Quantitative Correlation Inequalities via Semigroup Interpolation

Anindya De
University of Pennsylvania, Philadelphia, PA, USA
anindyad@seas.upenn.edu

Shivam Nadimpalli
Columbia University, New York, NY, USA
sn2855@columbia.edu

Rocco A. Servedio
Columbia University, New York, NY, USA
ras2105@columbia.edu

──── **Abstract** ────

Most correlation inequalities for high-dimensional functions in the literature, such as 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. We give a general approach that can be used to bootstrap many *qualitative* correlation inequalities for functions over product spaces into *quantitative* statements. The approach combines a new extremal result about power series, proved using complex analysis, with harmonic analysis of functions over product spaces. We instantiate this general approach in several different concrete settings to obtain a range of new and near-optimal quantitative correlation inequalities, including:

- A quantitative version of Royen's celebrated Gaussian Correlation Inequality [23]. In [23] 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 $\{0,1\}^n$ [26]. 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 [26]; the only prior generalization of which we are aware is due to Keller [17, 15, 16], which extended [26] to product distributions over $\{0,1\}^n$. In the special case of p-biased distributions over $\{0,1\}^n$ that was considered by Keller, our new bound essentially saves a factor of $p \log(1/p)$ over the quantitative bounds given in [17, 15, 16]. We also give a quantitative version of the FKG inequality for monotone functions over the continuous domain $[0,1]^n$, answering a question of Keller [16].

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 [10, 19], 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 [23]. Other well-known correlation inequalities include the Fortuin-Kasteleyn-Ginibre (FKG) inequality [7], which is an important tool in statistical mechanics and probabilistic combinatorics; the Griffiths–Kelly–Sherman (GKS) inequality [9, 18], which is a correlation inequality for ferromagnetic spin systems; and various generalizations of the GKS inequality to quantum spin systems [8, 25].

1.1 Quantitative Correlation Inequalities

Here, we attempt 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 [26] 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 the analysis of Boolean functions [21]. For a Boolean function $f : \{0, 1\}^n \to \{0, 1\}$, the *influence* of coordinate i on f is defined to be $\mathbf{Inf}_i[f] := \mathbf{Pr}_{\boldsymbol{x} \sim U_n}[f(\boldsymbol{x}) \neq f(\boldsymbol{x}^{\oplus i})]$, 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:

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

■ **Table 1** Qualitative and quantitative correlation inequalities. Here γ denotes the standard Gaussian distribution $\mathcal{N}(0,1)^n$; $\{0,1\}_p^n$ denotes the p-biased hypercube (with no subscript corresponding to $p = 1/2$, i.e. the uniform distribution); π denotes any distribution over $\{0, \ldots, m-1\}$; and $[0,1]^n$ is endowed with the Lebesgue measure.

	Qualitative Bounds	Quantitative Bounds
Monotone $f, g \in L^2(\mathbb{R}^n, \gamma)$	[7]	[14]
Symmetric, convex $K, L \subseteq \mathbb{R}_\gamma^n$	[23]	Theorem 21
Convex $f, g \in L^2(\mathbb{R}^n, \gamma)$	[11]	Deferred to full version.
Monotone $f, g : \{0,1\}^n \to \{0,1\}$	[10, 19, 7]	[26]
Monotone $f, g : \{0,1\}_p^n \to \mathbb{R}$	[7]	[15], Theorem 29
Monotone $f, g : \{0, \ldots, m-1\}_\pi^n \to \mathbb{R}$	[7]	Theorem 29
Monotone $f, g \in L^2([0,1]^n)$	[22]	Deferred to full version.

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 [26] itself, Talagrand gives an example for which Equation (1) is tight up to constant factors.

Talagrand's result has proven to be influential in the theory of Boolean functions, and several works [15, 16, 17, 12] have obtained extensions and variants of this inequality for product distributions over $\{0,1\}^n$. An analogue of Talagrand's inequality in the setting of monotone functions over Gaussian space was obtained by Keller, Mossel and Sen [14] using a new notion of "geometric influences." Beyond these results, we are not aware of *quantitative* correlation inequalities in other settings, even though (as discussed above) a wide range of *qualitative* correlation inequalities are known. In particular, even for very simple and concrete settings such as the solid cube $[0,1]^n$ endowed with the uniform measure or the m-ary cube $\{0,1,\ldots,m-1\}^n$ with a product measure, no quantitative versions of the FKG inequality were known (see the discussion immediately following Theorem 4 of Keller [13]). As a final example, no quantitative version of the Gaussian Correlation Inequality was previously known.

1.2 Our Contributions

We establish a general framework to transfer qualitative correlation inequalities into quantitative correlation inequalities. We apply this general framework to obtain a range of new quantitative correlation inequalities, which include the following:

1. Quantitative versions of Royen's Gaussian Correlation Inequality and Hu's correlation inequality [11] for symmetric convex functions over Gaussian space;

2. A quantitative FKG inequality for a broad class of product distributions, including arbitrary product distributions over finite domains and the uniform distribution over $[0,1]^n$.

All these results are obtained in a unified fashion via simple proofs that are substantially different from previous works [26, 15, 16, 17, 12]. We also give several lower bound examples, including one which shows that our quantitative version of the Gaussian Correlation Inequality is within a logarithmic factor of the best possible bound.

We note that the special case of item 2 above with the uniform distribution on $\{0, 1\}^n$ essentially recovers Talagrand's correlation inequality [26]. In more detail, our bound is weaker than that obtained in [26] 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 [15, 16, 17] which have a $p \log(1/p)$ dependence (though, similar to the situation vis-a-vis [26], we lose a logarithmic factor in other dependencies). Finally, for the uniform distribution over $[0, 1]^n$, our result gives an answer to a question posed by Keller [16], who wrote "It seems tempting to find a generalization of Talagrand's result to the continuous setting, but it is not clear what is the correct notion of influences in the continuous case that should be used in such generalization."

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; 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:

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

(Note the similarity between Φ and the function Ψ mentioned earlier that arose in Talagrand's quantitative correlation inequality [26]; 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 $\mathbf{E}_\mu\left[f^2\right] \leq 1$ 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.[1] 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 semigroup whose stationary distribution is μ; this is defined more precisely in Section 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 18.) 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 [23] 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 $\{\rho^j\}_{j \geq 0}$. Let $\{\mathcal{W}_j\}_{j \geq 0}$ be the corresponding eigenspaces. Consequently, we can express $q(\rho) - q(0)$ as

$$q(\rho) - q(0) = \mathbf{E}_{\boldsymbol{x} \sim \mu}[f(\boldsymbol{x})\mathrm{T}_\rho g(\boldsymbol{x})] - \mathbf{E}_{\boldsymbol{x} \sim \mu}[f(\boldsymbol{x})] \cdot \mathbf{E}_{\boldsymbol{y} \sim \mu}[g(\boldsymbol{y})] = \sum_{j > 0} \rho^j \, \mathbf{E}[f_j(\boldsymbol{x})g_j(\boldsymbol{x})], \qquad (3)$$

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 (in fact, as we will soon see, a_{j^*} also gives a quantitative lower bound on this slack). For example, for the Gaussian Correlation Inequality, it turns out that $j^* = 2$ (for most of the other applications of our general framework in this paper, it turns out that $j^* = 1$).
2. For any i such that j^* does not divide i, $a_i = 0$.

---

[1] 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.

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 12, implies (see the proof of Theorem 14) that for any such power series $q(\cdot)$, there exists some $\rho^* \in [0, 1]$ such that

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

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 Equation (3), it follows that

$$q(1) - q(0) = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mu}[f(\boldsymbol{x})g(\boldsymbol{x})] - \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mu}[f(\boldsymbol{x})] \cdot \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \mu}[g(\boldsymbol{y})] \geq \frac{1}{C} \cdot \Phi(a_{j^*}), \tag{4}$$

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 (Equation (1)):

$$\mathbf{E}[fg] - \mathbf{E}[f]\mathbf{E}[g] \geq \frac{1}{C} \cdot \Phi\left(\sum_{i=1}^n \mathbf{Inf}_i[f]\mathbf{Inf}_i[g]\right), \tag{5}$$

for an absolute constant $C > 0$. While our bound is weaker than that of [26] by a log factor (recall the difference between Ψ and Φ), our methods are applicable to a much 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 [26]; for example, [26]'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.

▶ **Remark 3.** Coupled with the first property described above, Equation (4) shows that a_{j^*} not only qualitatively captures the "correlation gap"

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mu}[f(\boldsymbol{x})g(\boldsymbol{x})] - \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mu}[f(\boldsymbol{x})] \cdot \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \mu}[g(\boldsymbol{y})],$$

but also provides a quantitative lower bound on this gap.

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 [1].

▶ **Definition 4** (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.*[2]

It is well known that a Markov semigroup can be constructed from a Markov process and vice versa [1]. 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 5** (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 Section 5 and Section 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 Section 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 $(\otimes_{i=1}^n \bar{\mathrm{P}}_{t_i})_{t_i \geq 0}$ on $L^2(\Omega^n, \pi^{\otimes n})$. We write $\mathrm{P}_{\bar{t}}$ to denote this semigroup, and write P_t to denote the Markov semigroup $(\otimes_{i=1}^n \mathrm{P}_t)_{t \geq 0}$. We next define a decomposition of $L^2(\Omega^n, \pi^{\otimes n})$ that is particularly well-suited to the action of a Markov semigroup $(\mathrm{P}_t)_{t \geq 0}$.

▶ **Definition 6** (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. $\mathcal{W}_0 = \mathrm{span}(1)$ *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* $\lambda_{t_1} < \lambda_{t_2}$.

▶ **Remark 7.** 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 Proposition 19); its usage in the broader sense defined above is not standard (to our knowledge).

▶ **Remark 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 $\mathcal{W}_i$.

---

[2] 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.

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*

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

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 4, we see that Definition 9 coincides with Definition 1.

▶ **Example 10.** To provide intuition for Definition 9, a useful concrete example to consider is

- $\Omega = \{0, 1\}$ and $\pi = $ the uniform distribution on Ω, so $L^2(\Omega^n, \pi^{\otimes n})$ is the space of all real-valued functions on the Boolean cube $\{0, 1\}^n$ under the uniform distribution;
- $\mathcal{F}_{\mathrm{mon}} = $ the class of all monotone Boolean functions, i.e. all $f : \{0, 1\}^n \to \{0, 1\}$ such that if $x_i \leq y_i$ for all i then $f(x) \leq f(y)$;
- $(\mathrm{P}_t)_{t \geq 0}$ is defined by $\mathrm{P}_t = \mathrm{T}_{e^{-t}}$, where T_ρ is the Bonami-Beckner operator. We remind the reader that for any $f : \{0, 1\}^n \to \mathbb{R}$ and any $0 \leq \rho \leq 1$, the function $\mathrm{T}_\rho f(x)$ is defined to be $\mathbf{E}_{\boldsymbol{y} \sim N_\rho(x)}[f(\boldsymbol{y})]$, where "$\boldsymbol{y} \sim N_\rho(x)$" means that $\boldsymbol{y} \in \{0, 1\}^n$ is randomly chosen by independently setting each $\boldsymbol{y}_i$ to be x_i with probability ρ and to be uniform random with probability $1 - \rho$.

In this setting, as will be shown later, we have that for any two monotone functions $f, g \in \mathcal{F}$, the function $\mathbf{E}_{\boldsymbol{x} \sim \{\pm 1\}^n}[\mathrm{T}_\rho f(\boldsymbol{x}) g(\boldsymbol{x})]$ is a non-decreasing function of ρ; hence $\frac{\partial}{\partial t} \mathbf{E}_{\boldsymbol{x} \sim \{\pm 1\}^n}[\mathrm{P}_t f(\boldsymbol{x}) g(\boldsymbol{x})]$ is always at most 0 (note that as t increases $\rho = e^{-t}$ decreases), so $(\mathrm{P}_t)_{t \geq 0}$ is monotone compatible with $\mathcal{F}_{\mathrm{mon}}$.

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 [24].

▶ **Theorem 11** (Hadamard Three Circles Theorem). *Suppose f is holomorphic on the annulus* $\{z \in \mathbb{C} \mid r_1 \leq |z| \leq r_2\}$. *For* $r \in [r_1, r_2]$, *let* $M(r) := \max_{|z| = r} |f(z)|$. *Then*

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

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:[3]

---

[3]   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 12.

▶ **Lemma 12** (Main technical lemma). *Let $p(t) = \sum_{i=1}^{\infty} c_i t^i$ with $c_1 = 1$ and $\sum_{i=1}^{\infty} |c_i| \leq M$ where $M \geq 3/2$. Then:*

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

The proof given below is inspired by arguments with a similar flavor in [3, 4], 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:

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

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:

1. 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 $\left[\delta, \frac{16+9\delta}{25}\right] \subseteq [\delta, 1)$.
2. 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.
3. 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 $\mathbb{C}\backslash\{0\}$; in particular, it is holomorphic on the annulus defined by C_1 and C_3. Consequently, by Theorem 11, 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 12 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 wonder whether this might be the smallest achievable value. However, it turns out that the $1/\log^2 M$ dependence of Lemma 12 is in fact the best possible result; a proof of the following result can be found in the appendix to the full version of this paper.

▷ **Claim 13.** For sufficiently large M, there exists a real polynomial $p(t) = \sum_{i=1}^{d} c_i t^i$ with $c_1 = 1$ and $\sum_{i=1}^{d} |c_i| \le M$ such that

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

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 14, 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 Section 6 deals with finite product domains.

▶ **Theorem 14** (Main Theorem). *Consider a symmetric Markov semigroup* $(\mathrm{P}_t)_{t \ge 0}$ *on* $L^2(\Omega^n, \Pi^{\otimes n})$ *with a chaos decomposition*

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

Let $(\mathrm{P}_t)_{t \ge 0}$ *be monotone compatible with* $\mathcal{F} \subseteq L^2(\Omega^n, \Pi^{\otimes n})$, *where* $\|f\| \le 1$ *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*

$$\langle f, g \rangle - f_0 g_0 \ge \frac{1}{C} \cdot \Phi\left(\langle f_{j^*}, g_{j^*} \rangle\right), \tag{6}$$

where recall from Equation (2) *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 12 to obtain the lower bound.

Proof of Theorem 14. Fix $f, g \in \mathcal{F}$ and let us write $a_\ell := \langle f_\ell, g_\ell \rangle$. It follows from Definition 6 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

$$\langle \mathrm{P}_t f, g \rangle = \sum_{j^* | \ell} \lambda_t^\ell \langle f_\ell, g_\ell \rangle = \sum_{j^* | \ell} a_\ell \lambda_t^\ell. \tag{7}$$

Here we remark that the argument to $\Phi(\cdot)$ in the right hand side of Equation (6) is non-negative, i.e. $a_{j^*} \geq 0$. To see this, observe that

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

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

Returning to Equation (7), rearranging terms gives that

$$\langle \mathrm{P}_t f, g \rangle - f_0 g_0 = \sum_{\substack{\ell > 0 \\ j^* | \ell}} a_\ell \lambda_t^\ell = a_{j^*} p(\lambda_t^{j^*}), \qquad \text{where} \qquad p(\lambda_t^{j^*}) := \lambda_t^{j^*} + \frac{1}{a_{j^*}} \sum_{\substack{\ell > j^* \\ j^* | \ell}} a_\ell \lambda_t^\ell. \quad (8)$$

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

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

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 12 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 6 that as t ranges over $(0, \infty)$, λ_t and consequently u ranges over the interval $(0, 1]$. Together with Equation (8), 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   A Quantitative Extension of the Gaussian Correlation Inequality

In this section, we prove a quantitative versions of Royen's Gaussian Correlation Inequality (GCI) [23] for symmetric convex sets. We start by recalling some elementary facts about harmonic analysis over Gaussian space in Section 5.1, after which we derive our "robust" form of the Gaussian Correlation Inequality in Section 5.2 as a consequence of Theorem 14. In Section 5.3 we discuss how our robust GCI can be viewed as a Gaussian-space analogue of Talagrand's celebrated correlation inequality for monotone Boolean functions over the Boolean hypercube [26]. We analyze the tightness of our robust GCI in Section 5.4.

5.1   Harmonic (Hermite) Analysis over Gaussian space

Our notation and terminology presented in this subsection follows Chapter 11 of [21]. We say that an n-dimensional *multi-index* is a tuple $\alpha \in \mathbb{N}^n$, and we define

$$\operatorname{supp}(\alpha) := \{i : \alpha_i \neq 0\}, \qquad \#\alpha := |\operatorname{supp}(\alpha)|, \qquad |\alpha| := \sum_{i=1}^{n} \alpha_i. \tag{9}$$

We write $\mathcal{N}(0,1)^n$ to denote the n-dimensional standard Gaussian distribution. For $n \in \mathbb{N}_{>0}$, we write $L^2(\mathbb{R}^n, \gamma)$ to denote the space of functions $f : \mathbb{R}^n \to \mathbb{R}$ that have finite 2$^{\text{nd}}$ moment $\|f\|_2^2$ under the standard Gaussian measure γ, that is:

$$\|f\|_2^2 = \operatorname*{\mathbf{E}}_{\boldsymbol{z} \sim \mathcal{N}(0,1)^n} \left[f(\boldsymbol{z})^2 \right]^{1/2} < \infty.$$

We view $L^2(\mathbb{R}^n, \gamma)$ as an inner product space with $\langle f, g \rangle := \mathbf{E}_{\boldsymbol{z} \sim \mathcal{N}(0,1)^n}[f(\boldsymbol{z})g(\boldsymbol{z})]$ for $f, g \in L^2(\mathbb{R}^n, \gamma)$. We recall the "Hermite basis" for $L^2(\mathbb{R}, \gamma)$:

▶ **Definition 15** (Hermite basis). *The* Hermite polynomials $(h_j)_{j \in \mathbb{N}}$ *are the univariate polynomials defined as*

$$h_j(x) = \frac{(-1)^j}{\sqrt{j!}} \exp\left(\frac{x^2}{2}\right) \cdot \frac{d^j}{dx^j} \exp\left(-\frac{x^2}{2}\right).$$

▶ **Proposition 16** (Proposition 11.33, [21]). *The Hermite 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.

▶ **Proposition 17** (Plancharel's identity). *For $f, g \in L^2(\mathbb{R}^n, \gamma)$, we have:*

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

and as a special case we have Parseval's identity,

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

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

▶ **Definition 18** (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*

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

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:

▶ **Proposition 19** (Proposition 11.33, [21])**.** *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) [23]. 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,

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

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

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

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 [20] for further details):

▶ **Proposition 20** (Royen's Theorem, [23])**.** *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*

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

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 4. We can now state our main result:

▶ **Theorem 21** (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*

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

where recall from Equation (2) 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 $\mathcal{W}_i = \text{span}\{h_\alpha : |\alpha| = i\}$; the orthogonality of this decomposition follows from Proposition 16. From Proposition 19, it follows that this decomposition is in fact a chaos decomposition (recall Definition 6) with respect to the Ornstein-Uhlenbeck semigroup $(\mathrm{U}_\rho)_{\rho \in [0,1]}$.

If $K \in \mathcal{F}_{\mathrm{csc}}$, 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 Proposition 17, we have that

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

It follows that the hypotheses of Theorem 14 hold for $\mathcal{F}_{\mathrm{csc}}$ with $j^* = 2$; consequently, for $K, L \in \mathcal{F}_{\mathrm{csc}}$ we have

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

which completes the proof of the theorem. ◄

It is natural to ask whether Theorem 21 can be extended to a broader class of functions than $0/1$-valued indicator functions of centrally symmetric, convex sets $\mathcal{F}_{\mathrm{csc}}$. Indeed, the GCI implies the monotone compatibility of centrally symmetric, *quasiconcave*[4], non-negative functions (which is a larger family of functions than $\mathcal{F}_{\mathrm{csc}}$) with the Ornstein-Uhlenbeck semigroup. This allows us to once again use Theorem 14 to obtain a quantitative correlation inequality for this family of functions; we defer this to the full version of this paper.

5.3   Interpreting Theorem 21

Recall Talagrand's correlation inequality [26]: If $f, g : \{0, 1\}^n \to \{0, 1\}$ are monotone Boolean functions, then

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

where $\Psi(x) = \frac{x}{\log(e/x)}$. However (see Chapter 2 of [21]), for monotone $f : \{+1, -1\}^n \to \{+1, -1\}$, we have $\widehat{f}(i) = \mathbf{Inf}_i[f]$ where

$$\mathbf{Inf}_i[f] := \Pr_{\boldsymbol{x} \sim \{+1, -1\}^n}\left[f(\boldsymbol{x}) \neq f\left(\boldsymbol{x}^{\oplus i}\right)\right].$$

---

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

In other words, the degree-1 Fourier coefficient $\widehat{f}(i)$ captures the "dependence" of f on its ith coordinate, and the quantity $\sum_{i=1}^{n} \widehat{f}(i)\widehat{g}(i)$ captures the extent to which "both f and g simultaneously depend on the same coordinates". This intuitively explains why it is plausible for such a quantity to appear in Talagrand's inequality.

Inspired by the resemblance between our quantitative Gaussian correlation inequality 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. However, while functions on the Boolean hypercube have influences only along n "directions", there are infinitely many directions over Gaussian space. We make the following definition:

▶ **Definition 22** (Influences for $\mathcal{F}_{\text{csc}}$). *Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric, convex set. Given a unit vector $v \in S^{n-1}$, we define the* influence of K *along direction v as*

$$\mathbf{Inf}_v[K] := -\widetilde{K}(2v) = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathcal{N}(0,1)^n} [-K(\boldsymbol{x})h_2(v \cdot \boldsymbol{x})]$$

where $h_2(x) = \frac{x^2-1}{\sqrt{2}}$ is the degree-2 univariate Hermite polynomial (see Section 11.2 of [21]).

It follows from the proof of Theorem 14 that the quantity $\sum_{|\alpha|=2} \widetilde{K}(\alpha)\widetilde{L}(\alpha)$ for $K, L \in \mathcal{F}_{\text{csc}}$ is non-negative. In fact more is true: if K is a centrally symmetric, convex set, then each $\mathbf{Inf}_{e_i}[K]$ is itself non-negative. The proof of the following proposition can be found in the appendix to the full version of this paper.

▶ **Proposition 23** (Influences are non-negative). *If K is a centrally symmetric, convex set, then $\mathbf{Inf}_v[K] \geq 0$ for all $v \in S^{n-1}$, with equality holding if and only if $K(x) = K(y)$ whenever $x_{v^\perp} = y_{v^\perp}$ (the projection of x orthogonal to v coincides with that of y), except possibly on a set of measure zero.*

It is natural to define the "total influence of K" to be $\mathbf{Inf}[K] := \sum_{i=1}^{n} \mathbf{Inf}_{e_i}[K]$; we observe that this quantity is given by

$$\mathbf{Inf}[K] = -\sum_{i=1}^{n} \widetilde{K}(2e_i) = \frac{\mathbf{E}_{\boldsymbol{x} \sim \mathcal{N}(0,1)^n} \left[-f(\boldsymbol{x}) \cdot (\|\boldsymbol{x}\|^2 - n) \right]}{\sqrt{2}},$$

and hence it is invariant under orthogonal transformations (i.e. any orthonormal basis $v_1, \ldots, v_n$ could have been used in place of $e_1, \ldots, e_n$ in defining $\mathbf{Inf}[K]$).

The above discussion suggests that the notion of "influences" for centrally symmetric, convex sets in Gaussian space proposed in Definition 22 is indeed "influence-like". A forthcoming paper [5] will further explore this notion.

5.4 On the Tightness of Theorem 21

In [26], Talagrand gave the following family of example functions for which Equation (1) is tight up to constant factors: let $f, g : \{0,1\}^n \to \{0,1\}$ be given by

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

where $k \leq n/2$. Writing ϵ to denote $\mathbf{E}[f]$, we have $\epsilon^2 = \epsilon - \epsilon(1-\epsilon) = \mathbf{E}[fg] - \mathbf{E}[f]\mathbf{E}[g]$, and it can be shown that $\Psi\left(\sum_{i=1}^{n} \widehat{f}(i)\widehat{g}(i)\right) = \Theta(\epsilon^2)$, so Equation (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. In particular, let $K, L \in \mathcal{F}_{\mathrm{csc}}$ be n-dimensional origin-centered balls of radii r_1 and r_2 respectively such that $r_1 < r_2$ with

$$\mathbf{E}[K] = \epsilon \qquad \text{and} \qquad \mathbf{E}[L] = 1 - \epsilon$$

where the expectations are taken with respect to the n-dimensional Gaussian measure. As $K \subseteq L$, we have $\mathbf{E}[KL] - \mathbf{E}[K]\mathbf{E}[L] = \epsilon - \epsilon(1 - \epsilon) = \epsilon^2$. 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. 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 $i, j \in [n]$. It follows that

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

An application of the Berry-Esseen Central Limit Theorem (see [2, 6] or, for example, Section 11.5 of [21]) together with standard anti-concentration bounds on Gaussian tails (we omit the details here; a complete calculation can be found in the full version of this paper) gives that

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

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

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

which lets us conclude that Theorem 21 is tight to within a logarithmic factor.

6   Generalizing Talagrand's Inequality to Arbitrary Finite Product Domains

The main result of this section, Theorem 29, is an extension of Talagrand's correlation inequality [26] 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 [21]. We use multi-index notation for $\alpha \in \mathbb{N}^n$ as defined in Equation (9).

Let (Ω, π) be a finite probability space with $|\Omega| = m \geq 2$, 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 [21]). So every function $f : \Omega^n \to \mathbb{R}$ has a decomposition

$$f = \sum_{\alpha \in \mathbb{N}^n_{<m}} \widehat{f}(\alpha)\phi_\alpha. \tag{10}$$

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 [21]). 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*

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

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.

▶ **Proposition 25** (Proposition 8.28 of [21])). *We have $\mathrm{T}_\rho f = \sum_\alpha \rho^{\#\alpha} \widehat{f}(\alpha)\phi_\alpha$.*

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 14, 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 26** (coordinate-wise noise operators). *Let T_ρ^i be the operator on functions $f : \Omega^n \to \mathbb{R}$ defined by*

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

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

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

▶ **Lemma 27.** *Let $\Omega = \{0, 1, \ldots, m-1\}$ and let $f : \Omega^n \to \mathbb{R}$ be a monotone function. Then $\mathrm{T}_\rho^i f : \Omega^n \to \mathbb{R}$ is a monotone function.*

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

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

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

$$\underset{\boldsymbol{z} \sim N_\rho(x_i)}{\mathbf{E}} \left[f\big(x^{i \mapsto \boldsymbol{z}}\big) \right] = \delta f(x) + (1 - \delta) \underset{\boldsymbol{z} \sim \Omega^n}{\mathbf{E}} \left[f\big(x^{i \mapsto \boldsymbol{z}}\big) \right]$$

$$\le \delta f(y) + (1 - \delta) \underset{\boldsymbol{z} \sim \Omega^n}{\mathbf{E}} \left[f\big(y^{i \mapsto \boldsymbol{z}}\big) \right] = \underset{\boldsymbol{z} \sim N_\rho(y_i)}{\mathbf{E}} \left[f\big(y^{i \mapsto \boldsymbol{z}}\big) \right]. \qquad \blacktriangleleft$$

▶ **Lemma 28.** *Let $\Omega = \{0, 1, \ldots, m - 1\}$ and let $f, g : \Omega^n \to \mathbb{R}$ be monotone functions. Then $\langle T_\rho f, g \rangle$ is nondecreasing in $\rho \in [0, 1]$.*

Proof. We have

$$\langle T_{\rho_1, \ldots, \rho_n} f, g \rangle = \langle T_{\rho, 1, \ldots, 1} f, T_{1, \rho_2, \ldots, \rho_n} g \rangle = \langle T_{\rho_1}^1 f, h \rangle$$

where $h := T_{1, \rho_2, \ldots, \rho_n} g$. It follows from a repeated application of Lemma 27 that h is monotone. Now, note that

$$\langle T_{\rho_1}^1 f, h \rangle = \widehat{f}(\bar{0}) \cdot \widehat{h}(\bar{0}) + \sum_{\alpha_1 > 0} \rho_1 \widehat{f}(\alpha) \widehat{h}(\alpha) + \sum_{\substack{\bar{0} \ne \alpha \\ \alpha_1 = 0}} \widehat{f}(\alpha) \widehat{h}(\alpha)$$

where $\bar{0} = (0, \ldots, 0)$. By Cheybshev's order inequality, we know that $\langle T_1^1 f, h \rangle \ge \langle T_0^1 f, h \rangle = \widehat{f}(\bar{0}) \cdot \widehat{h}(\bar{0}) + \sum_{\bar{0} \ne \alpha, \alpha_1 = 0} \widehat{f}(\alpha) \widehat{h}(\alpha)$. From the above expression, we have:

$$\frac{\partial}{\partial \rho_1} \langle T_{\rho_1}^1 f, h \rangle = \sum_{\alpha_1 > 0} \widehat{f}(\alpha) \widehat{h}(\alpha)$$

which must be nonnegative since $\langle T_1^1 f, h \rangle \ge \langle T_0^1 f, h \rangle$, and so we can conclude that $\langle T_{\rho_1}^1 f, h \rangle$ is nondecreasing in ρ_1. The result then follows by repeating this for each coordinate. $\qquad \blacktriangleleft$

Let $\mathcal{F}_{\mathrm{mon}} \subseteq L^2(\Omega^n, \pi^{\otimes n})$ be the family of monotone functions $f : \Omega^n \to \mathbb{R}$. Then Lemma 28 shows that $\mathcal{F}_{\mathrm{mon}}$ is monotone compatible with the Bonami–Beckner noise operator. We can now prove our Talagrand-analogue for monotone functions over Ω^n:

▶ **Theorem 29.** *Let $\Omega = \{0, 1, \ldots, m - 1\}^n$ and let $\mathcal{F}_{\mathrm{mon}} \subseteq L^2(\Omega^n, \pi^{\otimes n})$ denote the family of monotone functions on Ω^n such that $\|f\| \le 1$ for all $f \in \mathcal{F}_{\mathrm{mon}}$. Then for $f, g \in \mathcal{F}_{\mathrm{mon}}$, 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 Equation (2) 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 $\mathcal{W}_i = \mathrm{span} \{\phi_\alpha : \# \alpha = i\}$; the orthogonality of this decomposition follows from the orthonormality of $(\phi_\alpha)_{\alpha \in \mathbb{N}_{<m}^n}$. Furthermore, this decomposition is a chaos decomposition with respect to the Bonami–Beckner operator $(T_\rho)_{\rho \in [0,1]}$. It follows that the hypotheses of Theorem 14 hold for $\mathcal{F}_{\mathrm{mon}}$ with $j^* = 1$, from which the result follows. $\qquad \blacktriangleleft$

Theorem 29 can be interpreted in terms of the *Efron–Stein decomposition* of a function (see Chapter 8 of [21]); a complete discussion of this can be found in the full version of this paper.

6.3 Comparison with Keller's Inequality for the p-biased Hypercube

In this subsection, we restrict our attention to the p-biased hypercube, i.e. $\{+1, -1\}_p^n := (\{+1, -1\}^n, \pi_p^{\otimes n})$ where $\pi_p(-1) = p$ and $\pi_p(+1) = 1 - p$. In this setting our Theorem 29 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 [15, 16] 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 30** (Theorem 7 of [15]; see also [17] 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 $|f(x)|, |g(x)| \leq 1$. Then*

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

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 Section 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 29 to Theorem 30, 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 $1/\log\left(1/\sum_i \widehat{f_p}(i)\widehat{g_p}(i)\right)$). Theorem 29 can be significantly stronger than Theorem 30 in a range of natural settings because of these differences. We show that for every $\omega(1)/n \leq p \leq 1/2$, 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 29 is at least an absolute constant independent of n and p; but (iii) the RHS of Theorem 30 is $\Theta(p \log(1/p))$. A proof can be found in the appendix to the full version of this paper.

──── **References** ────

1 D. Bakry, I. Gentil, and M. LeDoux. *Analysis and Geometry of Markov Diffusion Operators.* Springer, 2013. URL: `https://books.google.com/books?id=tf37sgEACAAJ`.

2 Andrew C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.

3 Peter Borwein and Tamás Erdélyi. Littlewood-type polynomials on subarcs of the unit circle. *Indiana University Mathematics Journal*, 46(4):1323–1346, 1997.

4 Peter Borwein, Tamás Erdélyi, and Géza Kós. Littlewood-type problems on $[0, 1]$. *Proceedings of the London Mathematical Society*, 3(79):22–46, 1999.

5 A. De, S. Nadimpalli, and R. Servedio. Influences for Centrally Symmetric, Convex Sets. In preparation, 2020.

6 Carl-Gustav Esseen. On the Liapunoff limit of error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, A:1–19, 1942.

7 C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22(2):89–103, 1971. URL: `https://projecteuclid.org:443/euclid.cmp/1103857443`.

8   G. Gallavotti. A proof of the Griffiths inequalities for the XY model. *Stud. Appl. Math*, 50(1):89–92, 1971.

9   R. Griffiths. Correlations in Ising ferromagnets. I. *Journal of Mathematical Physics*, 8(3):478–483, 1967.

10   T.E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Camb. Phil. Soc.*, 56:13–20, 1960.

11   Yaozhong Hu. Itô-wiener chaos expansion with exact residual and correlation, variance inequalities. *Journal of Theoretical Probability*, 10(4):835–848, 1997.

12   G. Kalai, N. Keller, and E. Mossel. On the correlation of increasing families. *Journal of Combinatorial Theory, Series A*, 144, November 2015. `doi:10.1016/j.jcta.2016.06.012`.

13   N. Keller. Lower bound on the correlation between monotone families in the average case. *Advances in Applied Mathematics*, 43(1):31–45, 2009.

14   N. Keller, E. Mossel, and A. Sen. Geometric Influences II: Correlation Inequalities and Noise Sensitivity. *Ann. Inst. H. Poincaré Probab. Statist.*, 50(4):1121–1139, November 2014. `doi:10.1214/13-AIHP557`.

15   Nathan Keller. Improved FKG Inequality for product measures on the discrete cube, 2008.

16   Nathan Keller. *Influences of variables on Boolean functions*. PhD thesis, Hebrew University of Jerusalem, 2009.

17   Nathan Keller. A simple reduction from a biased measure on the discrete cube to the uniform measure. *European Journal of Combinatorics*, 33:1943–1957, 2012.

18   D. Kelly and S. Sherman. General Griffiths' inequalities on correlations in Ising ferromagnets. *Journal of Mathematical Physics*, 9(3):466–484, 1968.

19   Daniel J Kleitman. Families of non-disjoint subsets. *Journal of Combinatorial Theory*, 1(1):153–155, 1966.

20   Rafał Latała and Dariusz Matlak. Royen's Proof of the Gaussian Correlation Inequality. *Geometric Aspects of Functional Analysis*, pages 265–275, 2017. `doi:10.1007/978-3-319-45282-1_17`.

21   R. O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

22   Christopher J. Preston. A generalization of the FKG inequalities. *Communications in Mathematical Physics*, 36(3):233–241, 1974.

23   Thomas Royen. A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions. *arXiv preprint*, 2014. `arXiv:1408.1028`.

24   Walter Rudin. *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., 1987.

25   M. Suzuki. Correlation inequalities and phase transition in the generalized X-Y model. *Journal of Mathematical Physics*, 14(7):837–838, 1973.

26   M. Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.