Binary Maximal Correlation Bounds and Isoperimetric Inequalities via Anti-Concentration

Dror Drach
Tel Aviv University
dror.drach@gmail.com

Or Ordentlich
Hebrew University of Jerusalem
or.ordentlich@mail.huji.ac.il

Ofer Shayevitz
Tel Aviv University
ofersha@eng.tau.ac.il

***Abstract*—This paper establishes a dimension-independent upper bound on the maximal correlation between Boolean functions of dependent random variables, in terms of the second and third singular values in their spectral decomposition, and the anti-concentration properties of the second singular vectors. This result has notable consequences, among which are: A strengthening of Witsenhausen's lower bound on the probability of disagreement between Boolean functions; a Poincaré inequality for bounded-cardinality functions; and improved lower bounds on the isoperimetric constant of Markov chains.**

I. Introduction

Studying the probability of disagreement between Boolean functions of dependent random variables arises as a fundamental problem in many mathematical disciplines [1, 2, 3, 4]. Often, the starting point for such a study is relating the disagreement probability to the *correlation* between the two functions. Specifically, for a pair of r.v.s (X, Y), the probability of disagreement between functions $f : \mathcal{X} \to \{0, 1\}$ and $g : \mathcal{Y} \to \{0, 1\}$ can be lower bounded by [1]

$$\Pr(f(X)\neq g(Y)) \geq 2\sqrt{p(1-p)q(1-q)}(1-\rho(f,g)), \quad (1)$$

where $\rho(f, g)$ is the correlation between $f(X)$ and $g(Y)$, and $p = \mathbb{E}f(X)$, $q = \mathbb{E}g(Y)$ are the associated biases.

Since uniform lower bounds are often desired, it is natural to define the *binary maximal correlation* between X and Y:

$$\rho_b(X;Y) \triangleq \max_{f,g} \mathbb{E}f(X)g(Y) \quad (2)$$

where the maximization is subject to $\mathbb{E}f(X) = Eg(Y) = 0$, $\mathbb{E}f^2(X) = Eg^2(Y) = 1$, and $|\text{range}(f)| = |\text{range}(g)| = 2$. Since $\rho(f, g) \leq \rho_b(X; Y)$, one can substitute the latter into (1) to obtain a uniform lower bound on $\Pr(f(X) \neq g(Y))$.

The special case where $f = g$ is of particular interest, being intimately related to the isoperimetric problem for Markov chains. Indeed, if W is a reversible Markov kernel over the discrete alphabet $[K]$, with invariant distribution μ, then the isoperimetric (Cheeger) constant[1] of W is [5]

$$h(W) \triangleq \min_{f:[K]\to\{-1,1\}} \frac{\Pr(f(X) \neq f(Y))}{\min(\Pr(f(X) = -1), \Pr(f(X) = 1))},$$

where $(X, Y) \sim \mu \times W$. From the discussion above, and noting that $\min\{p, 1 - p\} \leq 2p(1 - p)$, we see that the isoperimetric constant is lower bounded by

$$h(W) \geq 1 - \rho_b(X; Y), \quad (3)$$

which in fact can be shown to be tight up to a multiplicative factor of 2 in some cases, e.g., when W is a lazy reversible kernel and μ is uniform [6].

However, while the binary maximal correlation is clearly an interesting quantity to study, its finite cardinality constraint makes it difficult to estimate for large alphabets. In particular, it is not stable under tensor products, hence cannot be easily computed even in the simpler i.i.d. case,[2] which is often of much interest. These problems are typically circumvented by removing the cardinality constraint, and replacing $\rho_b(X; Y)$ with the *Hirschfeld-Gebelein-Rényi maximal correlation*:

$$\rho_m(X, Y) \triangleq \max_{f,g} \mathbb{E}[f(X)g(Y)],$$

where the maximization is subject to $\mathbb{E}f(X)=\mathbb{E}g(Y)=0$ and $\mathbb{E}f^2(X) = \mathbb{E}g^2(Y) = 1$ only. Unlike its binary counterpart, maximal correlation is easily computable via spectral methods, being equal to the second singular value in the spectral decomposition of the joint distribution. It is also stable under tensor products, in the sense that $\rho_m(X^n, Y^n) = \rho_m(X, Y)$ for i.i.d. pairs $(X^n, Y^n) \sim P_{XY}^{\otimes n}$ [1]. These properties make maximal correlation very useful, especially in high dimension. For example, inequality (1) together with tensorization, immediately imply Witsenhausen's lower bound [1]:

$$\Pr(f(X^n) \neq g(Y^n)) \geq 2\sqrt{p(1-p)q(1-q)}(1-\rho_m(X,Y)) \quad (4)$$

One can further use this to derive the famous isoperimetric inequality [3] $h(W^n) \geq n^{-1}(1 - \rho_m(X; Y))$, where W^n is the n-fold Cartesian product of W (defined in Section IV-C).

In this paper, we derive an upper bound on the binary maximal correlation, using spectral methods combined with anti-concentration techniques. In particular, our bound depends on the second and third singular values in the spectral decomposition, as well as on the concentration function of the singular vectors pertaining to the functions attaining the maximal correlation. More importantly, we derive a dimension-independent version of our bound, by estimating the concentration function of scaled tensor products of the aforementioned singular vectors. Our bounds are easy to compute, and are always at least as good, and typically improve upon, the trivial upper bound $\rho_b(X^n; Y^n) \leq \rho_m(X; Y)$.

These results have several notable consequences described in this paper: A strengthening of Witsenhausen's lower bound (4) on the probability of disagreement between Boolean functions; a dimension-independent Poincaré inequality for bounded-cardinality functions over reversible semigroups; and improved lower bounds on the isoperimetric (Cheeger) con-

[1]In the literature, the Cheeger constant is usually defined as $h(W)/2$.

[2]Witsenhausen [1] hinted that ρ_b might not generally tensorize. An unpublished manuscript [7] suggested that it does, but, unfortunately, the proof of [7, Theorem 2] contains a subtle error. Bradley [8] gave a counterexample confirming that ρ_b indeed does not tenzorize.

978-1-5386-8209-8/21/$31.00 ©2021 IEEE

stant of reversible Markov chains, with a version that is stable under Cartesian products.

II. Preliminaries

A. Maximal correlation with cardinality constraints

While we are mostly interested in the binary maximal correlation, we will prove our results in a more general setting where the functions f and g are restricted to output one of M and N distinct values, respectively. To that end, we define the (M,N)-*quantized maximal correlation* to be

$$\rho_m^{(M,N)}(X;Y) \triangleq \max_{f,g} \mathbb{E}[f(X)g(Y)],$$

where the maximization is subject to $\mathbb{E}f(X) = Eg(X) = 0$, $\mathbb{E}f^2(X) = Eg^2(X) = 1$, $|\text{range}(f)| \le M$ and $|\text{range}(g)| \le N$. Note that $\rho_b(X;Y) = \rho_m^{(2,2)}(X;Y)$, and $\rho_m^{(\infty,\infty)}(X;Y) = \rho_m(X;Y)$. We also write $\rho_m^{(\infty,M)}(X;Y)$ for the *one-sided quantized maximal correlation*, where there is no restriction on the cardinality of X.

B. Spectral decomposition of joint distributions

Let $(X,Y) \sim P_{XY}$. It is known [9] (see also [10]) that under some mild regularity conditions there are (possibly countably infinite) orthonormal sets of singular functions $\{f_i : \mathcal{X} \to \mathbb{R}\}_{i\ge1}$ and $\{g_j : \mathcal{Y} \to \mathbb{R}\}_{j\ge1}$ spanning $L^2(P_X)$ and $L^2(P_Y)$ respectively, satisfying[3]

$$\mathbb{E}[f_i(X)f_j(X)] = \mathbb{E}[g_i(Y)g_j(Y)] = \mathbb{1}(i=j)$$
$$\mathbb{E}[f_i(X)g_j(Y)] = \sigma_i \mathbb{1}(i=j),$$

for nonnegative singular values $1 = \sigma_1 \ge \sigma_2 \ge \cdots$, where $f_1 = g_1 \equiv 1$ and $\sigma_2 = \rho_m(X;Y)$. For finite alphabets, $\sigma_i = 0$ for any $i > \min\{|\mathcal{X}|,|\mathcal{Y}|\}$, and $f_i \equiv g_j \equiv 0$ for any $i > |\mathcal{X}|$ and $j > |\mathcal{Y}|$.

Let $(X^n,Y^n) \sim P_{XY}^{\otimes n}$. Then for any vector in $u = (u_1,...,u_n) \in \mathbb{N}^n$, define the functions

$$f_u(X^n) = \prod_{i=1}^n f_{u_i}(X_i), \quad g_u(Y^n) = \prod_{i=1}^n g_{u_i}(Y_i)$$

It is immediate to verify that $\{f_u\}_{u\in\mathbb{N}^n}$ and $\{g_u\}_{u\in\mathbb{N}^n}$ are orthonormal sets that span $L^2(P_X^{\otimes n})$ and $L^2(P_Y^{\otimes n})$ respectively, and furthermore,

$$\mathbb{E}[f_u(X^n)g_v(Y^n)] = \prod_{i=1}^n \sigma_{u_i}\mathbb{1}(u_i = v_i) = \sigma_u \mathbb{1}(u=v),$$

where $\sigma_u = \prod_{i=1}^n \sigma_{u_i}$. Therefore, any two functions $f \in L^2(P_X^{\otimes n})$ and $g \in L^2(P_Y^{\otimes n})$ can be written as

$$f(X^n) = \sum_{u\in\mathbb{N}^n} a_u f_u(X^n), \quad g(Y^n) = \sum_{u\in\mathbb{N}^n} b_u g_u(Y^n),$$

such that

$$\mathbb{E}[f(X^n)g(Y^n)] = \sum_{u\in\mathbb{N}^n} a_u b_u \sigma_u. \tag{5}$$

It is now easy to see that $\rho_m(X^n;Y^n) = \rho_m(X;Y) = \sigma_2$, and also that the scalar functions $f_2(X_k), g_2(Y_k)$ achieve the maximal correlation for any k.

[3]These orthonormal sets exist for any joint distribution with finite χ^2-information. In particular, this always holds for discrete alphabets, in which case the expansion corresponds to the standard singular value decomposition of the matrix with entries $P_{XY}(x,y)/\sqrt{P_X(x)P_Y(y)}$.

C. Anti-concentration

The *concentration function* of an r.v. X is defined as

$$Q(X,t) \triangleq \sup_{a\in\mathbb{R}} \Pr(a \le X \le a+t), \quad t \ge 0.$$

An *anti-concentration inequality* is an upper bound on $Q(X,t)$, indicating that X is not too concentrated in any interval. Such inequalities will play a central role in our proofs, essentially since a gap between ρ_m and ρ_b can often be attributed to the fact that the functions f_2, g_2 achieving ρ_m are far from being concentrated on only two values.

Throughout the paper, we make use of the following simple properties of $Q(X,t)$ [11]: 1) $t \mapsto Q(X,t)$ is non-decreasing; 2) $Q(\alpha X,t) = Q(X,t/\alpha)$ for any $\alpha > 0$; and 3) if X and Y are independent, then

$$Q(X+Y,t) \le \min\{Q(X,t),Q(Y,t)\}. \tag{6}$$

Since we are interested in dimension-independent bounds, and since the maximal correlation in n dimensions can be attained by any linear combination of the scalar functions f_2, g_2 applied to the coordinates, we are particularly interested in anti-concentration inequalities for weighted sums of i.i.d. r.vs. To that end, we introduce the following quantities. First, we define the n-*fold concentration function* of X to be

$$A_n(X,t) \triangleq \sup_{\mathbf{a}\in\mathbb{R}^n:\|\mathbf{a}\|_2=1} Q\left(\sum_{i=1}^n a_i X_i, t\right), \tag{7}$$

where $X_1,\dots,X_n$ are i.i.d copies of X. Then, we define the *asymptotic concentration function* of X to be

$$A(X,t) \triangleq \lim_{n\to\infty} A_n(X,t), \tag{8}$$

where the limit exists since $A_n(X,t)$ is non-decreasing in n.

Upper-bounding $A_n(X,t)$ is closely related to the Littlewood-Offord problem [12, 13], which is concerned with the maximal possible value of $Q(\sum \alpha_i X_i, t)$ for $t = 0$, where X_i are i.i.d $\text{Ber}(1/2)$ and $|\alpha_i| \ge 1$. This problem and its variations have been extensively studied in additive combinatorics [14]. Here however we are mostly interested in the moderate t regime. Furthermore, several works considered bounding $A(X,t)$ as a function of $Q(X_i,t)$ and related quantities, e.g., by Kolmogorov-Rogozin [15, 16] and Esseen [17]. One such bound is due to [18, Corollary 1.4] and yields

$$A_n(X,t) \le C_1 Q(X,t/2), \tag{9}$$

where $C_1 = \frac{12}{11}4\sqrt{2}$.

For the purpose of this paper, we will rely both on (9), as well as on a bound of a different flavor that we establish based on [17]: Let $X^* = X - X'$ be the symmetrized version of X, where X' is an independent copy of X. For any $t > 0$, define

$$D(X,t) \triangleq \mathbb{E}\min\left(X^2/t^2, 1\right), \tag{10}$$

where $D(X,0) = \Pr(X \ne 0)$.

Lemma 1. *Let* $t_\tau \triangleq Q(X,\tau)\sqrt{D(X^*;\tau)}C_2^{-1}\tau$, *where* $C_2 = (96/95)^2\sqrt{48\pi/11} \approx 3.7809$. *Then*

$$A_n(X,t_\tau) \le A(X,t_\tau) \le Q(X,\tau). \tag{11}$$

The proof for Lemma 1 appears in section V.

III. Main technical results

Let $(X,Y) \sim P_{XY}$, and let $f_2(x)$ and $g_2(y)$ be the singular functions that attain σ_2, where $\sigma_2 = \rho_m(X;Y)$ and σ_3

are respectively the second-largest and third-largest singular values in the decomposition of P_{XY}. We define the gap between these singular values as

$$\Delta \triangleq \sigma_2^2 - \max\{\sigma_2^4, \sigma_3^2\}.$$

Note that $\Delta > 0$ if and only if $\sigma_3 < \sigma_2 < 1$. Let

$$\theta_M^{(n)}(X) \triangleq \max_t \frac{t^2}{4} \left|1 - M \cdot A_n(X,t)\right|_+ , \quad (12)$$

be the anti-concentration coefficient of X in dimension n, where $|x|_+ \triangleq \max\{x, 0\}$. For any $a, b \in [0,1]$, define

$$d(a,b) \triangleq \max_{\substack{x,y \\ |x|\le a, |y|\le b}} xy\sigma_2 + \sqrt{(1-x^2)(1-y^2)} \max\{\sigma_2^2, \sigma_3\} \quad (13)$$

Theorem 1. *Let $f_2 = f_2(X)$ and $g_2 = g_2(Y)$. Then*

$$\rho_m^{(M,N)}(X^n;Y^n) \le d\left(|1+\theta_M^{(n)}(f_2)|^{-\frac{1}{2}}, |1+\theta_N^{(n)}(g_2)|^{-\frac{1}{2}}\right).$$

In particular,

$$\left|\rho_m^{(M,\infty)}(X^n;Y^n)\right|^2 \le |\rho_m(X;Y)|^2 - \Delta \cdot \frac{\theta_M^{(n)}(f_2)}{1+\theta_M^{(n)}(f_2)}.$$

Remark 1. *In one dimension, Theorem 1, as well as its corollaries and applications, all hold in a stronger form: For $n = 1$, one should use $\Delta = \sigma_2^2 - \sigma_3^2$ and $d(a,b) \triangleq \max_{\substack{x,y \\ |x|\le a, |y|\le b}} xy\sigma_2 + \sqrt{(1-x^2)(1-y^2)}\sigma_3$. This will become clear from the proof. For the sake of simplicity, however, we ignore this refinement when stating the results, and proceed with the general definitions of Δ and $d(a,b)$.*

The proof consists of two parts: First, we establish that $\mathbb{E}f(X^n)g(Y^n) \le d(h_x, h_y)$ for general functions f and g, where h_x is the projection of the function $f(X^n)$ onto the n-dimensional subspace pertaining to $\rho_m(X;Y)$, and h_y is the respective projection of $g(Y^n)$. Second, we show that bounded cardinality functions cannot generally have a large projections on these subspaces, unless $f_2(X), g_2(Y)$ are already very concentrated.

For specific distributions (e.g., Gaussian), one could sometimes either compute or easily upper bound $A_n(X,t)$ (or $A(X,t)$) to obtain a dimension-independent result.

Example 1 (**Gaussian case**). *Let (X,Y) be a standard ρ-correlated jointly Gaussian pair. The associated singular functions are the normalized Hermite polynomials, where $f_2(x) = x, g_2(y) = y$. Also, $\sigma_2 = \rho$ and $\sigma_3 = \rho^2$. Since Gaussians are closed under convolution, we can compute the asymptotic concentration function, and Theorem 1 yields*

$$\rho_b(X^n;Y^n) \le 0.9418\rho + (1-0.9418)\rho^2. \quad (14)$$

which is significantly tighter than the trivial $\rho_b(X^n;Y^n) \le \rho$. It is however well-known that the Gaussian binary maximal correlation tensorizes, and achieved by some 1-dim threshold functions [2]. In particular, this implies that $\rho_b(X^n;Y^n) = \frac{2}{\pi}\rho + O(\rho^2)$, which is stronger than our bound for small ρ.

Theorem 1 is dependent on $\theta_M^{(n)}(X)$ and is therefore not computable for large n in general. This can be addressed by applying either (9) or Lemma 1. Define

$$\eta_M^{(1)}(X) \triangleq \max_t t^2 \left|1 - MC_1 \cdot Q(X,t)\right|_+ , \quad (15)$$

where $C_1 = \frac{12}{11}4\sqrt{2}$. By (9) we have that $\theta_M^{(n)}(X) \ge \eta_M^{(1)}(X)$, independent of n. Similarly, define

$$\eta_M^{(2)}(X) \triangleq \max_t \frac{t^2}{4C_2^2} Q^2(X,t) D(X^*_t;t) |1 - MQ(X,t)|_+ \quad (16)$$

where $C_2 = (96/95)^2\sqrt{48\pi/11} \approx 3.7809$. By Lemma 1 we have that $\theta_M^{(n)}(X) \ge \eta_M^{(2)}(X)$, independent of n.

We will proceed using the coefficient

$$\eta_M(X) \triangleq \max\{\eta_M^{(1)}(X), \eta_M^{(2)}(X)\}. \quad (17)$$

Similarly, any other bound of the appropriate nature may be used in addition to $\eta_M^{(1)}(X)$ and $\eta_M^{(2)}(X)$.

Corollary 1 (**Dimension-independent bound**). *It holds that*

$$\rho_m^{(M,N)}(X^n;Y^n) \le d\left(|1+\eta_M(f_2)|^{-\frac{1}{2}}, |1+\eta_N(g_2)|^{-\frac{1}{2}}\right),$$

$$\left|\rho_m^{(M,\infty)}(X^n;Y^n)\right|^2 \le |\rho_m(X;Y)|^2 - \Delta \cdot \frac{\eta_M(f_2)}{1+\eta_M(f_2)}.$$

The bounds above are nontrivial if $\Delta > 0$, and $f_2(X)$ (or $g_2(Y)$, if relevant) has no mass points with probability $\ge 1/M$ (or $\ge 1/N$, respectively), as is often the case.

IV. Applications

A. Improved lower bound on the probability of disagreement

Corollary 1 immediately translates to the following improvement of (4) (and therefore of [1, Theorem 2]):

$$\Pr(f(X^n) \ne g(Y^n)) \ge 2\sqrt{p(1-p)q(1-q)} \cdot (1-d^\star)$$

where $d^\star = d\left(|1+\eta_2(f_2)|^{-\frac{1}{2}}, |1+\eta_2(g_2)|^{-\frac{1}{2}}\right)$.

B. A Finite-Cardinality Poincaré inequality

Let P_t be a reversible Markov semigroup with generator $\mathcal{L}$, and unique invariant distribution μ. The associated Dirichlet form is given by [19]

$$\mathcal{E}(f) \triangleq -\langle f, \mathcal{L}f\rangle_\mu. \quad (18)$$

For simplicity, we assume that the Dirichlet form above exists for any $f \in L^2(\mu)$ (which is true in the discrete case). The *Poincaré constant of cardinality M of P_t* is defined to be

$$C_M(P_t) \triangleq \sup_{f: 2\le|\mathrm{range}(f)|\le M} \frac{\mathsf{Var}_\mu(f)}{\mathcal{E}(f)}, \quad (19)$$

and so a Poincaré inequality $\mathsf{Var}_\mu(f) \le C_M \cdot \mathcal{E}(f)$ holds for all functions f with cardinality at most M. The case $C_\infty(P_t)$ corresponds to the standard Poincaré inequality.

Let X_t be the stationary Markov process generated by P_t with $X_0 \sim \mu$, i.e., where $\mathbb{E}(f(X_t)|X_0 = x) = P_t f(x)$. It is well-known that

$$\lim_{t\to 0} t^{-1} \log \rho_{\max}(X_0;X_t) = -1/C_\infty(P_t). \quad (20)$$

Similarly, it is easy to show that

$$\lim_{t\to 0} t^{-1} \log \rho^{(M,\infty)}(X_0;X_t) = -1/C_M(P_t). \quad (21)$$

Let $\{-\lambda_k\}_{k\ge 0}$ be the eigenvalues of the generator $\mathcal{L}$, where $\lambda_1 = 0$ and $\lambda_k \le \lambda_{k+1}$. Due to reversibility, we can expand $\mathcal{L} = U^*\mathrm{diag}(\{-\lambda_k\})U$, where U is a unitary operator w.r.t. $\langle\cdot,\cdot\rangle_\mu$, i.e., $U^*U = \mathsf{Id}$, where U^* is the adjoint of U. Since $P_t = e^{t\mathcal{L}} = U^*\mathrm{diag}(\{e^{-\lambda_k t}\})U \approx U^*\mathrm{diag}(\{1-\lambda_k t\})U = \mathsf{Id} - U^*\mathrm{diag}(\{\lambda_k t\})U$ for small t, the eigenvalues of P_t are roughly $1 - \lambda_k t$. Therefore, $-\frac{1}{t}\log\rho_{\max}(X_0;X_t) \approx -\frac{1}{t}\log(1-t\lambda_1) \approx \lambda_1$, and we conclude that $C_\infty = \lambda_1^{-1}$

as is well known. For the finite-cardinality case, we can use our new anti-concentration result to obtain, for small t,

$$
\begin{aligned}
&-(1/t)\cdot\log\rho^{(M,\infty)}(X_0;X_t)\\
&\gtrapprox -(1/t)\cdot\log\left(1-t\lambda_1-\frac{\theta_M^{(1)}}{1+\theta_M^{(1)}}\cdot t(\min(2\lambda_1,\lambda_2)-\lambda_1)\right)\\
&\approx \lambda_1+\frac{\theta_M^{(1)}}{1+\theta_M^{(1)}}\cdot(\min(2\lambda_1,\lambda_2)-\lambda_1)
\end{aligned}
$$

This immediately implies the following.

Theorem 2 (bounded-cardinality Poincaré inequalities).

$$C_M(P_t)\le\left(\lambda_1+\frac{\theta_M^{(1)}}{1+\theta_M^{(1)}}\cdot(\min(2\lambda_1,\lambda_2)-\lambda_1)\right)^{-1},$$

$$C_M(P_t^{\otimes n})\le\left(\lambda_1+\frac{\eta_M}{1+\eta_M}\cdot(\min(2\lambda_1,\lambda_2)-\lambda_1)\right)^{-1}$$

where $\theta_M^{(1)}$ and η_M are the one-dimensional (12) *and dimension-independent* (17) *anti-concentration coefficients of the r.v. induced by μ over the second eigenfunction of $\mathcal{L}$.*

Note that C_M is usually strictly smaller than C_∞, and that the cardinality-M Poincaré constant does not tensorize, i.e., $C_M(P_t^{\otimes n})>C_M(P_t)$ is possible.

C. A new isoperimetric inequality

Let W be a discrete-alphabet reversible Markov kernel over $[K]\times[K]$ with a unique invariant distribution μ. Let $(X,Y)\sim\mu\times W$. The isopermietric coefficient of W is

$$h(W)\triangleq\min_{f:[K]\to\{-1,1\}}\frac{\Pr(f(X)\neq f(Y))}{\min(\Pr(f(X)=-1),\Pr(f(X)=1))}.$$

The Cartesian product channel W^n over $[K]^n\times[K]^n$ picks a single coordinate uniformly at random, and passes it through W. Namely, for the input x^n, we have $Y_J\sim W(\cdot|x_J)$ for $J\sim$ Uniform($[K]$), and $Y_j=x_j$ for all $j\neq J$. We are interested in lower bounds on $h(W^n)$. It is known [20, 21] that

$$h(W^n)\ge n^{-1}\max\left(h(W)/2,\lambda_1\right), \tag{22}$$

where $\lambda_1=\lambda_1(L)$ is the first nonzero eigenvalue of the Laplacian operator $L=\mathsf{Id}-W$. We now show that this bound can sometimes be improved using spectral methods.

Theorem 3 (isoperimetric inequalities). *Let $\{\lambda_k\}_{k=0}^\infty$ be the eigenvalues of $L=\mathsf{Id}-W$, in increasing order. Then*

$$h(W)\ge\lambda_1+\frac{\theta_2^{(1)}}{1+\theta_2^{(1)}}\cdot(\min(2\lambda_1,\lambda_2)-\lambda_1), \tag{23}$$

$$h(W^n)\ge\frac{1}{n}\left[\lambda_1+\frac{\eta_2}{1+\eta_2}\cdot(\min(2\lambda_1,\lambda_2)-\lambda_1)\right] \tag{24}$$

where $\theta_M^{(1)}$ and η_M are the one-dimensional (12) *and dimension-independent* (17) *anti-concentration coefficients of the r.v. induced by μ over the second eigenfunction of L.*

Proof. Since the invariant distribution is not necessarily uniform, it is more convenient to work with linear operators rather than matrices, over the inner-product space induced by $\langle\cdot,\cdot\rangle_\mu$. Let P be the conditional expectation operator induced by the channel W, namely $Pf(x)=\mathbb{E}(f(Y)|X=x)$. Then W and P have the same eigenvalues due to operator similarity, and by reversibility are self-adjoint w.r.t. $\langle\cdot,\cdot\rangle_\mu$. Now, consider the discrete semigroup P_t generated by the operator $\mathcal{L}=P-\mathsf{Id}$, which (since $\mathcal{L}^T\mu=0$) has μ as the invariant distribution. Note that $-\lambda_k(\mathcal{L})=\lambda_k(L)\triangleq\lambda_k$.

Since $\mathcal{L}$ is self-adjoint, we can write $\mathcal{L}=U^*\mathrm{diag}(\{-\lambda_k\})U$ where the operator U is unitary w.r.t. $\langle\cdot,\cdot\rangle_\mu$, i.e., $U^*U=\mathsf{Id}$. The semigroup P_t generated by $\mathcal{L}$ is then given by the operator exponent $P_t=e^{\mathcal{L}t}=U^*\mathrm{diag}(\{e^{-\lambda_k t}\})U$. Note that for small t, $P_t\approx U^*\mathrm{diag}(\{1-\lambda_k t\})U=\mathsf{Id}+t\mathcal{L}=(1-t)\mathsf{Id}+tP$, and hence the induced channel is $(1-t)\mathsf{Id}+tW$.

The Dirichlet form associated with P_t is given by

$$
\begin{aligned}
\mathcal{E}(f)&=-\langle f,\mathcal{L}f\rangle_\mu &(25)\\
&=\langle f,(\mathsf{Id}-P)f\rangle_\mu &(26)\\
&=\mathbb{E}f^2(X)-\mathbb{E}(f(X)\mathbb{E}(f(Y)|X)) &(27)\\
&=\mathbb{E}f^2(X)-\mathbb{E}(f(X)f(Y)). &(28)
\end{aligned}
$$

For a Boolean $f:[K]\to\{-1,1\}$, this yields

$$\mathcal{E}(f)=1-(1-2\Pr(f(X)\neq f(Y)))=2\Pr(f(X)\neq f(Y)).$$

Also, we have that

$$
\begin{aligned}
\mathsf{Var}_\mu(f)&=4\Pr(f(X)=1)\cdot\Pr(f(X)=-1) &(29)\\
&\ge2\min(\Pr(f(X)=1),\Pr(f(X)=-1)). &(30)
\end{aligned}
$$

Hence, (23) now follows using Theorem 2 and

$$h(W)\ge\min_{f:[K]\to\{-1,1\}}\frac{\mathcal{E}(f)}{\mathsf{Var}_\mu(f)}=\frac{1}{C_2(P_t)}. \tag{31}$$

Now let $P_t^{\otimes n}$ be n independent copies of P_t, and let $\mathcal{E}_k$ be the Dirichlet form operating on the kth copy only, with all else kept fixed. It is known that $\mathcal{E}(f)=\mathbb{E}\left(\sum_{k=1}^n\mathcal{E}_k(f)\right)$, and for a Boolean $f:[K]^n\to\{-1,1\}$ this yields

$$
\begin{aligned}
\mathcal{E}(f)&=2n\mathbb{E}\left(\Pr(f(X^n)\neq f(Y^n)|J,X^{\sim J})\right) &(32)\\
&=2n\Pr(f(X^n)\neq f(Y^n)), &(33)
\end{aligned}
$$

where $X^{\sim k}$ is the vector X^n without its kth coordinate, and where Y^n is obtained from X^n via the Cartesian product channel W^n defined previously. Hence, as in (31), (24) now follows from $h(W^n)\ge\frac{1}{n\cdot C_2(P_t^{\otimes n})}$ and Theorem 2. ☐

Example 2. *Consider the kernel $W(y|x)=((x+y \bmod 3)+1)/6$ over $\{0,1,2\}$, which is reversible with a uniform invariant distribution. It is easy to check that $1=h(W)\ge\lambda_1\approx0.7113$. Theorem 3 gives the improvements $h(W)\ge\lambda_1+0.637\cdot10^{-2}$, and $h(W^n)\ge(\lambda_1+0.334\cdot10^{-4})/n$.*

V. PROOFS

A. Proof of Theorem 1

In product space, the singular values in descending order are $1,\sigma_2,\ldots,\sigma_2,\max\{\sigma_3,\sigma_2^2\},\ldots,\max\{\sigma_3,\sigma_2^2\},\ldots$, where the n occurrences of σ_2 correspond to the n singular functions $f_{u^{(i)}}(X^n)=f_2(X_i)$ and $g_{u^{(i)}}(Y^n)=g_2(Y_i)$, for $1\le i\le n$, where $u^{(i)}$ has its i-th coordinate equal to 2 and all other entries equal to 1. Therefore, any function $f\in L^2(P_X^{\otimes n})$ can be written as

$$f(x^n)=\phi(x^n)+\phi_\perp(x^n) \tag{34}$$

where $\phi(x^n)\triangleq\sum_{i=1}^n a_if_2(x_i)$ is the projection of f onto the n-dimensional subspace pertaining to σ_2, i.e $\mathbb{E}\phi(X^n)\phi_\perp(X^n)=0$. Similarly, any function $g\in L^2(P_Y^{\otimes n})$

can be written as $g(y^n) = \nu(y^n) + \nu_\perp(y^n)$, where $\nu(y^n) \triangleq \sum_{i=1}^n b_i g_2(y_i)$ and $\mathbb{E}\nu(Y^n)\nu_\perp(Y^n) = 0$.

From this point onward, we assume that $f(X^n)$ and $g(Y^n)$ are unbiased, unit variance functions. By (5) and the Cauchy–Schwarz inequality, it is clear that

$$\mathbb{E}\phi(X^n)\nu(Y^n) \le \sigma_2\sqrt{\mathbb{E}\phi^2(X^n)\mathbb{E}\nu^2(Y^n)}. \quad (35)$$

Moreover, since $\phi_\perp(X^n)$ and $\nu_\perp(Y^n)$ are orthogonal to the subspace corresponding to σ_1 and σ_2, we have

$$\begin{aligned}\mathbb{E}\phi_\perp(X^n)\nu_\perp(Y^n) &\le \sigma_*\sqrt{\mathbb{E}\phi_\perp^2(X^n)\mathbb{E}\nu_\perp^2(Y^n)}\\ &= \sigma_*\sqrt{(1-\mathbb{E}\phi^2(X^n))(1-\mathbb{E}\nu^2(Y^n))},\end{aligned}$$

$\sigma_* \triangleq \max\{\sigma_3, \sigma_2^2\}$. Finally, since $\mathbb{E}\phi(X^n)\nu_\perp(Y^n) = \mathbb{E}\phi_\perp(X^n)\nu(Y^n) = 0$, we conclude that

$$\begin{aligned}\mathbb{E}f(X^n)g(Y^n) &\le \sigma_2\sqrt{\mathbb{E}\phi^2(X^n)\mathbb{E}\nu^2(Y^n)}\\ &+ \sigma_*\sqrt{(1-\mathbb{E}\phi^2(X^n))(1-\mathbb{E}\nu^2(Y^n))}. \quad (36)\end{aligned}$$

If $f(X^n)$ and $g(Y^n)$ are such that $\mathbb{E}\phi^2(X^n) \le h_x^2$ and $\mathbb{E}\nu^2(Y^n) \le h_y^2$, inequality (36) is further upper bounded by

$$\mathbb{E}f(X^n)g(Y^n) \le d(h_x, h_y), \quad (37)$$

where $d(a,b)$ is defined in (13).

In general, (37) always holds with $h_x = h_y = 1$, in which case it collapses to $\mathbb{E}f(X^n)g(Y^n) \le \rho_m(X;Y)$. However, the next Lemma shows that for M-valued functions, $\mathbb{E}\phi^2(X^n)$ and $\mathbb{E}\nu^2(Y^n)$ have non trivial upper bounds.

Lemma 2. *Let $f : \mathcal{X}^n \to \{m_1, \dots, m_M\}$ be with $\mathbb{E}f(X^n) = 0$ and $\mathbb{E}f^2(X^n) = 1$. Write*

$$f(X^n) = \phi(X^n) + \phi_\perp(X^n),$$

where $\phi(X^n)$ and $\phi_\perp(X^n)$ are as previously defined. Then

$$\mathbb{E}\phi^2(X^n) \le \min_t \frac{1}{1 + \frac{t^2}{4}\left|1 - M\cdot A_n(f_2(X), t)\right|_+}. \quad (38)$$

Proof. Denote $\mathbb{E}\phi^2(X^n) = ||\phi||^2$. We have

$$\begin{aligned}1 - ||\phi||^2 &= \mathbb{E}\phi_\perp^2(X^n) = \mathbb{E}(\phi(X^n) - f(X^n))^2\\ &\ge \Pr\left(|\phi(X^n) - f(X^n)| > t/2\right)\cdot(t/2)^2\\ &= \left(1 - \Pr\left(|\phi(X^n) - f(X^n)| \le t/2\right)\right)\cdot t^2/4\\ &\ge \left(1 - \Pr\left(\phi(X^n) \in \bigcup_{i=1}^M (m_i - t/2, m_i + t/2)\right)\right) t^2/4\\ &\ge \left(1 - \sum_{i=1}^M \Pr\left(\phi(X^n) \in (m_i - t/2, m_i + t/2)\right)\right)\cdot t^2/4\\ &\ge \left(1 - M\max_\alpha \Pr\left(\phi(X^n) \in (\alpha, \alpha + t)\right)\right)\cdot t^2/4\\ &= (1 - MQ(\phi(X^n), t))\cdot t^2/4. \quad (39)\end{aligned}$$

Recall $Q(\alpha X, t) = Q(X, t/\alpha)$ and write $t = ||\phi||\tau$. Then

$$\begin{aligned}1 - ||\phi||^2 &\ge (1 - MQ(\phi(X^n), ||\phi||\tau))\,||\phi||^2\tau^2/4\\ &= \left(1 - MQ\left(\frac{1}{||\phi||}\phi(X^n), \tau\right)\right)||\phi||^2\tau^2/4\\ &\ge (1 - M\cdot A_n(f_2(X), \tau))\,||\phi||^2\tau^2/4.\end{aligned}$$

Rearranging, we have

$$||\phi||^2 \le \left[1 + (1 - M\cdot A_n(f_2(X), \tau))\,\tau^2/4\right]^{-1},$$

and Lemma 2 follows since above holds for any τ. □

Substituting the bound from Lemma 2 in (37), we see that $\rho_m^{(M,N)}(X^n;Y^n) \le d(h_M, h_N)$, where h_M^2 is upper-bounded by the r.h.s. of (38), and h_N^2 is similarly upper bounded using the corresponding N and $g_2(Y)$. Similarly, $\rho_m^{(M,\infty)}(X^n;Y^n) \le d(h_M, 1)$, where h_M^2 is upper bounded by the r.h.s. of (38). A simple computation yields $d(a,1) = \sqrt{\sigma_2^2 - (\sigma_2^2 - \sigma_*^2)(1-a^2)}$, concluding the proof of Theorem 1.

B. Proof of Lemma 1

We need the following anti-concentration inequality.

Theorem 4 ([17]). *Let $S = X_1 + \ldots + X_n$, for i.i.d. $X_1, \dots, X_n$. Then for any $0 < t_1, \dots, t_n < t$, it holds that*

$$Q(S,t) \le \frac{Ct}{\sqrt{\sum_{i=1}^n t_i^2 D(X_i^*; t_i)}},$$

For some positive constant C no larger than $(96/95)^2\sqrt{48\pi/11} \approx 3.7809$, where X^* is the symmetrized r.v. corresponding to X, and $D(X,t)$ is defined in (10).

Now, let $\mathbf{a} \in \mathbb{R}^n$ with $\|\mathbf{a}\|_2 = 1$. By Theorem 4

$$\begin{aligned}Q\left(\sum a_i X_i, t\right) &\le \frac{Ct}{\sqrt{\sum_{i=1}^n t_i^2 D((a_i X_i)^*; t_i)}}\\ &= \frac{Ct}{\sqrt{\sum_{i=1}^n t_i^2 D(|a_i|(X_i)^*; t_i)}} = \frac{Ct}{\sqrt{\sum_{i=1}^n t_i^2 D(X_i^*; \frac{t_i}{|a_i|})}},\end{aligned}$$

Without losing generality, assume that $|a_1| \ge |a_i|,\ 1 \le i \le n$. Pick $t_i = t|a_i|/c$ for some $c > |a_1|$, then

$$\begin{aligned}Q\left(\sum a_i X_i, t\right) &\le \frac{Ct}{\sqrt{\sum_{i=1}^n (t|a_i|/c)^2 D(X^*; \frac{t}{c})}}\\ &= C\cdot c/\sqrt{D(X^*; t/c)}. \quad (40)\end{aligned}$$

Next, let $t_\tau = c\cdot\tau$ for $c = C^{-1}Q(X,\tau)\sqrt{D(X^*;\tau)}$ and some $\tau > 0$. Note that this choice does not necessarily satisfy $c > |a_1|$, which is a necessary condition in Theorem 4. But if $c > |a_1|$ nevertheless, then Theorem 4 and (40) give

$$Q\left(\sum a_i X_i, t_\tau\right) \le Q(X,\tau). \quad (41)$$

If however $c \le |a_1|$, a different argument applies: Consider the concentration of the random variable $|a_1|X_1$. A consequence of inequality (6) is that $Q\left(\sum a_i X_i, t\right) \le Q(|a_1|X, t)$. Furthermore, by our assumption that $c \le |a_1|$ we have

$$Q\left(\sum a_i X_i, t\right) \le Q(|a_1|X, t) = Q\left(X, \frac{t}{|a_1|}\right) \le Q\left(X, \frac{t}{c}\right),$$

and thus, under the same choice of $t_\tau = c\cdot\tau$, we get (41) again. We therefore conclude that (41) holds whether or not $c \ge |a_1|$, where $t_\tau = Q(X,\tau)\sqrt{D(X^*;\tau)}C^{-1}\tau$. This holds for any $\mathbf{a} \in \mathbb{R}^n$ with $\|\mathbf{a}\|_2 = 1$, hence $A_n(X, t_\tau) \le Q(X,\tau)$.

VI. Acknowledgements

This work was supported by the ISF under Grants 1791/17 and 1495/18. The work of DD was further supported by the Yitzhak and Chaya Weinstein Research Institute for Signal Processing. We sincerely thank an anonymous referee for bringing reference [18] to our attention.

REFERENCES

[1] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, 1975.

[2] C. Borell, "Geometric bounds on the Ornstein-Uhlenbeck velocity process," *Probability Theory and Related Fields*, vol. 70, no. 1, pp. 1–13, 1985.

[3] N. Alon and V. D. Milman, "λ1, isoperimetric inequalities for graphs, and superconcentrators," *Journal of Combinatorial Theory, Series B*, vol. 38, no. 1, pp. 73–88, 1985.

[4] R. O'Donnell, *Analysis of Boolean functions*. Cambridge University Press, 2014.

[5] J. Cheeger, "A lower bound for the smallest eigenvalue of the laplacian," in *Proceedings of the Princeton conference in honor of Professor S. Bochner*, 1969, pp. 195–199.

[6] D. A. Levin and Y. Peres, *Markov chains and mixing times*. American Mathematical Soc., 2017, vol. 107.

[7] G. Kumar, "Binary renyi correlation a simpler proof of Witsenhausen's result and a tighter upper bound," 2010, preprint.

[8] R. C. Bradley, "A "cousin" of a theorem of Cs'aki and Fischer," *arXiv preprint arXiv:1603.08964*, 2016.

[9] H. Lancaster, "The structure of bivariate distributions," *The Annals of Mathematical Statistics*, vol. 29, no. 3, pp. 719–736, 1958.

[10] A. Makur, "Information contraction and decomposition," Ph.D. dissertation, Massachusetts Institute of Technology, 2019.

[11] V. V. Petrov, *Sums of independent random variables*. Springer Science & Business Media, 2012, vol. 82.

[12] J. Littlewood and A. Offord, "On the number of real roots of a random algebraic equation, iii, rec," *Math.[Mat. Sbornik] NS*, vol. 12, no. 54, pp. 277–286, 1943.

[13] P. Erdös, "On a lemma of Littlewood and Offord," *Bulletin of the American Mathematical Society*, vol. 51, no. 12, pp. 898–902, 1945.

[14] T. Tao and V. H. Vu, *Additive combinatorics*. Cambridge University Press, 2006, vol. 105.

[15] A. N. Kolmogorov, "Two uniform limit theorems for sums of independent random variables," *Theory of Probability & Its Applications*, vol. 1, no. 4, pp. 384–394, 1956.

[16] B. A. Rogozin, "On the increase of dispersion of sums of independent random variables," *Theory of Probability & Its Applications*, vol. 6, no. 1, pp. 97–99, 1961.

[17] C. Esseen, "On the concentration function of a sum of independent random variables," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 9, no. 4, pp. 290–308, 1968.

[18] M. Rudelson and R. Vershynin, "Small ball probabilities for linear images of high-dimensional distributions," *International Mathematics Research Notices*, vol. 2015, no. 19, pp. 9594–9617, 2015.

[19] R. van Handel, "Probability in high dimension, apc 550 lecture notes," *Princeton University*, 2016.

[20] C. Houdré and P. Tetali, "Isoperimetric invariants for product markov chains and graph products," *Combinatorica*, vol. 24, no. 3, pp. 359–388, 2004.

[21] F. R. Chung and P. Tetali, "Isoperimetric inequalities for cartesian products of graphs," *Combinatorics Probability and Computing*, vol. 7, no. 2, pp. 141–148, 1998.