Probability Theory and Related Fields (2022) 183:649–675
https://doi.org/10.1007/s00440-022-01120-5

Quantitative correlation inequalities via extremal power series

Anindya De[1] · Shivam Nadimpalli[2] · Rocco A. Servedio[2]

Received: 6 April 2021 / Revised: 3 February 2022 / Accepted: 11 February 2022 /
Published online: 14 March 2022

Abstract

Many correlation inequalities for high-dimensional functions in the literature, such as the Harris–Kleitman inequality, the Fortuin–Kasteleyn–Ginibre inequality and the celebrated Gaussian Correlation Inequality of Royen, are *qualitative* statements which establish that any two functions of a certain type have non-negative correlation. Previous work has used Markov semigroup arguments to obtain quantitative extensions of some of these correlation inequalities. In this work, we augment this approach with a new extremal bound on power series, proved using tools from complex analysis, to obtain a range of new and near-optimal quantitative correlation inequalities. These new results include: A quantitative version of Royen's celebrated Gaussian Correlation Inequality (Royen, 2014). In (Royen, 2014) Royen confirmed a conjecture, open for 40 years, stating that any two symmetric convex sets must be non-negatively correlated under any centered Gaussian distribution. We give a lower bound on the correlation in terms of the vector of degree-2 Hermite coefficients of the two convex sets, conceptually similar to Talagrand's quantitative correlation bound for monotone Boolean functions over $\{0, 1\}^n$ (Talagrand in Combinatorica 16(2):243–258, 1996). We show that our quantitative version of Royen's theorem is within a logarithmic factor of being optimal. A quantitative version of the well-known FKG inequality for monotone functions over any finite product probability space. This is a broad generalization of Talagrand's quantitative correlation bound for functions from $\{0, 1\}^n$ to $\{0, 1\}$ under the uniform distribution (Talagrand in Combinatorica 16(2):243–258, 1996). In the special case of p-biased distributions over $\{0, 1\}^n$ that was considered

✉ Shivam Nadimpalli
sn2855@columbia.edu

Anindya De
anindyad@seas.upenn.edu

Rocco A. Servedio
ras2105@columbia.edu

1 University of Pennsylvania, Philadelphia, USA

2 Columbia University, New York, USA

by Keller, our new bound essentially saves a factor of $p\log(1/p)$ over the quantitative bounds given in Keller (Eur J Comb 33:1943–1957, 2012; Improved FKG inequality for product measures on the discrete cube, 2008; Influences of variables on Boolean functions. PhD thesis, Hebrew University of Jerusalem, 2009).

Mathematics Subject Classification 52 (Convex and discrete geometry) · 60 (Probability theory and stochastic processes)

1 Introduction

Correlation inequalities are theorems stating that for certain classes of functions and certain probability distributions $\mathcal{D}$, any two functions f, g in the class must be non-negatively correlated with each other under $\mathcal{D}$, i.e. it must be the case that $\mathbf{E}_{\mathcal{D}}[fg] - \mathbf{E}_{\mathcal{D}}[f]\,\mathbf{E}_{\mathcal{D}}[g] \geq 0$. Inequalities of this type have a long history, going back at least to a well-known result of Chebyshev, "Chebyshev's order inequality," which states that for any two nondecreasing sequences $a_1 \leq \cdots \leq a_n$, $b_1 \leq \cdots \leq b_n$ and any probability distribution p over $[n] = \{1, \ldots, n\}$, it holds that

$$\sum_{i=1}^{n} a_i b_i p_i \geq \left(\sum_{i=1}^{n} a_i p_i\right)\left(\sum_{i=1}^{n} b_i p_i\right).$$

Modern correlation inequalities typically deal with high-dimensional rather than one-dimensional functions. Results of this sort have proved to be of fundamental interest in many fields such as combinatorics, analysis of Boolean functions, statistical physics, and beyond.

Perhaps the simplest high-dimensional correlation inequality is the well known Harris–Kleitman theorem [14, 21], which states that if $f, g : \{0,1\}^n \to \{0,1\}$ are monotone functions (meaning that $f(x) \leq f(y)$ whenever $x_i \leq y_i$ for all i) then $\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] \geq 0$, where expectations are with respect to the uniform distribution over $\{0,1\}^n$. The Harris-Kleitman theorem has a one-paragraph proof by induction on n; on the other end of the spectrum is the Gaussian Correlation Inequality (GCI), which states that if $K, L \subseteq \mathbb{R}^n$ are any two symmetric convex sets and $\mathcal{D}$ is any centered Gaussian distribution over $\mathbb{R}^n$, then $\mathbf{E}_{\mathcal{D}}[KL] - \mathbf{E}_{\mathcal{D}}[K]\,\mathbf{E}_{\mathcal{D}}[L] \geq 0$ (where we identify sets with their 0/1-valued indicator functions). This was a famous conjecture for four decades before it was proved by Thomas Royen in 2014 [29]. Other well-known correlation inequalities include the Fortuin–Kasteleyn–Ginibre (FKG) inequality [11], which is an important tool in statistical mechanics and probabilistic combinatorics; the Griffiths–Kelly–Sherman (GKS) inequality [13, 24], which is a correlation inequality for ferromagnetic spin systems; and various generalizations of the GKS inequality to quantum spin systems [12].

1.1 Quantitative correlation inequalities

The agenda of the current work is to obtain *quantitative* correlation inequalities. Consider the following representative example: For two monotone Boolean functions $f, g : \{0,1\}^n \to \{0,1\}$, as discussed above, the Harris-Kleitman theorem states that

$\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] \geq 0$. It is easy to check that the Harris–Kleitman inequality is tight if and only if f and g depend on disjoint sets of variables. One might therefore hope to get an improved bound by measuring how much f and g depend simultaneously on the same coordinates. Such a bound was obtained by Talagrand [31] in an influential paper (appropriately titled "How much are increasing sets correlated?"). To explain Talagrand's main result, we recall the standard notion of *influence* from Boolean function analysis [28]. For a Boolean function $f : \{0,1\}^n \to \{0,1\}$, the *influence* of coordinate i on f is defined to be $\mathbf{Inf}_i[f] := \Pr_{\boldsymbol{x}\sim U_n}[f(\boldsymbol{x}) \neq f(\boldsymbol{x}^{\oplus i})]$, where U_n is the uniform distribution on $\{0,1\}^n$ and $\boldsymbol{x}^{\oplus i}$ is obtained by flipping the ith bit of $\boldsymbol{x}$. Talagrand proved the following *quantitative* version of the Harris–Kleitman inequality:

$$\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] \geq \frac{1}{C} \cdot \Psi\left(\sum_{i=1}^{n} \mathbf{Inf}_i[f]\mathbf{Inf}_i[g]\right) \tag{1}$$

where $\Psi(x) := x/\log(e/x)$, $C > 0$ is an absolute constant, and the expectations are with respect to the uniform measure. A simple corollary of this result is that $\mathbf{E}[fg] = \mathbf{E}[f]\,\mathbf{E}[g]$ if and only if the sets of influential variables for f and g are disjoint. In [31] itself, Talagrand gives an example for which Eq. (1) is tight up to constant factors.

Talagrand's result has proven to be influential in the theory of Boolean functions, and several works [17–20] have obtained extensions and variants of this inequality for product distributions over $\{0,1\}^n$. Keller et al. [23] proved the following alternative strengthening of the Harris–Kleitman inequality:

$$\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] \geq \frac{1}{C} \cdot \sum_{i=1}^{n} \Upsilon(\mathbf{Inf}_i[f])\Upsilon(\mathbf{Inf}_i[g]) \tag{2}$$

where $\Upsilon(x) := x/\sqrt{\log(e/x)}$, $C > 0$ is an absolute constant, and the expectations are with respect to the uniform measure. They also obtained analogues of Eqs. (1) and (2) in the setting of monotone functions over $\mathbb{R}^n$ using a new notion of "geometric influence" for functions over Gaussian space [22, 23].[1] The paper [23] crucially relies on semigroup methods such as reverse hypercontractivity and integration by parts. In fact, Mossel [27] (see Remark 5.5) showed that the semigroup method can be extended to yield analogues of Eqs. (1) and (2) for monotone functions over any poset where an appropriate semigroup can be defined; the approach in [27] follows the work of Cordero-Erausquin and Ledoux [6] who used semigroup methods to prove Talagrand's generalization of the KKL inequality. Finally, Eldan [9] recently gave a new approach towards obtaining Eqs. (1) and (2) via stochastic calculus, and gave a refinement of Eq. (1) in terms of the degree-2 Fourier coefficients of the functions.

[1] We note that both Eqs. (1) and (2) are sharp, and neither implies the other; see Sect. 4.1 of [23] or the discussion following Theorem 1.4 in [20].

1.2 Our contributions

In this paper, we establish a general framework to obtain a range of new quantitative correlation inequalities. Similar to [23, 27], our approach also uses the semigroup framework. However, unlike the previous papers, the key fact exploited here is that for the semigroups of interest in this paper, the corresponding Laplacian has integer eigenvalues. This is useful for us because it means that the action of the so-called "noise operator" can be expressed as a power series in terms of the "noise rate". We then use a new extremal bound on power series, proved using elementary tools from complex analysis to obtain several new quantitative correlation inequalities including:

1. Quantitative versions of Royen's Gaussian Correlation Inequality and Hu's correlation inequality [16] for symmetric convex functions over Gaussian space. (We also give a lower bound example which shows that our quantitative version of the Gaussian Correlation Inequality is within a logarithmic factor of the best possible bound.)
2. A quantitative FKG inequality for a broad class of product distributions, including arbitrary product distributions over finite domains.

These results are obtained in a unified fashion via simple proofs that we view as quite different from previous works [17–20, 23, 31]. In particular, unlike several of these earlier papers, our proofs do not use semigroup tools such as reverse hypercontractivity or integration by parts.

We note that the special case of the second item above with the uniform distribution on $\{0, 1\}^n$ essentially recovers Talagrand's correlation inequality [31]. In more detail, our bound is weaker than that obtained in [31] by a logarithmic factor, but our proof is significantly simpler and easily generalizes to other domains. For p-biased distributions over $\{0, 1\}^n$, our bound avoids any dependence on p compared to the results of Keller [17–19] which have a $p \log(1/p)$ dependence (though, similar to the situation vis-a-vis [31], we lose a logarithmic factor in other dependencies). Finally, we note that our framework allows us to obtain two seemingly incomparable quantitative versions of the FKG inequality for monotone functions over the solid cube $[0, 1]^n$ endowed with the uniform measure; we refer the interested reader to the preprint version of this paper for more details of these last two results.

1.3 The approach

We start with a high level meta-observation before explaining our framework and techniques in detail. While the statements of the Harris–Kleitman inequality, the FKG inequality, and the Gaussian Correlation Inequality have a common flavor, the proofs of these results are extremely different from each other. As noted earlier, the Harris–Kleitman inequality admits a simple inductive proof which is only a few lines long: Given two monotone Boolean functions $f, g : \{0, 1\}^n \to \{0, 1\}$, we can write

$$\mathbf{E}[f \cdot g] - \mathbf{E}[f] \cdot \mathbf{E}[g] = \mathbf{E}\left[\frac{f_1 - f_0}{2}\right] \cdot \mathbf{E}\left[\frac{g_1 - g_0}{2}\right] + \sum_{i \in \{0,1\}} \left(\frac{\mathbf{E}[f_i \cdot g_i] - \mathbf{E}[f_i] \cdot \mathbf{E}[g_i]}{2}\right)$$

where $f_i(x) := f(i, x_2, \ldots, x_n)$ and $g_i(x) := g(i, x_2, \ldots, x_n)$ for $i \in \{0, 1\}$. Monotonicity of f and g ensures that the first term is non-negative, whereas the inductive hypothesis ensures that each summand in the second term is non-negative as well.

In contrast, the Gaussian Correlation Inequality was an open problem for nearly four decades, and no inductive proof for it is known. Thus, at first glance, it is not clear how one might come up with a common framework to obtain quantitative versions of these varied qualitative inequalities. Our approach circumvents this difficulty by using the qualitative inequalities essentially as “black boxes.” This allows us to extend the qualitative inequalities into quantitative ones while essentially sidestepping the difficulties of proving the initial qualitative statements themselves.

1.3.1 Our general framework

In this subsection we give an overview of our general framework and the high-level ideas underlying it, with our quantitative version of the Gaussian Correlation Inequality serving as a running example throughout for concreteness.

We begin by defining a function $\Phi : [0, 1] \to [0, 1]$ which will play an important role in our results:

$$\Phi(x) := \min\left\{x, \frac{x}{\log^2(1/x)}\right\}. \tag{3}$$

(Note the similarity between Φ and the function Ψ mentioned earlier that arose in Talagrand’s quantitative correlation inequality [31]; the difference is that Φ is smaller by essentially a logarithmic factor in the small-x regime.)

Let $\mathcal{F}$ be a family of real-valued functions on some domain (endowed with measure μ) with $\mathbf{E}_{\mu}\left[f^2\right] \le 1$ for all $f \in \mathcal{F}$. For example, the Gaussian Correlation Inequality is a correlation inequality for the family $\mathcal{F}_{\mathrm{csc}}$ of centrally symmetric, convex sets (identified with their 0/1-indicator functions), and μ is the standard Gaussian measure $\mathcal{N}(0, 1)^n$, usually denoted γ.[2] A *quantitative* correlation inequality for $f, g \in \mathcal{F}$ gives a (non-negative) lower bound on the quantity $\mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})g(\boldsymbol{x})] - \mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})]\,\mathbf{E}_{\boldsymbol{y}\sim\mu}[g(\boldsymbol{y})]$, typically in terms of some measure of “how much f and g simultaneously depend on the same coordinates.” Our general approach establishes such a quantitative inequality in two main steps:

Step 1 For this step, we require an appropriate family of “noise operators” $(\mathrm{T}_\rho)_{\rho\in[0,1]}$ with respect to the measure μ. Very briefly, each of these operators T_ρ will be a (re-indexed version of a) symmetric Markov operator whose stationary distribution is μ;

[2] Since convexity is preserved under linear transformation, no loss of generality is incurred in assuming that the background measure is the standard normal distribution $\mathcal{N}(0, 1)^n$ rather than an arbitrary centered Gaussian.

this is defined more precisely in Sect. 4. (Looking ahead, we will see, for example, that in the case of the GCI, the appropriate noise operator is the Ornstein-Uhlenbeck noise operator, defined in Definition 15.) The crucial property we require of the family $(\mathrm{T}_\rho)_{\rho\in[0,1]}$ with respect to $\mathcal{F}$ is what we refer to as *monotone compatibility*:

Definition 1 (*Monotone compatibility*). A class of functions $\mathcal{F}$ and background measure μ is said to be *monotone compatible* with respect to a family of noise operators $(\mathrm{T}_\rho)_{\rho\in[0,1]}$ if (i) for all $f, g \in \mathcal{F}$, the function

$$q(\rho) := \mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})\mathrm{T}_\rho g(\boldsymbol{x})]$$

is a non-decreasing function of ρ, and (ii) for $\rho = 1$ we have $\mathrm{T}_1 = \mathrm{Id}$ (the identity operator).

The notion of monotone compatibility should be seen as a mild extension of qualitative correlation inequalities. As an example, in the case of the Gaussian Correlation Inequality, Royen's proof [29] in fact shows that that the family $\mathcal{F}_{\mathrm{csc}}$ is monotone compatible with Ornstein-Uhlenbeck operators.

Step 2 We express the operator T_ρ in terms of its eigenfunctions. In all the cases we consider in this paper, the eigenvalues of the operator T_ρ are $\left\{\rho^j\right\}_{j\geq 0}$. Let $\{\mathcal{W}_j\}_{j\geq 0}$ be the corresponding eigenspaces. Consequently, we can express $q(\rho) - q(0)$ as

$$\begin{aligned} q(\rho) - q(0) &= \mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})\mathrm{T}_\rho g(\boldsymbol{x})] - \mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})] \cdot \mathbf{E}_{\boldsymbol{y}\sim\mu}[g(\boldsymbol{y})] \\ &= \sum_{j>0} \rho^j\, \mathbf{E}[f_j(\boldsymbol{x})g_j(\boldsymbol{x})], \end{aligned} \tag{4}$$

where f_j (respectively g_j) is the projection of f (respectively g) on the space $\mathcal{W}_j$. To go back to our running example, for the Gaussian Correlation Inequality, $\mathcal{W}_j$ is the subspace spanned by degree-j Hermite polynomials on $\mathbb{R}^n$.

Define $a_j := \mathbf{E}[f_j(\boldsymbol{x})g_j(\boldsymbol{x})]$, so $q(\rho) = \sum_{j\geq 0} a_j\rho^j$. Now, corresponding to any family $\mathcal{F}$ and noise operators $(\mathrm{T}_\rho)_{\rho\in[0,1]}$, there will be a unique $j^* \in \mathbb{N}$ such that the following properties hold:

1. If $a_{j^*} = 0$, then $\mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})g(\boldsymbol{x})] = \mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})] \cdot \mathbf{E}_{\boldsymbol{y}\sim\mu}[g(\boldsymbol{y})]$. In other words, a_{j^*} *qualitatively* captures the "slack" in the correlation inequality. For example, for the Gaussian Correlation Inequality, it turns out that $j^* = 2$ (and over the Boolean hypercube $\{-1, 1\}^n$, it turns out that $j^* = 1$).
2. For any i such that j^* does not divide i, $a_i = 0$.

Now, from the fact that the spaces $\{\mathcal{W}_j\}$ are orthonormal and the fact that every $f \in \mathcal{F}$ has $\mathbf{E}_\mu\left[f^2\right] \leq 1$, it follows that $\sum_{j>0} |a_j| \leq 1$. Our main technical lemma, Lemma 11, implies (see the proof of Theorem 13) that for any such power series $q(\cdot)$, there exists some $\rho^* \in [0, 1]$ such that[3]

$$q(\rho^*) - q(0) \geq \frac{1}{C} \cdot \Phi(a_{j^*}).$$

[3] Looking ahead (see Eq. (9)), it will be immediate from the monotone compatibility of $\mathcal{F}$ with T_ρ that $a_{j^*} \geq 0$.

The proof crucially uses tools from complex analysis. As the class $\mathcal{F}$ is monotone compatible with the operators $(\mathrm{T}_\rho)_{\rho\in[0,1]}$, recalling Eq. (4), it follows that

$$q(1) - q(0) = \mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})g(\boldsymbol{x})] - \mathbf{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})] \cdot \mathbf{E}_{\boldsymbol{y}\sim\mu}[g(\boldsymbol{y})] \geq \frac{1}{C} \cdot \Phi(a_{j^*}), \tag{5}$$

which is the desired quantitative correlation inequality for $\mathcal{F}$.

Remark 2 We emphasize the generality of our framework; the argument sketched above can be carried out in a range of different concrete settings. For example, by using the Harris-Kleitman qualitative correlation inequality for monotone Boolean functions in place of the GCI, and the Bonami-Beckner noise operator over $\{0, 1\}^n$ in place of the Ornstein-Uhlenbeck noise operator, the above arguments give a simple proof of the following (slightly weaker) version of Talagrand's correlation inequality (Eq. (1)):

$$\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] \geq \frac{1}{C} \cdot \Phi\left(\sum_{i=1}^{n} \mathbf{Inf}_i[f]\mathbf{Inf}_i[g]\right), \tag{6}$$

for an absolute constant $C > 0$. While our bound is weaker than that of [31] by a log factor (recall the difference between Ψ and Φ), our methods are applicable to a wider range of settings (such as the GCI and the other applications given in this paper). Finally, we emphasize that our proof strategy is really quite different from that of [31]; for example, [31]'s proof relies crucially on bounding the degree-2 Fourier weight of monotone Boolean functions by the degree-1 Fourier weight, whereas our strategy does not analyze the degree-2 spectrum of monotone Boolean functions at all.

1.4 Organization

The rest of the paper is organized as follows: Sect. 2 recalls the necessary background on Markov semigroups and functional analysis, and recalls a well-known result from complex analysis that we will require to prove our main lemma. In Sect. 3, we prove our main technical lemma, Lemma 11, which is at the heart of our approach to quantitative correlation inequalities. Section 4 presents our general approach to quantitative correlation inequalities, Theorem 13, which we instantiate with concrete examples in subsequent sections as follows: In Sect. 5, we obtain quantitative analogues of several correlation inequalities over Gaussian space; in particular, we give robust forms of Royen's Gaussian Correlation Inequality (GCI) [29], present an extension of the quantitative GCI to quasiconcave functions, and also obtain a robust form of Hu's correlation inequality for convex functions [16]. In Sect. 6, we obtain an analogue of Talagrand's correlation inequality [31] in the setting of monotone functions over finite product spaces. We note that this setting includes the Boolean hypercube $\{0, 1\}^n$, wherein we obtain a generalization of Talagrand's inequality to real-valued functions and p-biased distributions.

2 Preliminaries

In this section we give preliminaries setting notation, recalling useful background on noise operators and orthogonal decomposition of functions over product spaces, and recalling a well-known result that we will require from complex analysis.

2.1 Noise operators and orthogonal decompositions

Let (Ω, π) be a probability space; we do not require Ω to be finite, and we assume without loss of generality that π has full support.

The background we require for noise operators on functions in $L^2(\Omega, \pi)$ is most naturally given using the language of "Markov semigroups." Our exposition below will be self-contained; for a general and extensive resource on Markov semigroups, we refer the interested reader to [5].

Definition 3 (*Markov semigroup*). A collection of linear operators $(\mathrm{P}_t)_{t\geq 0}$ on $L^2(\Omega, \pi)$ is said to be a *Markov semigroup* if

1. $\mathrm{P}_0 = \mathrm{Id}$;
2. for all $s, t \in [0, \infty)$, we have $\mathrm{P}_s \circ \mathrm{P}_t = \mathrm{P}_{s+t}$; and
3. for all $t \in [0, \infty)$ and all $f, g \in L^2(\Omega, \pi)$, the following hold:
   (a) *Identity*: $\mathrm{P}_t 1 = 1$ where 1 is the identically-1 function.
   (b) *Positivity*: $\mathrm{P}_t f \geq 0$ almost everywhere if $f \geq 0$ almost everywhere.[4]

It is well known that a Markov semigroup can be constructed from a Markov process and vice versa [5]. We call a Markov semigroup *symmetric* if the underlying Markov process is time-reversible; the following definition is an alternative elementary characterization of symmetric Markov semigroups. (Recall that for $f, g \in L^2(\Omega, \pi)$ the inner product $\langle f, g\rangle$ is defined as $\mathbf{E}_{\boldsymbol{x}\sim\pi}[f(\boldsymbol{x})g(\boldsymbol{x})]$.)

Definition 4 (*Symmetric Markov semigroup*). A Markov semigroup $(\mathrm{P}_t)_{t\geq 0}$ on $L^2(\Omega, \pi)$ is *symmetric* if for all $t \in [0, \infty)$, the operator P_t is self-adjoint; equivalently, for all $t \in [0, \infty)$ and all $f, g \in L^2(\Omega, \pi)$, we have $\langle f, \mathrm{P}_t g\rangle = \langle \mathrm{P}_t f, g\rangle$.

We note that the families of noise operators $(\mathrm{U}_\rho)_{\rho\in[0,1]}$ and $(\mathrm{T}_\rho)_{\rho\in[0,1]}$ that we consider in Sects. 5 and 6 respectively will be parametrized by $\rho \in [0, 1]$ where $\rho = e^{-t}$ for $t \in [0, \infty)$, as is standard in theoretical computer science. (For example, the Bonami-Beckner noise operator operator T_ρ mentioned in the Introduction, which is a special case of the T_ρ operator defined in Sect. 6, corresponds to P_t for $(\mathrm{P}_t)_{t\geq 0}$ a suitable Markov semigroup and $\rho = e^{-t}$.)

Given a Markov semigroup $(\mathrm{P}_t)_{t\geq 0}$ on the probability space (Ω, π), we can naturally define the Markov semigroup $\left(\otimes_{i=1}^n \mathrm{P}_{t_i}\right)_{t_i\geq 0}$ on $L^2\left(\Omega^n, \pi^{\otimes n}\right)$; we will often abuse notation and denote this operator on $L^2\left(\Omega^n, \pi^{\otimes n}\right)$ by simply P_t. We next define a decomposition of $L^2\left(\Omega^n, \pi^{\otimes n}\right)$ that is particularly well-suited to the action of a Markov semigroup $(\mathrm{P}_t)_{t\geq 0}$.

[4] Note that this implies the following *order* property: if $f \geq g$ almost everywhere, then $\mathrm{P}_t f \geq \mathrm{P}_t g$ almost everywhere.

Definition 5 (*Chaos decomposition*). Consider a Markov semigroup $(\mathrm{P}_t)_{t\geq 0}$ on $L^2(\Omega^n, \pi^{\otimes n})$. We call an orthogonal decomposition of

$$L^2(\Omega^n, \pi^{\otimes n}) = \bigoplus_{i=0}^{\infty} \mathcal{W}_i$$

a *chaos decomposition* with respect to the Markov semigroup $(\mathrm{P}_t)_{t\geq 0}$ if

1. $\mathcal{W}_0 = \text{span}\{1\}$ where 1 is the identically-1 function (i.e. $\mathcal{W}_0 = \mathbb{R}$).
2. For all $t \geq 0$, there exists $\lambda_t \in [0, 1]$ such that if $f \in \mathcal{W}_i$, then $\mathrm{P}_t f = \lambda_t^i f$.
3. If $t_1 > t_2$, then $\lambda_{t_1} < \lambda_{t_2}$.

The term "chaos decomposition" is used in the literature to describe the spectral decomposition of $L^2(\mathbb{R}^n, \gamma)$ with respect to the Laplacian of the Ornstein–Uhlenbeck semigroup (see Fact 16); its usage in the broader sense defined above is not standard (to our knowledge).

Remark 6 The semigroup property (Item 2 of Definition 3) together with strict monotonicity (Item 3 of Definition 5) together imply that in fact $\lambda_t = \lambda_*^t$ for some $\lambda_* \in (0, 1)$. (This follows immediately from Cauchy's functional equation.) In other words, Item 2 in Definition 5 can be restated as follows: There exists $\lambda_* \in (0, 1)$ such that for all $t \geq 0$, if $f \in \mathcal{W}_i$ then we have $\mathrm{P}_t f = \lambda_*^{it} f$.

Example 7 To provide intuition for Definition 5, note that

- Over Gaussian space, a natural chaos decomposition with respect to the Ornstein–Uhlenbeck semigroup is given by the basis of Hermite polynomials (Sect. 5.1);
- For real-valued functions over finite product domains, a natural chaos decomposition with respect to the Bonami–Beckner semigroup is given by the Efron–Stein decomposition (Sect. 6.1).

Notation 8 Given an orthogonal decomposition $L^2(\Omega^n, \pi^{\otimes n}) = \bigoplus_i \mathcal{W}_i$, for $f \in L^2(\Omega^n, \pi^{\otimes n})$ we will write $f = \oplus_i f_i$ where f_i is the projection of f onto $\mathcal{W}_i$.

We note that $\lambda_0 = 1$, and as $1 \in \mathcal{W}_0$, it follows that $f_0 = \langle f, 1\rangle$. We revisit the definition of monotone compatibility given in the introduction in the language of Markov semigroups:

Definition 9 (*Monotone compatibility*). Let $(\mathrm{P}_t)_{t\geq 0}$ be a Markov semigroup on $L^2(\Omega^n, \pi^{\otimes n})$. We say that $(\mathrm{P}_t)_{t\geq 0}$ is *monotone compatible* with a family of functions $\mathcal{F} \subseteq L^2(\Omega^n, \pi^{\otimes n})$ if for all $f, g \in \mathcal{F}$, we have

$$\frac{\partial}{\partial t}\langle \mathrm{P}_t f, g\rangle \leq 0$$

when the above derivative exists.

Recalling that our noise operators such as $(\mathrm{T}_\rho)_{\rho\in[0,1]}$ are reparameterized versions of the Markov semigroup operators $(\mathrm{P}_t)_{t\geq 0}$ under the reparameterization $\mathrm{T}_\rho = \mathrm{P}_t$ with $\rho = e^{-t}$, and recalling item 1 in Definition 3, we see that Definition 9 coincides with Definition 1.

2.2 Complex analysis

Let $U \subseteq \mathbb{C}$ be a connected, open set. Recall that a function $f : U \to \mathbb{C}$ is said to be *holomorphic* if at every point in U it is complex differentiable in a neighborhood of the point. For U a connected closed set, f is said to be holomorphic if it is holomorphic in an open set containing U. Our main technical lemma appeals to the following classical result, a proof of which can be found in [30].

Theorem 10 (Hadamard Three Circles Theorem). *Suppose f is holomorphic on the annulus $\{z \in \mathbb{C} \mid r_1 \le |z| \le r_2\}$. For $r \in [r_1, r_2]$, let $M(r) := \max_{|z|=r} |f(z)|$. Then*

$$\log\left(\frac{r_2}{r_1}\right) \log M(r) \le \log\left(\frac{r_2}{r}\right) \log M(r_1) + \log\left(\frac{r}{r_1}\right) \log M(r_2).$$

3 A new extremal bound for power series with bounded length

Given a complex power series $p(t) = \sum_{i=1}^{\infty} c_i t^i$ where $c_i \in \mathbb{C}$, its *length* is defined to be the sum of the absolute values of its coefficients, i.e. $\sum_{i=1}^{\infty} |c_i|$. Our main technical lemma is a lower bound on the sup-norm of complex power series with no constant term and bounded length:[5]

Lemma 11 (Main Technical Lemma). *Let $p(t) = \sum_{i=1}^{\infty} c_i t^i$ with $c_1 = 1$ and $\sum_{i=1}^{\infty} |c_i| \le M$ where $M \ge 3/2$. Then:*

$$\sup_{t \in [0,1]} |p(t)| \ge \frac{\Theta(1)}{\log^2 M}.$$

The proof given below is inspired by arguments with a similar flavor in [2, 3], where the Hadamard Three Circles Theorem is used to prove various extremal bounds on polynomials.

Proof Consider the meromorphic map (easily seen to have a single pole at $z = 0$) given by

$$h(z) = A\left(z + \frac{1}{z}\right) + B,$$

which maps origin-centered circles to ellipses centered at B. Let $0 < \delta < c$ be a parameter that we will fix later, where $0 < c < 1$ is an absolute constant that will be specified later. We impose the following constraints on A and B:

$$-2A + B = \delta \qquad \frac{17}{4}A + B = 1,$$

[5] The "3/2" in the lemma below could be replaced by any constant bounded above 1; we use 3/2 because it is convenient in our later application of Lemma 11.

and note that these constraints imply that $A = \frac{4(1-\delta)}{25}$ and $B = \frac{8+17\delta}{25}$.

We define three circles in the complex plane that we will use for the Hadamard Three Circles Theorem:

(i) Let C_1 be the circle centered at 0 with radius 1. Note that for all $z \in C_1$, the value $h(z)$ is a real number in the interval $\left[\delta, \frac{16+9\delta}{25}\right] \subseteq [\delta, 1)$.
(ii) Let $r > 1$ be such that $h(-r) = 0$, so $r + \frac{1}{r} = \frac{8+17\delta}{4-4\delta} = 2 + \Theta(\delta)$ and hence $r = 1 + \Theta(\sqrt{\delta})$, which is less than 4. Define C_2 to be the circle centered at 0 with radius r.
(iii) Let C_3 be the circle centered at 0 with radius 4. Note that $|h(z)| \leq 1$ for $z \in C_3$.

Define $q(t) := \frac{p(t)}{t}$. Note that $q(0) = c_1 = 1$ and that for all $z \in \mathbb{C}$ such that $|z| \leq 1$, we have $|q(z)| \leq M$. Define $\psi(z) := q(h(z))$. Note that ψ is holomorphic on $\{z \in \mathbb{C} : |z| \leq 4\} \setminus \{0\}$; in particular, it is holomorphic on the (closed) annulus defined by C_1 and C_3. Consequently, by Theorem 10, we have:

$$\log\left(\frac{4}{1}\right)\log\alpha(r) \leq \log\left(\frac{4}{r}\right)\log\alpha(1) + \log\left(\frac{r}{1}\right)\log\alpha(4)$$

with $\alpha(r) := \sup_{|z|=r} |\psi(z)|$. As $h(-r) = 0$, we have $\psi(-r) = 1$ and so $\log\alpha(r) \geq 0$. Consequently, the left hand side of the above inequality is non-negative, which implies:

$$1 \leq \alpha(1)^{\log\left(\frac{4}{r}\right)} \cdot \alpha(4)^{\log r}.$$

As $\log\left(\frac{4}{r}\right) = \Theta(1)$, $\log r = \log\left(1 + \Theta\left(\sqrt{\delta}\right)\right) = \Theta\left(\sqrt{\delta}\right)$, and $\alpha(4) \leq M$, we get:

$$1 \leq \alpha(1)^{\Theta(1)} \cdot M^{\Theta\left(\sqrt{\delta}\right)}, \qquad \text{and hence} \qquad M^{-\Theta\left(\sqrt{\delta}\right)} \leq \alpha(1).$$

By (i) and the definition of α, we have:

$$\sup_{t\in[\delta,1)} q(t) \geq M^{-\Theta\left(\sqrt{\delta}\right)} \qquad \text{and hence} \qquad \sup_{t\in[0,1]} p(t) \geq \sup_{\delta\in[0,1]} \delta M^{-\Theta\left(\sqrt{\delta}\right)}.$$

Setting $\delta = \frac{\Theta(1)}{\log^2 M}$, we get that

$$\sup_{t\in[0,1]} |p(t)| \geq \frac{\Theta(1)}{\log^2 M},$$

and the lemma is proved. □

It is natural to wonder whether Lemma 11 is quantitatively tight. The polynomial $p(t) = t(1-t)^{\log M}$ is easily seen to have length M and $\sup_{t\in[0,1]} p(t) = \Theta(1/\log M)$, and it is tempting to conjecture that this might be the smallest achievable value. However, it turns out that the $1/\log^2 M$ dependence of Lemma 11 is in fact the best possible result, as shown by the following claim which we prove in Appendix A:

Claim 12 *For sufficiently large M, there exists a real polynomial $p(t) = \sum_{i=1}^{d} c_i t^i$ with $c_1 = 1$ and $\sum_{i=1}^{d} |c_i| \leq M$ such that*

$$\sup_{t \in [0,1]} p(t) \leq O\left(\left(\frac{1}{\log M}\right)^2\right).$$

4 A general approach to quantitative correlation inequalities

This section presents our general approach to obtaining *quantitative* correlation inequalities from *qualitative* correlation inequalities. While our main result, Theorem 13, is stated in an abstract setting, subsequent sections will instantiate this result in concrete settings that provided the initial impetus for this work. Section 5 deals with the setting of centrally symmetric, convex sets over Gaussian space, and Sect. 6 deals with finite product domains.

Theorem 13 (Main Theorem). *Consider a symmetric Markov semigroup $(\mathrm{P}_t)_{t\geq 0}$ on $L^2\left(\Omega^n, \Pi^{\otimes n}\right)$ with a chaos decomposition*

$$L^2\left(\Omega^n, \Pi^{\otimes n}\right) = \bigoplus_{\ell} \mathcal{W}_\ell.$$

Let $(\mathrm{P}_t)_{t\geq 0}$ be monotone compatible with $\mathcal{F} \subseteq L^2\left(\Omega^n, \Pi^{\otimes n}\right)$, where $\|f\| \leq 1$ for all $f \in \mathcal{F}$. Furthermore, suppose that there exists $j^* \in \mathbb{N}_{>0}$ such that every $f \in \mathcal{F}$ has a decomposition as

$$f = \bigoplus_{\ell=0}^{\infty} f_{\ell \cdot j^*},$$

i.e. $f_\ell = 0$ for $j^* \nmid \ell$. Then for all $f, g \in \mathcal{F}$, we have

$$\langle f, g\rangle - f_0 g_0 \geq \frac{1}{C} \cdot \Phi\left(\langle f_{j^*}, g_{j^*}\rangle\right), \tag{7}$$

where recall from Eq. (3) that $\Phi : [0,1] \to [0,1]$ is $\Phi(x) = \min\left\{x, \frac{x}{\log^2(1/x)}\right\}$ and $C > 0$ is a universal constant.

The proof of the above theorem uses an interpolating argument along the Markov semigroup, and appeals to Lemma 11 to obtain the lower bound.

Proof of Theorem 13 Fix $f, g \in \mathcal{F}$ and let us write $a_\ell := \langle f_\ell, g_\ell\rangle$. It follows from Definition 5 that f_ℓ, g_ℓ are eigenfunctions of P_t with eigenvalue λ_t^ℓ. This, together with the assumption that $f = \oplus_{j^*|\ell} f_\ell$ and $g = \oplus_{j^*|\ell} g_\ell$, implies that for $t > 0$ we have

$$\langle \mathrm{P}_t f, g\rangle = \sum_{j^*|\ell} \lambda_t^\ell \langle f_\ell, g_\ell\rangle = \sum_{j^*|\ell} a_\ell \lambda_t^\ell. \tag{8}$$

Here we remark that the argument to $\Phi(\cdot)$ in the right hand side of Eq. (7) is non-negative, i.e. $a_{j^*} \geq 0$. To see this, observe that

$$a_{j^*} = \frac{\partial}{\partial \lambda_t^{j^*}} \langle \mathrm{P}_t f, g \rangle = \frac{\partial}{\partial t} \langle \mathrm{P}_t f, g \rangle \cdot \frac{\partial t}{\partial \lambda_*^{t \cdot j^*}} \geq 0 \tag{9}$$

where we used the monotone compatibility of $\mathcal{F}$ with $(\mathrm{P}_t)_{t \geq 0}$, Property 3 of Definition 5, and Remark 6.

Returning to Eq. (8), rearranging terms gives that

$$\langle \mathrm{P}_t f, g \rangle - f_0 g_0 = \sum_{\substack{\ell > 0 \\ j^* | \ell}} a_\ell \lambda_t^\ell = a_{j^*} p(\lambda_t^{j^*}) \qquad \text{where}$$
$$p(\lambda_t^{j^*}) := \lambda_t^{j^*} + \frac{1}{a_{j^*}} \sum_{\substack{\ell > j^* \\ j^* | \ell}} a_\ell \lambda_t^\ell. \tag{10}$$

As $\lambda_t \in [0, 1]$, we re-parametrize $u := \lambda_t^{j^*}$ and write $b_\ell := \frac{a_{\ell j^*}}{a_{j^*}}$ for ease of notation; this gives us

$$p(u) = u + \sum_{\ell \geq 2} b_\ell u^\ell.$$

By the Cauchy–Schwarz inequality, we have

$$a_\ell^2 = \langle f_\ell, g_\ell \rangle^2 \leq \langle f_\ell, f_\ell \rangle \langle g_\ell, g_\ell \rangle = \|f_\ell\|^2 \|g_\ell\|^2, \qquad \text{and hence} \qquad |a_\ell| \leq \|f_\ell\| \|g_\ell\|.$$

Once again using the Cauchy–Schwarz inequality, we get

$$\sum_\ell |a_\ell| \leq \sum_{\ell = 0} \|f_\ell\| \cdot \|g_\ell\| \leq \sqrt{\left(\sum_\ell \|f_\ell\|^2 \right) \cdot \left(\sum_\ell \|g_\ell\|^2 \right)} \leq 1$$

where the last inequality follows from the assumption that $\|f\| \leq 1$ for all $f \in \mathcal{F}$. This implies that

$$\sum_\ell |b_\ell| = \frac{1}{|a_{j^*}|} \sum_\ell \left| a_{\ell \cdot j^*} \right| \leq \frac{1}{|a_{j^*}|} = \frac{1}{a_{j^*}}.$$

where the last equality holds because of $a_{j^*} \geq 0$ as shown earlier. If $a_{j^*} > 2/3$ then $\sum_{\ell \geq 2} |b_i| \leq 1/2$ while $b_1 = 1$, from which it easily follows that $\sup_{u \in [0,1]} p(u) \geq 1/2$. If $a_{j^*} < 2/3$ then the power series $p(u)$ satisfies the assumptions of Lemma 11 with $M = \frac{1}{a_{j^*}}$. This gives us

$$\sup_{u\in[0,1]} p(u) \geq \min\left\{\frac{1}{2}, \Theta\left(\frac{1}{\log^2\left(a_{j^*}^{-1}\right)}\right)\right\}.$$

It follows from Definition 5 that as t ranges over $(0, \infty)$, $\lambda_t = \lambda_*^t$ (cf. Remark 6) and consequently u ranges over the interval $(0, 1]$. Together with Eq. (10), this implies that

$$\sup_{t\in(0,\infty)} \langle \mathrm{P}_t f, g\rangle - f_0 g_0 = \sup_{t\in(0,\infty)} a_{j^*} \cdot p(\lambda_t) = a_{j^*} \cdot \sup_{u\in(0,1]} p(u)$$
$$\geq \Theta\left(\min\left\{a_{j^*}, \frac{a_{j^*}}{\log^2\left(a_{j^*}^{-1}\right)}\right\}\right).$$

However, because of monotone compatibility, we have that $\langle \mathrm{P}_t f, g\rangle$ is decreasing in t. As $\mathrm{P}_0 = \mathrm{Id}$, we can conclude that

$$\langle f, g\rangle - f_0 g_0 \geq \Theta\left(\min\left\{a_{j^*}, \frac{a_{j^*}}{\log^2\left(a_{j^*}^{-1}\right)}\right\}\right),$$

which completes the proof. □

5 Robust correlation inequalities over Gaussian space

In this section we prove quantitative versions of Royen's Gaussian Correlation Inequality (GCI) [29] for symmetric convex sets and Hu's inequality for symmetric convex functions [16].[6] We start by recalling some elementary facts about harmonic analysis over Gaussian space, after which we derive our "robust" form of the GCI in Sect. 5.2 as a consequence of Theorem 13. We analyze the tightness of our robust GCI in Sect. 5.3. In Sect. 5.4, we state and prove our quantitative version of Hu's correlation inequality for symmetric convex functions over Gaussian space.

5.1 Harmonic (Hermite) analysis over Gaussian space

Our notation and terminology presented in this subsection follows Chapter 11 of [28]. We say that an n-dimensional *multi-index* is a tuple $\alpha \in \mathbb{N}^n$, and we define

$$\mathrm{supp}(\alpha) := \{i : \alpha_i \neq 0\}, \qquad \#\alpha := |\mathrm{supp}(\alpha)|, \qquad |\alpha| := \sum_{i=1}^n \alpha_i. \tag{11}$$

We write $\mathcal{N}(0, 1)^n$ to denote the n-dimensional standard Gaussian distribution. It is a standard fact (see, for example, Proposition 11.33 of [28]) that the univariate *Hermite*

[6] Note that the 0/1 indicator function of a convex set is not a convex function.

polynomials $(h_j)_{j\in\mathbb{N}}$ form a complete, orthonormal basis for $L^2(\mathbb{R}, \gamma)$. For $n > 1$ the collection of n-variate polynomials given by $(h_\alpha)_{\alpha\in\mathbb{N}^n}$ where

$$h_\alpha(x) := \prod_{i=1}^{n} h_{\alpha_i}(x)$$

forms a complete, orthonormal basis for $L^2(\mathbb{R}^n, \gamma)$. Given a function $f \in L^2(\mathbb{R}^n, \gamma)$ and $\alpha \in \mathbb{N}^n$, we define its *Hermite coefficient on* α as $\widetilde{f}(\alpha) = \langle f, h_\alpha \rangle$. It follows that f is uniquely expressible as $f = \sum_{\alpha\in\mathbb{N}^n} \widetilde{f}(\alpha) h_\alpha$ with the equality holding in $L^2(\mathbb{R}^n, \gamma)$; we will refer to this expansion as the *Hermite expansion* of f. One can check that Parseval's and Plancharel's identities hold in this setting.

Fact 14 (Plancharel's and Parseval's identities). *For $f, g \in L^2(\mathbb{R}^n, \gamma)$, we have:*

$$\langle f, g \rangle = \mathbf{E}_{z\sim\mathcal{N}(0,1)^n}[f(z)g(z)] = \sum_{\alpha\in\mathbb{N}^n} \widetilde{f}(\alpha)\widetilde{g}(\alpha), \qquad \text{(Plancherel)}$$

$$\langle f, f \rangle = \mathbf{E}_{z\sim\mathcal{N}(0,1)^n}[f(z)^2] = \sum_{\alpha\in\mathbb{N}^n} \widetilde{f}(\alpha)^2. \qquad \text{(Parseval)}$$

Next we recall the standard Gaussian noise operator (parameterized so that the noise rate ρ ranges over [0, 1]):

Definition 15 (*Ornstein-Uhlenbeck semigroup*). We define the *Ornstein-Uhlenbeck semigroup* as the family of operators $(\mathrm{U}_\rho)_{\rho\in[0,1]}$ on the space of functions $f \in L^1(\mathbb{R}^n, \gamma)$ given by

$$\mathrm{U}_\rho f(x) := \mathbf{E}_{\boldsymbol{g}\sim\mathcal{N}(0,1)^n}\left[f\left(\rho \cdot x + \sqrt{1-\rho} \cdot \boldsymbol{g}\right)\right].$$

The Ornstein-Uhlenbeck semigroup is sometimes referred to as the family of *Gaussian noise operators* or *Mehler transforms*. The Ornstein-Uhlenbeck semigroup acts on the Hermite expansion as follows:

Fact 16 (Proposition 11.33, [28]). *For $f \in L^2(\mathbb{R}^n, \gamma)$, the function* $\mathrm{U}_\rho f$ *has Hermite expansion*

$$\mathrm{U}_\rho f = \sum_{\alpha\in\mathbb{N}^n} \rho^{|\alpha|} \widetilde{f}(\alpha) h_\alpha.$$

5.2 A robust extension of the Gaussian correlation inequality

We start by making a crucial observation regarding Royen's proof of the Gaussian correlation inequality (GCI) [29]. Recall that the GCI states that if K and L are the indicator functions of two centrally symmetric (i.e. $K(x) = 1$ implies $K(-x) = 1$), convex sets, then they are non-negatively correlated under the Gaussian measure; that is,

$$\mathbf{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)^n}[K(\boldsymbol{x})L(\boldsymbol{x})] - \mathbf{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)^n}[K(\boldsymbol{x})]\,\mathbf{E}_{\boldsymbol{y}\sim\mathcal{N}(0,1)^n}[K(\boldsymbol{y})] \geq 0.$$

In order to prove this, Royen interpolates between $\mathbf{E}[K]\,\mathbf{E}[L]$ and $\mathbf{E}[KL]$ via the Ornstein-Uhlenbeck semigroup, and shows that this interpolation is monotone nondecreasing; indeed, note that

$$\langle \mathrm{U}_1 K, L\rangle = \mathbf{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)^n}[K(\boldsymbol{x})L(\boldsymbol{x})], \qquad \text{and that}$$
$$\langle \mathrm{U}_0 K, L\rangle\, \mathbf{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)^n}[K(\boldsymbol{x})]\,\mathbf{E}_{\boldsymbol{y}\sim\mathcal{N}(0,1)^n}[K(\boldsymbol{y})].$$

Thus, Royen's main result can be interpreted as follows (we refer the interested reader to a simplified exposition of Royen's proof by Latała and Matlak [25] for further details):

Proposition 17 (Royen's Theorem, [29]) *Let $\mathcal{F}_{\mathrm{csc}} \subseteq L^2(\mathbb{R}^n, \gamma)$ be the family of indicators of centrally symmetric, convex sets, and let $(\mathrm{U}_\rho)_{\rho\in[0,1]}$ be the Ornstein-Uhlenbeck semigroup. Then for $K, L \in \mathcal{F}_{\mathrm{csc}}$, we have*

$$\frac{\partial}{\partial\rho}\langle \mathrm{U}_\rho K, L\rangle \geq 0 \qquad \textit{for all } 0 < \rho < 1.$$

In particular, $\mathcal{F}_{\mathrm{csc}}$ is monotone compatible with $(\mathrm{U}_\rho)_{\rho\in[0,1]}$.

Recall that we are parametrizing the Ornstein-Uhlenbeck semigroup by $\rho \in [0, 1]$ where $\rho = e^{-t}$ for $t \in [0, \infty)$; see the discussion following Definition 3. We can now state our main result:

Theorem 18 (Quantitative GCI). *Let $\mathcal{F}_{\mathrm{csc}} \subseteq L^2(\mathbb{R}^n, \gamma)$ be the family of indicators of centrally symmetric, convex sets. Then for $K, L \in \mathcal{F}_{\mathrm{csc}}$, we have*

$$\mathbf{E}[KL] - \mathbf{E}[K]\,\mathbf{E}[L] \geq \frac{1}{C}\cdot\Phi\left(\sum_{|\alpha|=2}\widetilde{K}(\alpha)\widetilde{L}(\alpha)\right) \tag{12}$$

where recall from Eq. (3) that $\Phi : [0, 1] \to [0, 1]$ is $\Phi(x) = \min\left\{x, \frac{x}{\log^2(1/x)}\right\}$ and $C > 0$ is a universal constant.

Proof Consider the orthogonal decomposition

$$L^2(\mathbb{R}^n, \gamma) = \bigoplus_{i=0}^{\infty}\mathcal{W}_i$$

where $\mathcal{W}_i = \mathrm{span}\{h_\alpha : |\alpha| = i\}$; the orthogonality of this decomposition follows from the orthonormality of the Hermite polynomials. From Fact 16, it follows that this decomposition is in fact a chaos decomposition (recall Definition 5) with respect to the Ornstein-Uhlenbeck semigroup $(\mathrm{U}_\rho)_{\rho\in[0,1]}$.

If $K \in \mathcal{F}_{\text{csc}}$, then $K(x) = K(-x)$ as K is the indicator of a centrally symmetric set; in other words, K is an even function. Consequently, its Hermite expansion is given by

$$K = \bigoplus_{\substack{i=0 \\ |\alpha|=2i}}^{\infty} h_\alpha.$$

Furthermore, from Fact 14, we have that

$$\|K\|^2 = \sum_{\alpha\in\mathbb{N}^n} \widetilde{K}(\alpha)^2 = \mathbf{E}\left[K^2\right] \le 1.$$

It follows that the hypotheses of Theorem 13 hold for $\mathcal{F}_{\text{csc}}$ with $j^* = 2$; consequently, for $K, L \in \mathcal{F}_{\text{csc}}$ we have

$$\langle \mathrm{U}_1 K, L\rangle - \langle \mathrm{U}_0 K, L\rangle = \mathbf{E}[KL] - \mathbf{E}[K]\,\mathbf{E}[L] \ge \frac{1}{C}\cdot\Phi\left(\sum_{|\alpha|=2} \widetilde{K}(\alpha)\widetilde{L}(\alpha)\right),$$

which completes the proof of the theorem. □

Remark 19 It is natural to ask whether Theorem 18 can be extended to a broader class of functions than 0/1-valued indicator functions of centrally symmetric, convex sets $\mathcal{F}_{\text{csc}}$. Indeed, the GCI implies the monotone compatibility of centrally symmetric, *quasiconcave*[7], non-negative functions (which is a larger family of functions than $\mathcal{F}_{\text{csc}}$) with the Ornstein-Uhlenbeck semigroup. This allows us to once again use Theorem 13 to obtain a quantitative correlation inequality for this family of functions.

Remark 20 Inspired by the resemblance between Eq. (12) and Talagrand's correlation inequality, we believe that the (negated) degree-2 Hermite coefficients of centrally symmetric, convex sets over Gaussian space are natural analogues of the degree-1 Fourier coefficients (i.e. the coordinate influences) of monotone Boolean functions. A forthcoming manuscript [8] explores this notion further.

5.3 On the tightness of theorem 18

In [31], Talagrand gave the following family of example functions for which Eq. (1) is tight up to constant factors: let $f, g : \{0,1\}^n \to \{0,1\}$ be given by

$$f(x) = \begin{cases} 1 & \sum_i x_i \ge n-k \\ 0 & \text{otherwise} \end{cases}, \qquad \text{and} \qquad g(x) = \begin{cases} 1 & \sum_i x_i > k \\ 0 & \text{otherwise} \end{cases}$$

[7] A function $f : \mathbb{R}^n \to \mathbb{R}$ is *quasiconcave* if for all $\lambda \in [0,1]$ we have $f(\lambda x + (1-\lambda)y) \ge \min\{f(x), f(y)\}$.

where $k \leq n/2$. Writing ε to denote $\mathbf{E}[f]$, we have $\varepsilon^2 = \varepsilon - \varepsilon(1-\varepsilon) = \mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g]$, and it can be shown that $\Psi\left(\sum_{i=1}^n \widehat{f}(i)\widehat{g}(i)\right) = \Theta(\varepsilon^2)$, so Eq. (1) is tight up to constant factors. We note that in this example f and g are the indicator functions of Hamming balls, and that $f \subseteq g$ (i.e. $f(x) = 1$ implies that $g(x) = 1$). Motivated by this example, we consider an analogous pair of functions in the setting of centrally symmetric, convex sets over Gaussian space, where we use origin-centered balls of different radii in place of Hamming balls. The main result of this subsection is that such an example witnesses that Theorem 18 can be tight up to a logarithmic factor (corresponding to the log factor difference between Φ and Ψ). In what follows, all expectations and probabilities are with respect to the n-dimensional Gaussian measure. As before, we will identify centrally symmetric, convex sets with their indicator functions.

Let $K, L \in \mathcal{F}_{\mathrm{csc}}$ be n-dimensional origin-centered balls of radii r_1 and r_2 respectively such that $r_1 < r_2$, $\mathbf{E}[K] = \varepsilon$, and $\mathbf{E}[L] = 1 - \varepsilon$. As $K \subseteq L$, we have $\mathbf{E}[KL] - \mathbf{E}[K]\,\mathbf{E}[L] = \varepsilon - \varepsilon(1-\varepsilon) = \varepsilon^2$. Since $K(x_1, \ldots, x_n) = K(x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n)$ for all $x \in \mathbb{R}^n$ and all $i \in [n]$, it easily follows that $\widetilde{K}(e_i + e_j) = \mathbf{E}[K(\boldsymbol{x})\boldsymbol{x}_i\boldsymbol{x}_j] = 0$ for all $i \neq j$, and the same is true for L. It follows that

$$\sum_{|\alpha|=2} \widetilde{K}(\alpha) = \sum_{i=1}^{n} \widetilde{K}(2e_i)$$

and similarly for L. Furthermore, as K, L are rotationally invariant, we have $\widetilde{K}(2e_i) = \widetilde{K}(2e_j)$ and $\widetilde{L}(2e_i) = \widetilde{L}(2e_j)$ for all $1 \leq i, j \leq n$. By definition, we have

$$-\widetilde{K}(2e_i) = \langle K, -h_2(x_i)\rangle = \mathbf{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)^n}\left[K(\boldsymbol{x})\frac{\left(1-\boldsymbol{x}_i^2\right)}{\sqrt{2}}\right]$$

as $h_2(x) = \frac{x^2-1}{\sqrt{2}}$. Now, note that

$$\begin{aligned} -\sum_{i=1}^{n} \widetilde{K}(2e_i) &= \frac{1}{\sqrt{2}}\sum_{i=1}^{n} \mathbf{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)^n}\left[K(\boldsymbol{x})\left(1-\boldsymbol{x}_i^2\right)\right] \\ &= \frac{1}{\sqrt{2}}\mathbf{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)^n}\left[K(\boldsymbol{x})\left(\sum_{i=1}^{n} 1-\boldsymbol{x}_i^2\right)\right] \\ &= \frac{1}{\sqrt{2}}\mathbf{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)^n}\left[K(\boldsymbol{x})\left(n-\|\boldsymbol{x}\|^2\right)\right]. \end{aligned}$$

In order to obtain a lower bound on the above quantity, we will show that $(n - \|\boldsymbol{x}\|^2)$ is "large" with non-trivial probability for $\boldsymbol{x} \in K$; we will do so by approximating $(n - \|\boldsymbol{x}\|^2)$ by a Gaussian distribution, and then appealing to the Berry-Esseen Central Limit Theorem (see [4, 10] or, for example, Sect. 11.5 of [28]). By the Berry-Esseen theorem, we have that for $t \in \mathbb{R}$,

$$\left| \Pr_{\boldsymbol{x} \sim \mathcal{N}(0,1)^n} \left[\frac{\|\boldsymbol{x}\|^2 - n}{\sqrt{n}} \leq t \right] - \Pr_{\boldsymbol{y} \sim \mathcal{N}(0,1)} \left[\boldsymbol{y} \leq t \right] \right| \leq \frac{c_1}{\sqrt{n}} \tag{13}$$

for some absolute constant c_1. We assume that $\varepsilon \gg c_1/\sqrt{n}$. By standard anti-concentration of the lower tail of the Gaussian distribution, we have that $\Pr_{\boldsymbol{y} \sim \mathcal{N}(0,1)} \left[\boldsymbol{y} \leq t \right] \geq \frac{\varepsilon}{2}$ for $t = -c_2\sqrt{\ln\left(\frac{2}{\varepsilon}\right)}$ where c_2 is an absolute constant. Then it follows from Eq. (13) that

$$\Pr_{\boldsymbol{x} \sim \mathcal{N}(0,1)^n} \left[\frac{\|\boldsymbol{x}\|^2 - n}{\sqrt{n}} \leq -c_2\sqrt{\ln\left(\frac{2}{\varepsilon}\right)} \right] \geq \frac{\varepsilon}{2} \pm \frac{c_1}{\sqrt{n}} \gtrsim \frac{\varepsilon}{2}$$

which can be rewritten as

$$\Pr_{\boldsymbol{x} \sim \mathcal{N}(0,1)^n} \left[\|\boldsymbol{x}\|^2 \leq n - c_2\sqrt{n \ln\left(\frac{2}{\varepsilon}\right)} \right] \gtrsim \frac{\varepsilon}{2}.$$

As $\mathbf{E}[K] = \varepsilon$, it follows that

$$\mathbf{E}_{\boldsymbol{x} \sim \mathcal{N}(0,1)^n} \left[K(\boldsymbol{x})\left(n - \|\boldsymbol{x}\|^2\right) \right] = \Omega\left(\varepsilon\sqrt{n \ln\left(\frac{2}{\varepsilon}\right)} \right)$$

from which we have $-\widetilde{K}(2e_i) \geq \Omega\left(\varepsilon\sqrt{\frac{1}{n}\ln\left(\frac{2}{\varepsilon}\right)}\right)$ for all $i \in [n]$. A similar calculation for L gives that $-\widetilde{L}(2e_i) \geq \Omega\left(\varepsilon\sqrt{\frac{1}{n}\ln\left(\frac{2}{\varepsilon}\right)}\right)$, from which it follows that $\sum_{i=1}^{n} \widetilde{K}(2e_i)\widetilde{L}(2e_i) = \Omega\left(\varepsilon^2 \ln\left(\frac{2}{\varepsilon}\right)\right)$. Recalling Eq. (3), we get that for small enough ε, the quantity

$$\Phi\left(\sum_{|\alpha|=2} \widetilde{K}(\alpha)\widetilde{L}(\alpha) \right) = \Omega\left(\frac{\varepsilon^2}{\log(2/\varepsilon)} \right),$$

which lets us conclude that Theorem 18 is tight to within a logarithmic factor.

5.4 A quantitative extension of Hu's inequality for convex functions

In this section, we consider the following special case of Hu's inequality [16]:

Theorem 21 (Hu's inequality). *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be centrally symmetric, convex functions. Then*

$$\mathbf{E}_{\boldsymbol{x} \sim \mathcal{N}(0,1)^n}[f(\boldsymbol{x})g(\boldsymbol{x})] - \mathbf{E}_{\boldsymbol{x} \sim \mathcal{N}(0,1)^n}[f(\boldsymbol{x})]\,\mathbf{E}_{\boldsymbol{y} \sim \mathcal{N}(0,1)^n}\left[g(\boldsymbol{y})\right] \geq 0.$$

As in Sect. 5.2, we will obtain a quantitative extension of Theorem 21 by appealing to Theorem 13. The Markov semigroup we will use here will once again be the Ornstein–Uhlenbeck semigroup $(\mathrm{U}_\rho)_{\rho\in[0,1]}$—monotone compatibility of this semigroup with the family of centrally symmetric, convex functions (which we will denote $\mathcal{F}_{\mathrm{cvx}}$) was proved by Hargé [15].

Fact 22 (Proof of Theorem 2.1, [15]). *Let $\mathcal{F}_{\mathrm{cvx}}$ denote the family of centrally symmetric, convex functions with $\|f\| \le 1$ for all $f \in \mathcal{F}_{\mathrm{cvx}}$. Then $\mathcal{F}_{\mathrm{cvx}}$ is monotone compatible with $(\mathrm{U}_\rho)_{\rho\in[0,1]}$.*

The proof of the following result is identical to that of Theorem 18 and is therefore omitted.

Theorem 23 (Quantitative Hu's inequality). *Let $\mathcal{F}_{\mathrm{cvx}} \subseteq L^2(\mathbb{R}^n, \gamma)$ be the family of centrally symmetric, convex functions. Then for $f, g \in \mathcal{F}_{\mathrm{cvx}}$, we have*

$$\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] \ge \frac{1}{C} \cdot \Phi\left(\sum_{|\alpha|=2} \widetilde{f}(\alpha)\widetilde{g}(\alpha)\right)$$

where recall from Eq. (3) that $\Phi : [0,1] \to [0,1]$ is $\Phi(x) = \min\left\{x, \frac{x}{\log^2(1/x)}\right\}$ and $C > 0$ is a universal constant.

6 A quantitative correlation inequality for arbitrary finite product domains

The main result of this section, Theorem 28, is an extension of Talagrand's correlation inequality [31] to real-valued functions on general, finite, product spaces. (Recall that Talagrand's inequality applies only to Boolean-valued functions on the domain $\{0,1\}^n$ under the uniform distribution.)

6.1 Harmonic analysis over finite product spaces

Our notation and terminology presented in this subsection follows Chapter 8 of [28]. We use multi-index notation for $\alpha \in \mathbb{N}^n$ as defined in Eq. (11).

Let (Ω, π) be a finite probability space with $|\Omega| = m \ge 2$, where we always assume that the distribution π over Ω has full support (i.e. $\pi(\omega) > 0$ for every $\omega \in \Omega$). We write $L^2(\Omega^n, \pi^{\otimes n})$ for the real inner product space of functions $f : \Omega^n \to \mathbb{R}$, with inner product $\langle f, g\rangle := \mathbf{E}_{\boldsymbol{x}\sim\pi^{\otimes n}}[f(\boldsymbol{x})g(\boldsymbol{x})]$.

It is easy to see that there exists an orthonormal basis for the inner product space $L^2(\Omega, \pi)$, i.e. a set of functions $\phi_0, \ldots, \phi_{m-1} : \Omega \to \mathbb{R}$, with $\phi_0 = 1$, that are orthonormal with respect to π. Moreover, such a basis extends to an orthonormal basis for $L^2(\Omega^n, \pi^{\otimes n})$ by a straightforward n-fold product construction: given a multi-index

$\alpha \in \mathbb{N}^n_{<m}$, if we define $\phi_\alpha \in L^2(\Omega^n, \pi^{\otimes n})$ as

$$\phi_\alpha(x) := \prod_{i=1}^{n} \phi_{\alpha_i}(x_i),$$

then the collection $(\phi_\alpha)_{\alpha\in\mathbb{N}^n_{<m}}$ is an orthonormal basis for $L^2(\Omega^n, \pi^{\otimes n})$ (see Proposition 8.13 of [28]). So every function $f : \Omega^n \to \mathbb{R}$ has a decomposition

$$f = \sum_{\alpha\in\mathbb{N}^n_{<m}} \widehat{f}(\alpha)\phi_\alpha. \tag{14}$$

This can be thought of as a "Fourier decomposition" for f, in that it satisfies both Parseval's and Plancharel's identities (see Proposition 8.16 of [28]). We now proceed to define a noise operator for finite product spaces.

Definition 24 (*Noise operator for finite product spaces*). Fix a finite product probability space $L^2(\Omega^n, \pi^{\otimes n})$. For $\rho \in [0, 1]$ we define the *noise operator for* $L^2(\Omega^n, \pi^{\otimes n})$ as the linear operator

$$\mathrm{T}_\rho f(x) := \mathbf{E}_{\boldsymbol{y}\sim N_\rho(x)}[f(\boldsymbol{y})],$$

where "$\boldsymbol{y} \sim N_\rho(x)$" means that $\boldsymbol{y} \in \Omega^n$ is randomly chosen as follows: for each $i \in [n]$, with probability ρ set $\boldsymbol{y}_i$ to be x_i and with the remaining $1-\rho$ probability set $\boldsymbol{y}_i$ by independently making a draw from π.

It is easy to check that $\mathrm{T}_\rho f = \sum_\alpha \rho^{\#\alpha} \widehat{f}(\alpha)\phi_\alpha$ (Proposition 8.28 of [28]).

6.2 A quantitative correlation inequality for finite product domains

Throughout this subsection, let $\Omega = \{0, 1, \ldots, m-1\}$ endowed with the natural ordering (though any m-element totally ordered set would do). We will consider monotone functions on $(\Omega^n, \pi^{\otimes})$; while our results hold in the more general setting of functions on $(\Omega^n, \otimes_{i=1}^n \pi_i)$, we stick to the setting of $L^2(\Omega^n, \pi^{\otimes n})$ for ease of exposition.

In order to appeal to Theorem 13, we must first show that the family of monotone (nondecreasing) functions on Ω^n is monotone compatible with the Bonami–Beckner noise operator (see Definition 24). To this end, we define noise operators that act on each coordinate of the input:

Definition 25 (*coordinate-wise noise operators*). Let T^i_ρ be the operator on functions $f : \Omega^n \to \mathbb{R}$ defined by

$$\mathrm{T}^i_\rho f(x) = \mathbf{E}_{\boldsymbol{y}\sim N_\rho(x_i)}\big[f(x_1, \ldots, \boldsymbol{y}, \ldots, x_n)\big],$$

and define $\mathrm{T}_{\rho_1,\ldots,\rho_n} f := \mathrm{T}^1_{\rho_1} \circ \mathrm{T}^2_{\rho_2} \circ \ldots \circ \mathrm{T}^n_{\rho_n} f$.

This is well-defined as the operators $\mathrm{T}^i_{\rho_i}$ and $\mathrm{T}^j_{\rho_j}$ commute.

Lemma 26 *Let $\Omega = \{0, 1, \ldots, m-1\}$ and let $f : \Omega^n \to \mathbb{R}$ be a monotone function. Then $\mathrm{T}^i_\rho f : \Omega^n \to \mathbb{R}$ is a monotone function.*

Proof Suppose $x, y \in \Omega^n$ are such that $x_i \leq y_i$ for all $i \in [n]$. We wish to show that $\mathrm{T}^i_\rho f(x) \leq \mathrm{T}^i_\rho f(y)$, which is equivalent to showing

$$\mathbf{E}_{z\sim N_\rho(x_i)}\left[f\left(x^{i\mapsto z}\right)\right] \leq \mathbf{E}_{z\sim N_\rho(y_i)}\left[f\left(y^{i\mapsto z}\right)\right].$$

Indeed, because of the monotonicity of f, via the natural coupling we have

$$\begin{aligned}\mathbf{E}_{z\sim N_\rho(x_i)}\left[f\left(x^{i\mapsto z}\right)\right] &= \delta f(x) + (1-\delta)\,\mathbf{E}_{z\sim\Omega^n}\left[f\left(x^{i\mapsto z}\right)\right] \\ &\leq \delta f(y) + (1-\delta)\,\mathbf{E}_{z\sim\Omega^n}\left[f\left(y^{i\mapsto z}\right)\right] \\ &= \mathbf{E}_{z\sim N_\rho(y_i)}\left[f\left(y^{i\mapsto z}\right)\right].\end{aligned}$$

□

Lemma 27 *Let $\Omega = \{0, 1, \ldots, m-1\}$ and let $f, g : \Omega^n \to \mathbb{R}$ be monotone functions. Then $\langle \mathrm{T}_\rho f, g\rangle$ is nondecreasing in $\rho \in [0, 1]$.*

Proof We have

$$\left\langle \mathrm{T}_{\rho_1,\ldots,\rho_n} f, g\right\rangle = \left\langle \mathrm{T}_{\rho,1,\ldots,1} f, T_{1,\rho_2,\ldots,\rho_n} g\right\rangle = \left\langle \mathrm{T}^1_{\rho_1} f, h\right\rangle$$

where $h := \mathrm{T}_{1,\rho_2,\ldots,\rho_n} g$. It follows from a repeated application of Lemma 26 that h is monotone. Now, note that

$$\left\langle \mathrm{T}^1_{\rho_1} f, h\right\rangle = \widehat{f}(\bar{0}) \cdot \widehat{h}(\bar{0}) + \sum_{\alpha_1>0} \rho_1 \widehat{f}(\alpha)\widehat{h}(\alpha) + \sum_{\substack{\bar{0}\neq\alpha \\ \alpha_1=0}} \widehat{f}(\alpha)\widehat{h}(\alpha)$$

where $\bar{0} = (0, \ldots, 0)$. By Cheybshev's order inequality, we know that $\left\langle \mathrm{T}^1_1 f, h\right\rangle \geq \left\langle \mathrm{T}^1_0 f, h\right\rangle = \widehat{f}(\bar{0}) \cdot \widehat{h}(\bar{0}) + \sum_{\bar{0}\neq\alpha,\alpha_1=0} \widehat{f}(\alpha)\widehat{h}(\alpha)$. From the above expression, we have:

$$\frac{\partial}{\partial\rho_1}\left\langle \mathrm{T}^1_{\rho_1} f, h\right\rangle = \sum_{\alpha_1>0} \widehat{f}(\alpha)\widehat{h}(\alpha)$$

which must be nonnegative since $\left\langle \mathrm{T}^1_1 f, h\right\rangle \geq \left\langle \mathrm{T}^1_0 f, h\right\rangle$, and so we can conclude that $\left\langle \mathrm{T}^1_{\rho_1} f, h\right\rangle$ is nondecreasing in ρ_1. The result then follows by repeating this for each coordinate. □

Let $\mathcal{F}_{\text{mon}} \subseteq L^2(\Omega^n, \pi^{\otimes n})$ be the family of monotone functions $f : \Omega^n \to \mathbb{R}$. Then Lemma 27 shows that $\mathcal{F}_{\text{mon}}$ is monotone compatible with the Bonami–Beckner noise operator. We can now prove our Talagrand-analogue for monotone functions over Ω^n:

Theorem 28 *Let* $\Omega = \{0, 1, \ldots, m-1\}^n$ *and let* $\mathcal{F}_{\text{mon}} \subseteq L^2(\Omega^n, \pi^{\otimes n})$ *denote the family of monotone functions on* Ω^n *such that* $\|f\| \le 1$ *for all* $f \in \mathcal{F}_{\text{mon}}$. *Then for* $f, g \in \mathcal{F}_{\text{mon}}$, *we have*

$$\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] \ge \frac{1}{C} \cdot \Phi\left(\sum_{\#\alpha=1} \widehat{f}(\alpha)\widehat{g}(\alpha)\right)$$

where recall from Eq. (3) *that* $\Phi : [0,1] \to [0,1]$ *is* $\Phi(x) = \min\left\{x, \frac{x}{\log^2(1/x)}\right\}$ *and* $C > 0$ *is a universal constant.*

Proof Consider the orthogonal decomposition

$$L^2(\Omega^n, \pi^{\otimes n}) = \bigoplus_{i=0}^{n} \mathcal{W}_i$$

where $\mathcal{W}_i = \text{span}\,\{\phi_\alpha : \#\alpha = i\}$; the orthogonality of this decomposition follows from the orthonormality of $(\phi_\alpha)_{\alpha \in \mathbb{N}^n_{<m}}$. Furthermore, this decomposition is a chaos decomposition with respect to the Bonami–Beckner operator $(\mathrm{T}_\rho)_{\rho \in [0,1])}$. It follows that the hypotheses of Theorem 13 hold for $\mathcal{F}_{\text{mon}}$ with $j^* = 1$, from which the result follows. □

6.3 Comparison with Keller's quantitative correlation inequality for the *p*-biased hypercube

In this subsection we restrict our attention to the p-biased hypercube $\{-1,1\}^n_p = (\{-1,1\}^n, \pi_p^{\otimes n})$ where $\pi_p(-1) = p$ and $\pi_p(+1) = 1-p$. In this setting our Theorem 28 generalizes Talagrand's inequality in two ways: it holds for real-valued monotone functions on $\{-1,1\}^n$ that have 2-norm at most 1 (rather than just monotone Boolean functions), and it holds for any p (as opposed to just $p = 1/2$). Keller [17, 18] has earlier given a generalization of Talagrand's inequality that holds for general p and for real-valued monotone functions with ∞-norm at most 1:

Theorem 29 (Theorem 7 of [17]; see also [19] for a slightly weaker version). *Let* $f, g \in L^2(\{0,1\}^n, \pi_p^{\otimes n})$ *be monotone functions such that for all* $x \in \{-1,1\}^n$, *we have* $|f(x)|, |g(x)| \le 1$. *Then*

$$\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] \ge \frac{1}{C} \cdot H(p) \cdot \Psi\left(\sum_{i=1}^{n} \widehat{f_p}(i)\widehat{g_p}(i)\right)$$

where $\widehat{f_p}(i)$ *is the* p*-biased degree-1 Fourier coefficient on coordinate* i, $\Psi : [0,1] \to [0,1]$ *is given by* $\Psi(x) = \frac{x}{\log(e/x)}$ *as in Sect.* 1.1, $C > 0$ *is a universal constant, and* $H : [0,1] \to [0,1]$ *is the binary entropy function* $H(x) = -x\log x - (1-x)\log(1-x)$.

Comparing Theorem 28 to Theorem 29, we see that the latter has an extra factor of $H(p)$, whereas the former shows that in fact no dependence on p is necessary (but the former has an extra factor of $\frac{1}{\log\left(1/\sum_i \widehat{f_p}(i)\widehat{g_p}(i)\right)}$). Theorem 28 can be significantly stronger than Theorem 29 in a range of natural settings because of these differences. In Appendix B we show that for every $\omega(1)/n \le p \le 1/2$, there is a pair of $\{-1, 1\}$-valued functions f, g (depending on p) such that under the p-biased distribution (i) the quantity $\mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g]$ is at least an absolute constant independent of n and p; (ii) the RHS of Theorem 28 is at least an absolute constant independent of n and p; but (iii) the RHS of Theorem 29 is $\Theta(p \log(1/p))$.

Acknowledgements A.D. is supported by NSF grants CCF 1910534 and CCF 1926872. S.N. is supported by NSF grants CCF-1563155 and by CCF-1763970. R.A.S. is supported by NSF grants CCF-1814873, IIS-1838154, CCF-1563155, and by the Simons Collaboration on Algorithms and Geometry. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF). This work was done while A.D. was participating in the "Probability, Geometry, and Computation in High Dimensions" program at the Simons Institute for the Theory of Computing.

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

A Proof of Claim 12

For $c \in \mathbb{N}$, let $T_c(x)$ denote the degree-c Chebyshev polynomial of the first kind. Define the univariate polynomial:

$$a_d(t) := \frac{T_{\sqrt{d}}\left(t\left(1+\frac{3}{d}\right)\right)}{T_{\sqrt{d}}\left(1+\frac{3}{d}\right)}$$

where d is a parameter (a perfect square) that we will set later. We make the following simple observations:

- $|a_d(t)| \le 1$ for all $t \in [0, 1]$, and $a(1) = 1$.
- Let $d \ge 4$. For $t \in \left[0, 1-\frac{3}{d}\right]$, we have $a_d(t) \in \left[-\frac{1}{4}, \frac{1}{4}\right]$. This follows from the fact that $\left(1-\frac{3}{d}\right)\left(1+\frac{3}{d}\right) < 1$, that $|T_{\sqrt{d}}(t)| \le 1$ for $|t| \le 1$, and that the derivative $T'_{\sqrt{d}}(t)$ is at least d for all $t \ge 1$.
- The sum of the absolute values of the coefficients of $a_d(t)$ is at most $2^{O\left(\sqrt{d}\right)}$. This is an easy consequence of standard coefficient bounds for Chebyshev polynomials (see e.g. Sect. 2.3.2 of [26]).

For simplicity, assume $\log^2 M = 4^k$ for some $k \in \mathbb{N}$. We define $b(t)$ as

$$b(t) := a_1(1-t)\cdot a_4(1-t)\cdot a_{16}(1-t)\cdots a_{\log^2 M}(1-t).$$

Note that $b(t)$ is a polynomial of degree $\sqrt{1}+\sqrt{4}+\sqrt{16}+\ldots+\sqrt{\log^2 M} = \Theta(\log M)$, and that $|b(t)| \le 1$ for all $t \in [0, 1]$. It follows from the third item above that the sum of the absolute values of the coefficients of $b(t)$ is at most

$$2^{O(\sqrt{1})+O(\sqrt{4})+\ldots+O\left(\sqrt{\log^2 M}\right)} = 2^{O(\log M)}.$$

Finally, we define

$$p(t) := t \cdot b(t).$$

In order to upper bound $|p(t)|$ for $t \in [0, 1]$, we first observe that if $t \leq \frac{1}{4^k}$ then we have $|p(t)| \leq \frac{1}{4^k}|b(t)| \leq \frac{1}{4^k} \leq \frac{1}{\log^2 M}$ as desired. Thus we may suppose that $t \in \left[\frac{1}{4^i}, \frac{1}{4^{i-1}}\right]$ for some $i \in \{1, \ldots, k\}$; in particular, let $t = \frac{1}{4^i} + \delta$ for $\delta \in \left[0, \frac{3}{4^i}\right]$. Now, for each $j \geq i+1$, we have

$$|a_{4^j}(1-t)| \leq \frac{1}{4} \qquad \text{which implies that} \qquad |a_{4^{(i+1)}}(t)| \cdot |a_{4^{(i+2)}}(t)| \cdots |a_{4^k}(t)| \leq \frac{1}{4^{k-i}}.$$

As $t \leq \frac{1}{4^{i-1}}$, it follows that

$$|p(t)| = |t \cdot b(t)| \leq \frac{1}{4^{i-1}} \cdot \frac{1}{4^{k-i}} = \frac{1}{4^{k-1}} = \Theta\left(\frac{1}{\log^2 M}\right),$$

and Claim 12 is proved. It follows that Lemma 11 is tight up to constant factors.

B Comparison of Theorem 28 and Theorem 29

Let $\omega(1)/n \leq p \leq 1/2$. Observe that under $\{-1, 1\}^n_p$ we have $\mathbf{E}[\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_n] = n(1-2p)$. We define $f : \{-1, 1\}^n \to \{-1, 1\}$ to be the "p-biased analogue of the majority function," i.e.

$$f(x) := \operatorname{sign}(x_1 + \cdots + x_n - n(1-2p)),$$

and we take $g = f$.

Since (as is well known) the median of the Binomial distribution $\mathrm{Bin}(n, p)$ differs from the mean by at most 1, it follows (using the Littlewood-Offord anticoncentration inequality described below) that $\mathbf{E}[f] = o(1)$, and hence we have (i): $\mathbf{E}[fg] - \mathbf{E}[f]\mathbf{E}[g] \geq 1 - o(1)$. To establish (ii) and (iii) it remains only to show that for any fixed $i \in [n]$ we have that the p-biased degree-1 Fourier coefficient $\widehat{f_p}(i)$ is at least $\Omega(1/\sqrt{n})$, or equivalently, that $\widehat{f_p}(1) + \cdots + \widehat{f_p}(n) = \Omega(\sqrt{n})$. To see this, we observe that this sum of degree-1 Fourier coefficients is

$$\sum_{i=1}^{n} \widehat{f_p}(i) = \mathbf{E}\left[f(\boldsymbol{x}) \cdot \sum_{i=1}^{n} \frac{\boldsymbol{x}_i - (1-2p)}{2\sqrt{p(1-p)}} \right]$$

$$= \frac{1}{2\sqrt{p(1-p)}} \mathbf{E}\left[\left|\left(\sum_{i=1}^{n} \boldsymbol{x}_i\right) - n(1-2p)\right|\right]. \tag{15}$$

We now recall the Littlewood-Offord anticoncentration inequality for the p-biased Boolean hypercube (see e.g. Theorem 5 of [7] or [1]). Specialized to our context, this says that for any real interval I of length at least 1, it holds that $\Pr\left[\sum_{i=1}^{n} \boldsymbol{x}_i \in I\right] \leq O(|I|)/\sqrt{np(1-p)}$. Taking I to be the interval of length $c\sqrt{np(1-p)}$ centered at $n(1-2p)$ for a suitably small positive constant c, it holds that

$$\Pr\left[\left|\left(\sum_{i=1}^{n} \boldsymbol{x}_i\right) - n(1-2p)\right| \geq c\sqrt{np(1-p)}\right] \geq \frac{1}{2}.$$

Consequently

$$\mathbf{E}\left[\left|\left(\sum_{i=1}^{n} \boldsymbol{x}_i\right) - n(1-2p)\right|\right] \geq \frac{c\sqrt{np(1-p)}}{2},$$

which together with Eq. (15) gives that $\sum_{i=1}^{n} \widehat{f_p}(i) \geq c\sqrt{n}/4$ as desired.

References

1. Aizenman, M., Germinet, F., Klein, A., Warzel, S.: On Bernoulli decompositions for random variables, concentration bounds, and spectral localization. Probab. Theory Relat. Fields **143**(1–2), 219–238 (2009)
2. Borwein, P., Erdélyi, T.: Littlewood-type polynomials on subarcs of the unit circle. Indiana Univ. Math. J. **46**(4), 1323–1346 (1997)
3. Borwein, P., Erdélyi, T., Kós, G.: Littlewood-type problems on [0, 1]. In: Proceedings of the London Mathematical Society, vol. 3, no. 79, pp. 22–46 (1999)
4. Berry, A.C.: The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Am. Math. Soc. **49**(1), 122–136 (1941)
5. Bakry, D., Gentil, I., LeDoux, M.: Analysis and Geometry of Markov Diffusion Operators. Springer, New York (2013)
6. Cordero-Erausquin, D., Ledoux, M.: Hypercontractive measures, Talagrand's inequality, and influences. In: Klartag, B., Mendelson, S., Milman, V. (Eds.) Geometric Aspects of Functional Analysis. Lecture Notes in Mathematics, vol 2050. Springer, Berlin, Heidelberg, pp. 169–189 (2012)
7. De, A., Diakonikolas, I., Servedio, R.A.: The inverse Shapley value problem. Games Econ. Behav. **105**, 122–147 (2017)
8. De, A., Nadimpalli, S., Servedio, R.A.: Convex influences. In preparation. (2021)
9. Eldan, R.: Second-order bounds on correlations between increasing families. arXiv:1912.11641 (2019)
10. Esseen, C.-G.: On the Liapunoff limit of error in the theory of probability. Ark. Mat. Astron. Fys. **A**, 1–19 (1942)
11. Fortuin, C.M., Kasteleyn, P.W., Ginibre, J.: Correlation inequalities on some partially ordered sets. Commun. Math. Phys. **22**(2), 89–103 (1971)
12. Gallavotti, G.: A proof of the Griffiths inequalities for the XY model. Stud. Appl. Math **50**(1), 89–92 (1971)
13. Griffiths, R.: Correlations in Ising ferromagnets. I. J. Math. Phys. **8**(3), 478–483 (1967)
14. Harris, T.E.: A lower bound for the critical probability in a certain percolation process. In: Proceedings of the Cambridge Philosophical Society, vol. 56, pp. 13–20 (1960)

15. Hargé, G.: Characterization of equality in the correlation inequality for convex functions, the U-conjecture. Ann. Inst. Henri Poincaré Probabilités et Stat. **41**(4), 753–765 (2005)
16. Hu, Y.: Itô-Wiener chaos expansion with exact residual and correlation, variance inequalities. J. Theor. Probab. **10**(4), 835–848 (1997)
17. Keller, N.: Improved FKG inequality for product measures on the discrete cube. (2008)
18. Keller, N.: Influences of variables on Boolean functions. PhD thesis, Hebrew University of Jerusalem, (2009)
19. Keller, N.: A simple reduction from a biased measure on the discrete cube to the uniform measure. Eur. J. Comb. **33**, 1943–1957 (2012)
20. Kalai, G., Keller, N., Mossel, E.: On the correlation of increasing families. J. Comb. Theory Ser. A **144**, 11 (2015)
21. Kleitman, D.J.: Families of non-disjoint subsets. J. Comb. Theory **1**(1), 153–155 (1966)
22. Keller, N., Mossel, E., Sen, A.: Geometric influences. Ann. Probab. **40**(3), 1135–1166 (2012)
23. Keller, N., Mossel, E., Sen, A.: Geometric influences II: correlation inequalities and noise sensitivity. Ann. Inst. l'IHP Poincaré Probab. Stat. **50**(4), 1121–1139 (2014)
24. Kelly, D., Sherman, S.: General Griffiths' inequalities on correlations in Ising ferromagnets. J. Math. Phys. **9**(3), 466–484 (1968)
25. Latała, R., Matlak D.: Royen's Proof of the Gaussian Correlation Inequality. In: Klartag B., Milman E. (Eds.) Geometric Aspects of Functional Analysis. Lecture Notes in Mathematics, vol 2169. Springer, Cham, pp. 265–275(2017)
26. Mason, J.C., Handscomb, D.C.: Chebyshev Polynomials. CRC Press, Boca Raton (2002)
27. Mossel, E.: Probabilistic aspects of voting, intransitivity and manipulation. (2020)
28. O'Donnell, R.: Analysis of Boolean Functions. Cambridge University Press, Cambridge (2014)
29. Royen, T.: A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions. arXiv preprint arXiv:1408.1028, (2014)
30. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill Inc., New York (1987)
31. Talagrand, M.: How much are increasing sets positively correlated? Combinatorica **16**(2), 243–258 (1996)