of the a 's in the interval $[k/2, k]$ it holds that $|X_a| > k^\alpha$ is at most

$$\exp(-\Theta(k^{1+2\alpha})).$$

Let's assume that at least half of the a 's in the interval $[k/2, k]$ satisfy that $|X_a| < k^\alpha$. We claim that in this case the number $H_{k/4}[-k^\alpha, k^\alpha]$ of pairs $a > b$ such that $X_a, X_b \in [-k^\alpha, k^\alpha]$ and $X_a - X_b < 1$ is $\Omega(k^{2-\alpha})$.

For the last claim, partition the interval $[-k^\alpha, k^\alpha]$ into sub-intervals of length 1 and note that at least $\Omega(k)$ of the points belong to sub-intervals which contain at least $\Omega(k^{1-\alpha})$ points. This implies that the number of pairs $a > b$ satisfying $|X_a - X_b| < 1$ is $\Omega(k^{2-\alpha})$.

Note that for such pair $a > b$ in order that $N_{a>b} > 0$ we need that $Z_{a>b} > -1$ which happens with constant probability.

We conclude that given that half of the X 's fall in $[-k^\alpha, k^\alpha]$ the probability of a linear order is bounded by

$$\exp(-\Omega(k^{2-\alpha})).$$

Thus overall we have bounded the probability by

$$\exp(-\Omega(k^{1+2\alpha})) + \exp(-\Omega(k^{2-\alpha})).$$

The optimal exponent is $\alpha = 1/3$ giving the desired upper bound.

For the lower bound we condition on X_a taking value in $(a, a+1)k^{-2/3}$. Each probability is at least $\exp(-O(k^{2/3}))$ and therefore the probability that all X_a take such values is

$$\exp(-O(k^{5/3})).$$

Moreover, conditioned on X_a taking such values the probability that

$$Z_{a>b} > X_b - X_a,$$

for all $a > b$ is at least

$$\left(\prod_{i=0}^{k-1} \Phi(i)^{k^{2/3}} \right)^k \geq \left(\prod_{i=0}^{\infty} \Phi(i) \right)^{k^{5/3}} = \exp(-O(k^{5/3})).$$

This proves the required result. $\square$

8 Applications to Hypergraphs and Additive Properties

In this section we prove Theorem 1.7 and give a few examples. The basic idea in the applications presented so far was that in order to bound correlation between k events of low influences, it suffices to know how to bound the correlation between the first $k-1$ and the last one. For low influence events, using the invariance principle, one obtains bounds coming from half spaces in Gaussian space, or majority functions in the discrete space.

The applications presented now will be of a different nature. We will be interested again in correlation between k events – however, we will restrict ourselves to correlation measures defined in such a way that any pair of events are un-correlated. While this is a much more restrictive setting, it allows us to obtain exact results and not just bounds. In other words, we obtain that such correlation measures for low-influence events depend only on the measure of the sets but not on any additional structure. While this may sound surprising, it in fact directly follows the invariance principle together with the fact that for jointly Gaussian random variables, pairwise independence implies independence. We first prove Theorem 1.7.

Proof. The proof follows immediately from Theorem 1.14, Proposition 1.16 and Lemma 2.9. $\square$

EXAMPLE 8.1. Consider the group $\mathbb{Z}_m$ for $m > 2$. We will be interested in linear relations over $\mathbb{Z}_m^k$. A *linear relation* over $\mathbb{Z}_m^k$ is given by $L = (L_0, \ell_1, \dots, \ell_k)$ such that $\ell_i \neq 0$ for all $i \geq 1$ and ℓ_i and m are co-prime for all $i \geq 1$. We will restrict ourselves to the case $k \geq 3$. We will write $L(x)$ to denote the logical statement that $\sum_{i=1}^k x_i \ell_i \bmod m \in L_0$. Given a set $A \subset \mathbb{Z}_m^n$ we will denote

$$L(A^k) = |\{x \in A^k : L(x)\}|,$$

and μ_L the uniform measure on $L(A^k)$. We note that for every linear relation we have that μ_L is pairwise smooth and that if the set L_0 is of size at least 2 then R is connected.

We now apply Theorem 1.7 to conclude that for low influence sets $A \subset \mathbb{Z}_m^n$, the number of k tuples $(x_1, \dots, x_k) \in A^k$ satisfying $\sum x_i \bmod m \in L_0^n$ is

$$(1 \pm o(1)) |A|^k \left(\frac{|L_0|}{m} \right)^n. \quad (53)$$

For general sets A we conclude that we have $\underline{A}^S \subset A \subset \overline{A}^T$ where $|S| = O(1)$, $T = O(1)$ and (53) holds for both $\underline{A}^S$ and $\overline{A}^T$.

EXAMPLE 8.2. We may consider much more general relations. For example we may take P be a polynomial in $x_1, \dots, x_r \in \mathbb{Z}_r$ and Q to be a polynomial in $x_{r+1}, \dots, x_k$ such that P and Q both have roots. Then we can look at the relation defined by the zeros of PQ . It is easy to check that R is connected. Therefore it follows that if $A \subset \mathbb{Z}_m^n$ is set all of whose influences are lower than τ then

$$c_1 \left(\tau, \frac{|A|}{|\mathbb{Z}_m^n|} \right) |R^n| \leq |A^k \cap R^n| \leq c_2 \left(\tau, \frac{|A|}{|\mathbb{Z}_m^n|} \right) |R^n|$$

where c_1 and c_2 are two positive functions. Again if A is not of low influences then there exist finite sets of coordinates S and T such that $\underline{A}^S \subset A \subset \overline{A}^T$ and the conclusion holds for S and T .

References

- [AM1] P. AUSTIN, E. MOSSEL, Approximation resistant predicates from pairwise independence, in “23rd Annual IEEE Conference on Computational Complexity”, Los Alamitos, CA, USA, IEEE Computer Society (2008), 249–258.
- [AM2] P. AUSTIN, E. MOSSEL, Approximation resistant predicates from pairwise independence, Computational Complexity, to appear.
- [B] W. BECKNER, Inequalities in Fourier analysis, Ann. of Math. (2) 10:1 (1975), 159–182.
- [Be] C.E. BELL, A random voting graph almost surely has a hamiltonian cycle when the number of alternatives is large, Econometrica 49:6 (1981), 1597–1603.
- [Bo] A. BONAMI, Étude des coefficients de Fourier des fonctions de $L^p(G)$, Ann. Inst. Fourier (Grenoble), 20:2 (1971), 335–402.
- [Bor] C. BORELL, Geometric bounds on the Ornstein–Uhlenbeck velocity process, Z. Wahrsch. Verw. Gebiete 70:1 (1984), 1–13.

- [DMR] I. DINUR, E. MOSSEL, O. REGEV, Conditional hardness for approximate coloring, Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing (STOC 2006) (2006), 344–353.
- [DS] I. DINUR, S. SAFRA, The importance of being biased, Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing, ACM, New York (2002), 33–42 (electronic).
- [ES] B. EFRON, C. STEIN, The jackknife estimate of variance, *Annals of Statistics* 9:3 (1996), 586–596.
- [FK] A. FRIEZE, R. KANNAN, Quick approximation to matrices and applications, *Combinatorica* 19:2 (1999), 175–220.
- [G] H. GEBELEIN, Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung, *Z. Angew. Math. Mech.* 21 (1941), 364–379.
- [Gr] B. GREEN, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.* 15:2 (2005), 340–376.
- [J] S. JANSON, Gaussian Hilbert Spaces, *Cambridge Tracts in Mathematics* 129, Cambridge University Press, Cambridge (1997).
- [KKL] J. KAHN, G. KALAI, N. LINIAL, The influence of variables on boolean functions, Proceedings of the 29th Annual Symposium on Foundations of Computer Science (1988), 68–80.
- [Ka1] G. KALAI, A Fourier-theoretic perspective on the Condorcet paradox and Arrow’s theorem, *Adv. in Appl. Math.* 29:3 (2002), 412–426.
- [Ka2] G. KALAI, Social indeterminacy, *Econometrica* 72 (2004), 1565–1581.
- [Kh] S. KHOT, On the power of unique 2-prover 1-round games, Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing, ACM, New York (2002), 767–775 (electronic).
- [KhKMO1] S. KHOT, G. KINDLER, E. MOSSEL, R. O’DONNELL, Optimal inapproximability results for max-cut and other 2-variable csps? Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (2004), 146–154.
- [KhKMO2] S. KHOT, G. KINDLER, E. MOSSEL, R. O’DONNELL, Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput.* 37 (2007), 319–357.
- [L] J.W. LINDBERG, Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Math. Zeit.* 15 (1922), 211–235.
- [MOO1] E. MOSSEL, R. O’DONNELL, K. OLESZKIEWICZ, Noise stability of functions with low influences: invariance and optimality (extended abstract), Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23–25 October 2005, Pittsburgh, PA, USA, IEEE Computer Society (2005), 21–30.
- [MOO2] E. MOSSEL, R. O’DONNELL, K. OLESZKIEWICZ, Noise stability of functions with low influences: invariance and optimality, *Ann. Math.*, to appear.
- [NW] R.G. NIEMI, H.F. WEISBERG, A mathematical solution for the probability of paradox of voting, *Behavioral Science* 13 (1968), 317–323.
- [O] R. O’DONNELL, Hardness amplification within NP, Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing, ACM, New York (2002), 751–760 (electronic).
- [Ol] K. OLESZKIEWICZ, On a nonsymmetric version of the Khinchine–Kahane inequality, *Progress in Probability* 56 (2003), 156–168.

- [R] P. RAGHAVENDRA, Optimal algorithms and inapproximability results for every CSP? Proceedings of the 40th Annual Symposium on Theory of Computing (STOC) (2008), 245–254.
- [Ré] A. RÉNYI, On measures of dependence, Acta Math. Acad. Sci. Hungar. 10 (1959), 441–451.
- [RiR] Y. RINOTT, V. ROTAR'. A remark on quadrant normal probabilities in high dimensions, Statistics and Probability Letters 51:1 (2001), 47–51.
- [Ro] V.I. ROTAR', Limit theorems for multilinear forms and quasipolynomial functions, Teor. Veroyatnost. i Primenen. 20:3 (1975), 527–546.
- [ST] A. SAMORODNITSKY, L. TREVISAN, Gowers uniformity, influence of variables, and PCPs, STOC '06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing, ACM Press, New York, NY, USA (2006), 11–20.
- [Sh] W. SHEPPARD, On the application of the theory of error to cases of normal distribution and normal correlation, Phil. Trans. Royal Soc. London, 192 (1899), 101–168.
- [T] T. TAO, Szemerédi's regularity lemma revisited, Contrib. Discrete Math. 1:1 (2006), 8–28 (electronic).
- [W] P. WOLFF, Hypercontractivity of simple random variables, Studia Math. 180:2 (2007), 219–236.

ELCHANAN MOSSEL, Faculty of Mathematics and Computer Science, Weizmann Institute, Rehovot Israel 76100

and

Statistics and Computer Science Department, University of California at Berkeley, USA

`mossel@stat.berkeley.edu`

Received: December 18, 2008

Accepted: March 31, 2009

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.