Electronic Colloquium on Computational Complexity, Report No. 84 (2015)

Dictatorship is the Most Informative Balanced Function at the Extremes

Or Ordentlich, Ofer Shayevitz and Omri Weinstein *

Abstract

Suppose X is a uniformly distributed n-dimensional binary vector and Y is obtained by passing X through a binary symmetric channel with crossover probability α. A recent conjecture by Courtade and Kumar postulates that $I(f(X);Y) \leq 1 - h(\alpha)$ for any Boolean function f. In this paper, we prove the conjecture for all $\alpha \in [0, \underline{\alpha}_n]$, and under the restriction to balanced functions, also for all $\alpha \in [\frac{1}{2} - \overline{\alpha}_n, \frac{1}{2}]$, where $\underline{\alpha}_n, \overline{\alpha}_n \to 0$ as $n \to \infty$. In addition, we derive an upper bound on $I(f(X);Y)$ which holds for all balanced functions, and improves upon the best known bound for all $\frac{1}{3} < \alpha < \frac{1}{2}$.

1 Introduction

Let X be an n-dimensional binary vector uniformly distributed over $\{0,1\}^n$, and Y be the output of passing each component of X through a binary symmetric channel with crossover probability $\alpha \leq 1/2$. The following was recently conjectured by Courtade and Kumar [1].

*Or Ordentlich and Ofer Shayevitz are with the Department of Electrical Engineering - Systems at the Tel Aviv University, {ordent,ofersha}@eng.tau.ac.il. Omri weinstein is with the Department of Computer Science, Princeton University, oweinste@cs.princeton.edu. The work of O. Ordentlich was supported by the Admas Fellowship Program of the Israel Academy of Science and Humanities. The work of O. Shayevitz was supported by an ERC grant no. 639573, aa CIG grant no. 631983, and an ISF grant no. 1367/14. The work of O. Weinstein was supported by a Simons Fellowship award in TCS and a Siebel scholarship.

ISSN 1433-8092

Conjecture 1 *For any Boolean function $f : \{0,1\}^n \mapsto \{0,1\}$ it holds that*

$$I\left(f(X);Y\right) \leq 1 - h(\alpha), \tag{1}$$

where $h(p) \triangleq -p\log p - (1-p)\log(1-p)$ is the binary entropy function.[1]

For a dictatorship function, $f(X) = X_i$ the conjectured upper bound (1) is attained with equality. Therefore, the conjecture can be interpreted as postulating that dictatorship is the most "informative" Boolean function, i.e., it achieves the maximal $I(f(X);Y)$.

So far, the best known bound that holds universally for all Boolean functions is

$$I\left(f(X);Y\right) \leq (1-2\alpha)^2. \tag{2}$$

This bound can be established through various techniques, including an application of Mrs. Gerber's Lemma [2–4], the strong data-processing inequality [5, 6] and standard Fourier analysis as described below. For very noisy channels, i.e., in the limit of $\alpha \to \frac{1}{2}$, this bound implies that

$$\lim_{\alpha\to\frac{1}{2}} \frac{I(f(X):Y)}{1-h(\alpha)} \leq \frac{2}{\log(e)} \approx 1.3863. \tag{3}$$

In the limit of $\alpha \to 0$, which corresponds to very clean channels, the ratio $(1-2\alpha)^2/(1-h(\alpha))$ approaches 1, but the derivative of $(1-2\alpha)^2$ at $\alpha = 0$ is finite, whereas the conjectured bound has an infinite (negative) slope at $\alpha = 0$.

In this paper, we prove Conjecture 1 for very clean channels, and under the restriction to balanced functions, also for very noisy channels. In addition, we derive an upper bound on $I(f(X);Y)$ which holds for all balanced functions, and improves upon (2) for all $\frac{1}{3} < \alpha < \frac{1}{2}$. Specifically, we obtain the following results, proved in sections 5, 3 and 4, respectively.

Theorem 1 (Dictatorship is optimal for very clean channels) *Let $f : \{0,1\}^n \mapsto \{0,1\}$ be any Boolean function. There exists $\underline{\alpha}_n > 0$ such that $I(f(X);Y) \leq 1 - h(\alpha)$ for all $\alpha \in [0, \underline{\alpha}_n]$. In particular, dictatorship is the most informative function in the noise interval $\alpha \in [0, \underline{\alpha}_n]$.*

[1] All logarithms are taken to base 2.

Theorem 2 *For any balanced Boolean function $f(x) : \{0,1\}^n \mapsto \{0,1\}$, and any $\frac{1}{2}\left(1-\frac{1}{\sqrt{3}}\right) \leq \alpha \leq \frac{1}{2}$, we have that*

$$I(f(X);Y) \leq \frac{\log(e)}{2}(1-2\alpha)^2 + 9\left(1-\frac{\log(e)}{2}\right)(1-2\alpha)^4.$$

Theorem 3 (Dictatorship is optimal for very noisy channels) *Let $f : \{0,1\}^n \mapsto \{0,1\}$ be any balanced Boolean function. There exists $\overline{\alpha}_n > 0$ such that $I(f(X);Y) \leq 1 - h(\alpha)$ for all $\alpha \in \left[\frac{1}{2} - \overline{\alpha}_n, \frac{1}{2}\right]$. In particular, dictatorship is the most informative balanced function in the noise interval $\alpha \in \left[\frac{1}{2} - \overline{\alpha}_n, \frac{1}{2}\right]$.*

The proof of Theorem 1 is based on the observation that, for α small enough, $\pi_y^f \triangleq \Pr\left(f(X) \neq f(Y)|Y=y\right)$ depends almost exclusively on the number of binary vectors $x \in \{0,1\}^n$ with Hamming distance 1 from y for which $f(x) \neq f(y)$ (i.e., the number of sensitive coordinates of the vector y). This in turn implies that, for very small α, the conditional entropy $H(f(X)|Y)$ depends on f mostly through the expectation of this quantity. This expectation is also known as the *Total Influence* of f, and is minimized by the dictatorship function.

For the proof of Theorem 2, we first lower bound the conditional entropy $H(f(X)|Y)$ in terms of the second and fourth moments of the random variable $(1-2P_Y^f)$, where $P_y^f \triangleq \Pr\left(f(X)=0|Y=y\right)$. Specifically, we show that

$$H(f(X)|Y) \geq 1 - \frac{\log(e)}{2}\mathbb{E}(1-2P_Y^f)^2 + \left(1-\frac{\log(e)}{2}\right)\mathbb{E}(1-2P_Y^f)^4. \quad (4)$$

To upper bound the second and fourth moments in (4), we use basic Fourier analysis of Boolean functions along with a simple application of the Hypercontractivity Theorem [7–9], in order to derive universal upper bounds on $\mathbb{E}(1-2P_Y^f)^{2k}$ that hold for all balanced Boolean functions.[2] In particular, these bounds show that $\mathbb{E}(1-2P_Y^f)^2 \leq (1-2\alpha)^2$ and $\mathbb{E}(1-2P_Y^f)^4 \leq 9(1-2\alpha)^2$. Plugging these bounds in (4) yields the Theorem.

[2]We remark that using similar techniques it is possible to obtain an upper bound on $I(f(X);Y)$ of a similar form that holds for any q-biased function (i.e., any function for which $\mathbb{E}(f(X)) = 1-2q$). However, the obtained bound is not maximized at $q=\frac{1}{2}$, and therefore cannot be used to establish the conjecture in the limit of $\alpha \to \frac{1}{2}$ for all Boolean functions.

A straightforward consequence of Theorem 2 is that for any *balanced* Boolean function f,

$$\lim_{\alpha \to \frac{1}{2}} \frac{I(f(X):Y)}{1-h(\alpha)} \leq 1, \tag{5}$$

which shows that the conjectured bound (1) becomes accurate for balanced functions, in the limit $\alpha \to \frac{1}{2}$.

Theorem 3 is essentially a corollary of Theorem 2. To prove the result, we rely on the fact that dictatorship is the only balanced Boolean function that attains the upper bound $\mathbb{E}(1-2P_Y^f)^2 \leq (1-2\alpha)^2$ with equality (i.e., dictatorships are the unique maximizers of the second moment). Furthermore, since the set of Boolean functions from $\{0,1\}^n \mapsto \{0,1\}$ is discrete, there exists a constant $c_n > 0$, independent of α, such that for any function g other than dictatorship, $\mathbb{E}(1-2P_Y^g)^2 < (1-c_n)(1-2\alpha)^2$. For α close enough to $\frac{1}{2}$ the second moment of $(1-2P_Y^f)^2$ dominates $H(f(X)|Y)$, and this shows that no function can obtain conditional entropy $H(f(X)|Y)$ smaller than dictatorship.

2 Preliminaries

To prove our results, it will be more convenient to map the additive group $\{0,1\}^n$ to the (isomorphic) multiplicative group $\{-1,1\}^n$. Specifically, let X be an n-dimensional binary vector uniformly distributed over $\{-1,1\}^n$ and Y be the output of passing each component of X through a binary symmetric channel with crossover probability $\alpha \leq 1/2$. Thus, for all $i \in \{1,\cdots,n\}$ we have $Y_i = X_i \cdot Z_i$, where $\{Z_i\}_{i=1}^n$ is an i.i.d. sequence of binary random variables statistically independent of $\{X_i\}_{i=1}^n$, with $\Pr(Z_i = -1) = \alpha$ and $\Pr(Z_i = 1) = 1-\alpha$. Note that Y is also uniformly distributed over $\{-1,1\}^n$.

Let $f : \{-1,1\}^n \mapsto \{-1,1\}$ be a *balanced* Boolean function, i.e.,

$$\Pr\left(f(X) = -1\right) = \frac{1}{2}.$$

Note that this condition is equivalent to $\mathbb{E}_X(f(X)) = 0$. For each $y \in \{-1,1\}^n$ define the posterior distribution of $f(X)$ given the observation y

$$P_y^f \triangleq \Pr(f(X) = -1 | Y = y),$$

and note that

$$\mathbb{E}(f(X)|Y=y) = 1 - 2P_y^f. \tag{6}$$

In what follows, we will extensively use Fourier analysis of real functions on the hypercube $\{-1,1\}^n$. Let X be a random vector uniformly distributed on $\{-1,1\}^n$ and define $[n] \triangleq \{1,2,\ldots,n\}$. The Fourier-Walsh transform of a function $f : \{-1,1\}^n \mapsto \mathbb{R}$ is given by

$$f(x) = \sum_{S\subseteq[n]} \hat{f}(S) \prod_{i\in S} x_i, \tag{7}$$

where

$$\hat{f}(S) = \mathbb{E}_X \left(f(X) \prod_{i\in S} X_i \right) \tag{8}$$

is the correlation of f with the "parity" function on the subset S. It is easy to verify that the basis $\left\{\varphi_S(x) = \prod_{i\in S} x_i\right\}_{S\subseteq[n]}$ is orthonormal with respect to the inner product $< f,g >= \mathbb{E}\left(f(X)g(X)\right)$, which implies that for any two functions $f,g : \{-1,1\}^n \mapsto \mathbb{R}$ it holds that

$$\mathbb{E}\left(f(X)g(X)\right) = \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S). \tag{9}$$

In particular, Parseval's identity gives $\sum_S \hat{f}^2(S) = \mathbb{E}\left(f^2(X)\right)$. Thus, if f is a Boolean function, i.e., $f : \{-1,1\}^n \mapsto \{-1,1\}$, we have $\sum_S \hat{f}^2(S) = 1$.

The following proposition gives four simple properties of the Fourier coefficients that will be useful in the sequel.

Proposition 1 (Basic properties of the Fourier transform) *Let $f : \{-1,1\}^n \mapsto \{-1,1\}$ be a Boolean function. We have that*

(i) f is balanced if and only if $\hat{f}(\emptyset) = 0$;

(ii) If $|\hat{f}(S)| > 0$ then $|\hat{f}(S)| \geq 2^{-n}$;

(iii) Any balanced function f which is not a dictatorship function must satisfy $\sum_{S:\,|S|\geq 2} \hat{f}^2(S) > 0$;

(iv) For $x, y \in \{-1,1\}^n$ write $x \sim y$ to denote that x and y differ in exactly one coordinate (i.e., edges of the hypercube), and define

$$K_y^f \triangleq \left|\{x \in \{-1,1\}^n \ : \ x \sim y, \ f(x) \neq f(y)\}\right|. \tag{10}$$

It holds that

$$2^{-n} \sum_{y\{-1,1\}^n} K_y^f = \sum_{S \subseteq [n]} |S| \hat{f}^2(S).$$

We remark that the quantity $2^{-n}\sum_y K_y^f$ is sometimes called the total influence *of f [7].*

Proof.

(i) By definition $\hat{f}(\emptyset) = \mathbb{E}f(X) = 0$ for f balanced.

(ii) We note that the sum $\sum_{x\in\{-1,1\}^n} f(x) \prod_{i\in S} x_i$ is an integer, and the claim follows immediately.

(iii) Let f be a balanced Boolean function ($\hat{f}(\emptyset) = 0$) with $\sum_{S:\,|S|\geq 2} \hat{f}^2(S) = 0$. Then $f(X) = \sum_{i\in[n]} \hat{f}_{\{i\}} X_i$. Since there always exists some $x \in \{-1,1\}^n$ for which $f(x) = \sum_i |\hat{f}(i)|$, we must have that $\sum_i |\hat{f}(i)| = 1$. On the other hand, by Parseval's identity we have $\sum_{i=1} \hat{f}^2(i) = 1$. Clearly, the last two equations can simultaneously hold if and only if there is a unique i for which $|\hat{f}_i| > 0$. Hence, f must be a dictatorship.

(iv) This statement is well-known, see, e.g., [7]. We give the proof for completeness. For every coordinate $i \in [n]$, let $\mathbf{1}_{f(y)\neq f(y^{\oplus i})}$ denote the indicator of the event that f takes different values on y and on y with its i'th bit flipped ($y^{\oplus i}$). Let us define the (real-valued) function

$$g^i(y) \triangleq \frac{f(y) - f(y^{\oplus i})}{2},$$

and observe that

- $g_i^2(y) = \mathbf{1}_{f(y)\neq f(y^{\oplus i})}$.

- The Fourier coefficients of g_i satisfy, for any $S \subseteq [n]$:

$$\hat{g}_i(S) = \begin{cases} \hat{f}(S) & \text{if } i \in S \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

By definition of K_y^f, we thus have

$$\begin{aligned} 2^{-n} \sum_{y\{-1,1\}^n} K_y^f &= 2^{-n} \sum_y \sum_{i\in[n]} \mathbf{1}_{f(y)\neq f(y^{\oplus i})} \\ &= 2^{-n} \sum_y \sum_{i\in[n]} g_i^2(y) \\ &= \sum_{i\in[n]} \mathbb{E} g_i^2(Y) \\ &= \sum_{i\in[n]} \sum_{S\subseteq[n]} \hat{g}_i{}^2(S) \qquad (12) \\ &= \sum_{i\in[n]} \sum_{S\subseteq[n]:i\in S} \hat{f}^2(S) \qquad (13) \\ &= \sum_{S\subseteq[n]} \sum_{i\in S} \hat{f}^2(S) \\ &= \sum_{S\subseteq[n]} |S| \hat{f}^2(S). \qquad (14) \end{aligned}$$

where (12) follows by Parseval's inequality, and (13) by (11).

■

For any function $f : \{-1,1\}^n \mapsto \mathbb{R}$ with Fourier coefficients $\{\hat{f}(S)\}$ and $\rho \in \mathbb{R}^+$, the *noise operator* $T_\rho f : \{-1,1\}^n \mapsto \mathbb{R}$ is defined as [7]

$$(T_\rho f)(x) \triangleq \sum_{S\subseteq[n]} \hat{f}(S) \rho^{|S|} \prod_{i\in S} x_i. \tag{15}$$

Recall that in our setting X and Y are the input and output of a binary symmetric channel with crossover probability α. Thus we can use the Fourier representation of f to write $\mathbb{E}\left(f(X)|Y=y\right)$ as follows [7]

$$\mathbb{E}\left(f(X)|Y=y\right) = \mathbb{E}\left(\sum_{S\subseteq[n]} \hat{f}(S) \prod_{i\in S} X_i \;\middle|\; Y=y\right)$$

$$
\begin{aligned}
&= \sum_{S\subseteq[n]} \hat{f}(S)\mathbb{E}\left(\prod_{i\in S} y_i Z_i\right) \\
&= \sum_{S\subseteq[n]} \hat{f}(S)\mathbb{E}\left(\prod_{i\in S} Z_i\right)\prod_{i\in S} y_i \\
&= \sum_{S\subseteq[n]} \hat{f}(S)\prod_{i\in S}\mathbb{E}\left(Z_i\right)\prod_{i\in S} y_i \\
&= \sum_{S\subseteq[n]} \hat{f}(S)(1-2\alpha)^{|S|}\prod_{i\in S} y_i, \qquad (16) \\
&= (T_{1-2\alpha}f)(y), \qquad (17)
\end{aligned}
$$

where we have used the fact that $\{Z_i\}$ is an i.i.d. sequence. Recalling equation (6), this also yields

$$
1-2P_y^f = (T_{1-2\alpha}f)(y) = \sum_{S\subseteq[n]} \hat{f}(S)(1-2\alpha)^{|S|}\prod_{i\in S} y_i. \qquad (18)
$$

The following well-known theorem will play an important role in the derivation of Theorem 2.

Theorem 4 (Hypercontractivity Theorem, [7–9]) *Let $1 \le p < q < \infty$. Then for all $\rho \le \sqrt{\frac{p-1}{q-1}}$, and all $g : \{-1,1\}^n \mapsto \mathbb{R}$, it holds that*

$$
\left[\mathbb{E}((T_\rho g)^q(X))\right]^{\frac{1}{q}} \le \left[\mathbb{E}(g^p(X))\right]^{\frac{1}{p}}. \qquad (19)
$$

3 The Intermediate Noise Regime

In this section we prove Theorem 2, which gives a universal upper bound on the mutual information $I(f(X);Y)$, that holds for any *balanced* Boolean function. The mutual information can be expressed as

$$
I(f(X);Y) = H(f(X)) - H(f(X)|Y) = 1 - \mathbb{E}_Y h(P_Y^f). \qquad (20)
$$

We note that $h(\cdot)$ admits the following Taylor series

$$h\left(\frac{1-p}{2}\right) = 1 - \sum_{k=1}^{\infty} \frac{\log(e)}{2k(2k-1)} p^{2k}, \tag{21}$$

and can be lower bounded by replacing p^{2k} with p^{2t} for all $k > t$, i.e.,

$$\begin{aligned} h\left(\frac{1-p}{2}\right) &\geq 1 - \sum_{k=1}^{t-1} \frac{\log(e)}{2k(2k-1)} p^{2k} - p^{2t} \sum_{k=t}^{\infty} \frac{\log(e)}{2k(2k-1)} \\ &= 1 - \sum_{k=1}^{t-1} \frac{\log(e)}{2k(2k-1)} p^{2k} - \left(1 - \sum_{k=1}^{t-1} \frac{\log(e)}{2k(2k-1)}\right) p^{2t}, \end{aligned} \tag{22}$$

where we have used the fact that $h(0) = 0$. Using (22), for any $t \in \mathbb{N}$, we can further lower bound $\mathbb{E}_Y h(P_Y^f)$ as

$$\begin{aligned} \mathbb{E}_Y h(P_Y^f) &= \mathbb{E}_Y h\left(\frac{1-(1-2P_Y^f)}{2}\right) \\ &\geq 1 - \sum_{k=1}^{t-1} \frac{\log(e)}{2k(2k-1)} \mathbb{E}_Y\left((1-2P_Y^f)^{2k}\right) \\ &\quad - \left(1 - \sum_{k=1}^{t-1} \frac{\log(e)}{2k(2k-1)}\right) \mathbb{E}_Y\left((1-2P_Y^f)^{2t}\right). \end{aligned} \tag{23}$$

Thus, any upper bound on the first t even moments $\mathbb{E}[(1-2P_Y^f)^{2k}]$, $k = 1, \ldots, t$, would directly translate to an upper bound on $I(f(X);Y)$.

Our simple argument yields an upper bound on all even moments of the random variable $1-2P_Y^f$, using a simple trick combined with the Hypercontractivity Theorem. Formally, we prove the following lemma.

Lemma 1 *Let $k \geq 1$ be an integer satisfying $(1-2\alpha)\sqrt{2k-1} \leq 1$. For any balanced Boolean function $f : \{-1,1\}^n \mapsto \{-1,1\}$ we have that*

$$\mathbb{E}_Y\left((1-2P_Y^f)^{2k}\right) \leq (2k-1)^k (1-2\alpha)^{2k}.$$

For the proof, we will need the following proposition.

Proposition 2 *For any balanced Boolean function $f : \{-1,1\}^n \mapsto \{-1,1\}$ and any $0 \leq \rho \leq 1$ we have that*

$$\mathbb{E}_Y\left((T_\rho f)^2(Y)\right) \leq \rho^2.$$

Proof. By the definition of the operator T_ρ:

$$\begin{aligned}
\mathbb{E}_Y\left((T_\rho f)^2(Y)\right) &= \mathbb{E}_Y\left(\left(\sum_{S\subseteq[n]} \hat{f}(S)\rho^{|S|}\prod_{i\in S} Y_i\right)^2\right) \\
&= \mathbb{E}_Y\Bigg(\sum_{S_1\in[n]} \hat{f}(S_1)\rho^{|S_1|}\prod_{i\in S_1} Y_i \sum_{S_2\in[n]} \hat{f}(S_2)\rho^{|S_2|}\prod_{j\in S_2} Y_j\Bigg) \\
&= \sum_{S_1\in[n]} \hat{f}(S_1)\rho^{|S_1|}\sum_{S_2\in[n]} \hat{f}(S_2)\rho^{|S_2|}\mathbb{E}_Y\left(\prod_{i\in S_1} Y_i\prod_{j\in S_2} Y_j\right) \\
&= \sum_{S_1\in[n]} \hat{f}(S_1)\rho^{|S_1|}\sum_{S_2\in[n]} \hat{f}(S_2)\rho^{|S_2|}\mathbf{1}_{S_1=S_2} \\
&= \sum_{S\subseteq[n]} \hat{f}^2(S)\rho^{2|S|}, \qquad (24)
\end{aligned}$$

where $\mathbf{1}_{S_1=S_2}$ is the indicator function on the event $S_1 = S_2$. Recalling that $\sum_{S\subseteq[n]} \hat{f}^2(S) = 1$ and our assumptions that $\rho \leq 1$ and that $\hat{f}(\emptyset) = \mathbb{E}(f(X)) = 0$, the "weight" assignment $\hat{f}^2(S)$ that maximizes $\sum_{S\subseteq[n]} \hat{f}^2(S)\rho^{2|S|}$ puts all the weight on characters S whose cardinality is $|S| = 1$. Hence,

$$\sum_{S\subseteq[n]} \hat{f}^2(S)\rho^{2|S|} \leq \rho^2,$$

as desired. ■

Proof of Lemma 1. Let $f : \{-1,1\}^n \mapsto \{-1,1\}$ be a boolean function and let $g \triangleq T_{1-2\alpha}f$. Let $k \geq 1$ be an integer, and let $\rho = \sqrt{1/(2k-1)} = \sqrt{(2-1)/(2k-1)}$. By (18), we have

$$\begin{aligned}
\mathbb{E}_Y((1-2P_Y^f)^{2k}) &= \mathbb{E}_Y\left(g^{2k}(Y)\right) \\
&= \mathbb{E}_Y\left(\left(T_\rho(T_{1/\rho}g)\right)^{2k}(Y)\right) \\
&\leq \left[\mathbb{E}_Y\left((T_{1/\rho}g)^2(Y)\right)\right]^k \qquad (25)
\end{aligned}$$

$$= \left[\mathbb{E}_Y\left((T_{(1-2\alpha)\sqrt{2k-1}}f)^2(Y)\right)\right]^k, \tag{26}$$

where (25) follows from the Hypercontractivity Theorem taken with $p = 2, q = 2k$ (as ρ satisfies the premise of the theorem), and (26) from the definition of the function g. Invoking Proposition 2 with $\rho = (1-2\alpha)\sqrt{2k-1} \leq 1$, we obtain that for any balanced Boolean function and any $k \geq 1$

$$\mathbb{E}_Y\left((1-2P_Y^f)^{2k}\right) \leq (2k-1)^k(1-2\alpha)^{2k} \tag{27}$$

as desired. ■

The following is an immediate consequence of Lemma 1 and (23).

Proposition 3 *For any balanced Boolean function $f : \{-1,1\}^n \mapsto \{-1,1\}$, any integer $t \geq 1$ and any $\frac{1}{2}\left(1 - \frac{1}{\sqrt{2t-1}}\right) \leq \alpha \leq \frac{1}{2}$, we have that*

$$I(f(X);Y) \leq \sum_{k=1}^{t-1} \frac{\log(e)}{2k(2k-1)}(2k-1)^k(1-2\alpha)^{2k} + \left(1 - \sum_{k=1}^{t-1}\frac{\log(e)}{2k(2k-1)}\right)(2t-1)^t(1-2\alpha)^{2t}. \tag{28}$$

Theorem 2 now follows by evaluating (28) with $t = 2$. Note that for balanced functions the upper bound $(1-2\alpha)^2$, which was the best known bound hitherto, is obtained as a special case of Proposition 3 by setting $t = 1$. It is easy to verify that for $\frac{1}{3} < \alpha < \frac{1}{2}$ the upper bound in Theorem 2 is tighter. See Figure 1 for a comparison between the bounds.

4 The High Noise Regime

In this section we prove Theorem 3 and establish the optimality of the dictatorship function among all balanced functions in the very noisy regime. the proof is essentially a simple consequence of Theorem 2 and the discreteness of the space of Boolean functions from $\{-1,1\}^n \mapsto \{-1,1\}$.

By equation (23) applied with $t = 2$, we have that for any balanced Boolean function (and any α),

$$I\left(f(X);Y\right) \leq \frac{\log(e)}{2}\mathbb{E}_Y\left((1-2P_Y^f)^2\right) + \left(1 - \frac{\log(e)}{2}\right)\mathbb{E}_Y\left((1-2P_Y^f)^4\right).$$

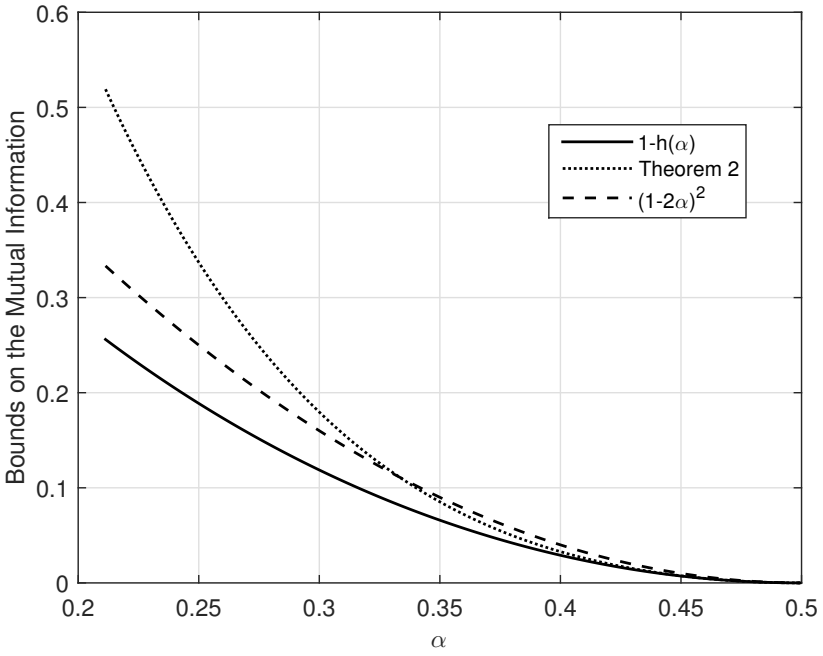

Figure 1: A comparison between the bound from Theorem 2, the conjectured bound and best known previous upper bound on $I(f(X);Y)$. Note that the bound from Theorem 2 is only valid for balanced functions.

Lemma 1 (applied with $k = 2$) implies that for $\frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) \leq \alpha \leq \frac{1}{2}$

$$\mathbb{E}_Y\left((1-2P_Y^f)^4\right) \leq (2\cdot 2-1)^2(1-2\alpha)^4 = 9(1-2\alpha)^4. \tag{29}$$

Furthermore, by equations (18) and (24), we have that for α in this range

$$\begin{aligned}\mathbb{E}_Y\left((1-2P_Y^f)^2\right) &= \sum_{S\subseteq[n]} \hat{f}^2(S)(1-2\alpha)^{2|S|} \\ &\leq \left(\sum_{|S|=1} \hat{f}^2(S)\right)(1-2\alpha)^2 + \left(\sum_{|S|\geq 2} \hat{f}^2(S)\right)(1-2\alpha)^4, \end{aligned} \tag{30}$$

where we have used the fact that $\hat{f}(\emptyset) = 0$ for balanced functions (Proposition 1). Combining the above three inequalities and using the fact that

$\sum_{S\subseteq[n]} \hat{f}^2(S) = 1$, yields

$$I\left(f(X);Y\right) \leq \frac{\log(e)}{2}\left(1-\sum_{|S|\geq 2}\hat{f}^2(S)\right)(1-2\alpha)^2+$$
$$+\left(9\left(1-\frac{\log(e)}{2}\right)+\frac{\log(e)}{2}\sum_{|S|\geq 2}\hat{f}^2(S)\right)(1-2\alpha)^4. \tag{31}$$

Now, suppose that f is a balanced Boolean function which is *not* a dictatorship function. This implies (in fact, equivalent to) $\sum_{|S|\geq 2}\hat{f}^2(S) > 0$. By Proposition 1, it therefore must be the case that

$$\sum_{|S|\geq 2}\hat{f}^2(S) > 2^{-2n} = 4^{-n}.$$

Therefore, for such f, the RHS of (31) can be upper bounded by

$$\frac{\log(e)}{2}(1-4^{-n})(1-2\alpha)^2+\left(9\left(1-\frac{\log(e)}{2}\right)+\frac{\log(e)}{2}4^{-n}\right)(1-2\alpha)^4. \tag{32}$$

It can be directly verified that for any $\alpha \in [0.5-\overline{\alpha}_n, 0.5]$, where $\overline{\alpha}_n \triangleq \frac{1}{4}\cdot 2^{-n}$, the expression in (32) is smaller than $\frac{\log(e)}{2}(1-2\alpha)^2 < 1-h(\alpha)$, and thus by (31), for such α we have

$$I\left(f(X);Y\right) \leq 1-h(\alpha), \tag{33}$$

which completes the proof.

5 The Low Noise Regime

In this section we prove Theorem 1, which shows that dictatorship is the most informative Boolean function in the very low noise regime.

Proof of Theorem 1. Let $\underline{\alpha}_n \triangleq \frac{2^{-2n}}{16n^2}$ and assume that $\alpha < \underline{\alpha}_n$. We begin by showing that for any α in this regime $I(g(X);Y) < 1-h(\alpha)$ for any biased function g. As $I(g(X);Y) \leq h\left(\Pr(g(X)=1)\right)$, it suffices to

show that $h(\alpha) \leq 1 - h\left(\Pr(g(X)=1)\right)$. Note that if g is biased, then $|\Pr(g(X)=1) - \frac{1}{2}| \geq 2^{-n}$, and therefore

$$\begin{aligned} 1 - h\left(\Pr(g(X)=1)\right) &\geq 1 - h\left(\frac{1}{2} - 2^{-n}\right) \\ &\geq 1 - \sqrt{4\left(\frac{1}{2} - 2^{-n}\right)\left(\frac{1}{2} + 2^{-n}\right)} \qquad (34) \\ &= 1 - \sqrt{1 - 4 \cdot 2^{-2n}}, \end{aligned}$$

where we have used the bound $h(p) \leq \sqrt{4p(1-p)}$ in (34). On the other hand, for α in this regime we have

$$\begin{aligned} h(\alpha) &\leq 2\alpha \log\left(\frac{1}{\alpha}\right) \\ &\leq 2\underline{\alpha}_n \log\left(\frac{1}{\underline{\alpha}_n}\right) \\ &= 2^{-2n}\left(\frac{1}{4n} + \frac{1}{2n^2} + \frac{\log(n)}{4n^2}\right) \\ &< 2^{-2n}. \end{aligned}$$

We therefore have that for any α in this regime and any biased function g

$$h(\alpha) < 2^{-2n} < 1 - \sqrt{1 - 4 \cdot 2^{-2n}} \leq 1 - h\left(\Pr(g(X)=1)\right),$$

as desired. Thus, it suffices to prove that $I(f(X);Y) \leq 1 - h(\alpha)$ for all balanced functions. In particular, we will show that $I(f(X);Y) < 1 - h(\alpha)$ for all functions f that are not a dictatorship.

Let $f : \{-1,1\}^n \mapsto \{-1,1\}$ be balanced, and assume it is not a dictatorship. Define

$$\pi_y^f \triangleq \Pr(f(X) \neq f(Y) | Y = y),$$

Now, note that $\pi_y^f \in \{P_y^f, 1 - P_y^f\}$, and hence $h(P_y^f) = h(\pi_y^f)$. Thus,

$$H(f(X)|Y) = \mathbb{E}h(\pi_Y^f), \qquad (35)$$

We first show that π_y^f is roughly equal to $K_y^f \alpha$ for α small enough.

Lemma 2 *For any $y \in \{-1,1\}^n$, it holds that*

$$K_y^f \alpha - n^2\alpha^2 \leq \pi_y^f \leq K_y^f \alpha + 2^n \alpha^2$$

Proof. Let $|x-y|$ denote the Hamming distance between x and y. Write

$$\begin{aligned}\pi_y^f &= \sum_{x:f(x)\neq f(y)} \alpha^{|x-y|}(1-\alpha)^{n-|x-y|} \\ &= K_y^f \cdot \alpha(1-\alpha)^{n-1} + \sum_{x:f(x)\neq f(y),|x-y|\geq 2} \alpha^{|x-y|}(1-\alpha)^{n-|x-y|}\end{aligned}$$

Therefore on the one hand we have

$$\pi_y^f - K_y^f\alpha \leq K_y^f\alpha((1-\alpha)^{n-1}-1) + 2^n\alpha^2(1-\alpha)^{n-2} \leq 2^n\alpha^2$$

and on the other hand, recalling also that $K_y^f \leq n$, we also have that

$$\pi_y^f - K_y^f\alpha \geq K_y^f\alpha((1-\alpha)^{n-1}-1) \geq -K_y^f\alpha^2(n-1) \geq -n^2\alpha^2$$

and the claim follows. ■

Our assumption that $\alpha \leq \underline{\alpha}_n$ combined with the upper bound in Lemma 2 (and recalling that $K_y^f \leq n$) guarantees that $\pi_y^f < \frac{1}{2}$. Thus, continuing with (35) and using the lower bound in Lemma 2, we can write

$$\begin{aligned}H(f(X)|Y) &\geq 2^{-n} \sum_{y:K_y^f\geq 1} h(K_y^f\alpha - n^2\alpha^2) \\ &= -2^{-n} \sum_{y:K_y^f\geq 1} (K_y^f\alpha - n^2\alpha^2)\log(K_y^f\alpha - n^2\alpha^2) \\ &\quad - 2^{-n} \sum_{y:K_y^f\geq 1} (1+n^2\alpha^2 - K_y^f\alpha)\log(1+n^2\alpha^2-K_y^f\alpha). \quad (36)\end{aligned}$$

The expectation $\mathbb{E}(K_Y^f)$ will play an important role in the remainder of the derivation. Note that by Proposition 1 we can write

$$\mathbb{E}K_Y^f = \sum_{s\subseteq[n]} |S|\hat{f}^2(S) = 1 + \delta_f$$

where $0 \leq \delta_f \leq n-1$. Moreover, by Proposition 1 and the assumption that f is not a dictatorship, $\delta_f \geq 2^{-2n}$.

We proceed by separately bounding the first and second summands in (36). For the first, we have

$$\begin{aligned}
&2^{-n} \sum_{y:K_y^f \geq 1} (K_y^f \alpha - n^2\alpha^2) \log(K_y^f \alpha - n^2\alpha^2) \\
&\quad \leq 2^{-n} \sum_{y:K_y^f \geq 1} (K_y^f \alpha - n^2\alpha^2) \log(K_y^f \alpha) \\
&\quad \leq 2^{-n} \sum_{y:K_y^f \geq 1} \left(K_y^f \alpha(\log\alpha + \log K_y^f) - n^2\alpha^2 \log\alpha\right) \\
&\quad \leq \alpha \log\alpha \mathbb{E}K_Y^f + \alpha \mathbb{E}(K_Y^f \log K_Y^f) - n^2\alpha^2 \log\alpha \\
&\quad \leq \alpha \log\alpha \mathbb{E}K_Y^f + \alpha \mathbb{E}K_Y^f \cdot \log \mathbb{E}K_Y^f - n^2\alpha^2 \log\alpha \qquad (37) \\
&\quad = \alpha(1+\delta_f)\log(\alpha(1+\delta_f)) - n^2\alpha^2\log\alpha \\
&\quad \leq \alpha(1+2^{-2n}) \cdot \log(\alpha(1+2^{-2n})) - n^2\alpha^2\log\alpha \qquad (38) \\
&\quad \leq \alpha\log\alpha + 2^{-2n}\alpha\log\alpha + \alpha(1+2^{-2n})\log(1+2^{-2n}) - n^2\alpha^2\log\alpha \\
&\quad \leq \alpha\log\alpha + (2^{-2n} - n^2\alpha)\alpha\log\alpha + \alpha(1+2^{-2n})2^{-2n}\log e \\
&\quad \leq \alpha\log\alpha + 2^{-2n}\alpha\left(\frac{\log\alpha}{2} + (1+2^{-2n})\log e\right) \qquad (39) \\
&\quad < \alpha\log\alpha \qquad (40)
\end{aligned}$$

where (37) is by virtue of Jensen's inequality, (38) follows since $x\log x$ is monotonically decreasing in the regime of interest, and (39) and (40) follows since $\alpha < \underline{\alpha}_n$.

We bound the second summand in (36) as follows

$$\begin{aligned}
&2^{-n} \sum_{y:K_y^f \geq 1} (1 + n^2\alpha^2 - K_y^f\alpha) \log(1 + n^2\alpha^2 - K_y^f\alpha) \\
&\quad \leq 2^{-n} \sum_{y:K_y^f \geq 1} (1+n^2\alpha^2)\log(1+n^2\alpha^2 - K_y^f\alpha) - K_y^f\alpha\log(1-K_y^f\alpha) \\
&\quad \leq -2^{-n} \sum_{y:K_y^f \geq 1} \log e \left((1+n^2\alpha^2)\frac{K_y^f\alpha - n^2\alpha^2}{1+n^2\alpha^2 - K_y^f\alpha} - (K_y^f\alpha)^2\right) \qquad (41)
\end{aligned}$$

$$\leq -2^{-n} \sum_{y:K_y^f \geq 1} \log e \left((1+n^2\alpha^2) \frac{K_y^f \alpha - n^2\alpha^2}{1+n^2\alpha^2} - (n\alpha)^2 \right) \tag{42}$$

$$= -2^{-n} \sum_{y:K_y^f \geq 1} \log e \left(K_y^f \alpha - 2n^2\alpha^2 \right)$$

$$\leq -\alpha \left(\mathbb{E} K_y^f - 2n^2\alpha \right) \log e$$

$$= -\alpha - \alpha(\delta_f - 2n^2\alpha) \log e$$

$$< -\alpha \log e$$

$$< (1-\alpha)\log(1-\alpha). \tag{43}$$

where we have used the fact that $-t \log e \leq \log(1-t) \leq -\frac{t}{1-t} \log e$ for $0 < t < 1$ in (41), the fact that $K_y^f \leq n$ in (42), and the fact that $(1-\alpha)\log(1-\alpha) \geq -\alpha \log e$ for any $\alpha \in (0,1)$.

From (36), (40) and (43), we conclude that for any $0 < \alpha < \underline{\alpha}_n$ it holds that $H(f(X)|Y) > h(\alpha)$, concluding the proof. ■

6 Discussion

In light of our results (Theorem 1 and Theorem 3), to prove Conjecture 1, it suffices to prove one the following weaker conjectures:

Conjecture 2 (Existence of a global optimal function) *There is a Boolean function that maximizes the mutual information, simultaneously for all noise parameters:*

$$\exists \ f \ \ \forall \, \alpha \in [0, 1/2] \quad \forall \ g: \quad I\left(g(X); Y\right) \leq I\left(f(X); Y\right).$$

Conjecture 3 (The information of dictatorship is unique) *For any value $\alpha \in (0, 1/2)$, the unique Boolean function f (up to permutation) for which*

$$I\left(f(X); Y\right) = 1 - h(\alpha),$$

is the dictatorship function $f(X) = X_1$.

Indeed, Conjecture 3 implies Conjecture 1 by a standard continuity argument (at $\underline{\alpha}_n$) combined with Theorem 1 (and for balanced functions, by a continuity argument (at $\overline{\alpha}_n$) and Theorem 3). Notice that this conjecture only requires ruling out strict equality, and is thus logically different than Conjecture 1.

limitations of our approach. A natural question is: What are the limits of the approach pursued in this paper? The result for the low-noise regime (Theorem 1) is clearly very specialized for tiny values of α, and so we focus the discussion below on the proof techniques of Theorems 2 and 3. To this end we note the following two limitations:

Firstly, the dictatorship function, which is conjectured to be optimal, satisfies $\mathbb{E}_Y\left((1-2P_Y^f)^{2k}\right) = (1-2\alpha)^{2k}$ for every $k \in \mathbb{N}$. The ratio between the bound in Lemma 1 on the k-th moment of any balanced Boolean function and $(1-2\alpha)^{2k}$ grows rapidly with k. For this reason, we only get mileage from applying the lemma with $k = 1, 2$ and not for higher moments.

The second limitation is that Lemma 1 upper bounds *each moment separately*, while we are seeking an upper bound on the entire distribution (weighted sum) of the moments: Quantifying the tradeoff between higher and lower moments seems to be one of the "brick walls" in proving the conjecture. For example, the dictatorship function has the largest second moment among all balanced functions, but it is not hard to see that the majority function, for example, has a much larger (relatively speaking) kth moment for very large values of k. To see this, note that for $k \gg 2^n$,

$$\mathbb{E}_Y\left((1-2P_Y^f)^{2k}\right) \gtrsim 2^{-n} \cdot \max_y |1-2P_y^f|^{2k}.$$

For the majority function, the maximum is attained at $y = (1, 1, , 1 \ldots, 1)$ for which $P_y^f \approx 2^{-nD\left(\frac{1}{2}||\alpha\right)}$ and consequently $\max_y |1-2P_y^{Maj}| \approx 1 - 2^{-nD\left(\frac{1}{2}||\alpha\right)}$. For dictatorship, on the other hand, $|1-2P_Y^f| = 1-2\alpha$ for every y, and therefore

$$\max_y |1-2P_y^{Maj}|^{2k} \gg \max_y |1-2P_y^{Dict}|^{2k}.$$

Therefore, one cannot hope to prove that there is a single function that simultaneously maximizes all moments; rather, the conjecture postulates that there is some tradeoff between these values and the largest mutual information is attained by functions that maximize lower moments at the expense of higher ones.

References

[1] T. Courtade and G. Kumar, "Which boolean functions maximize mutual information on noisy inputs?" *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4515–4525, Aug 2014.

[2] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–i," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 769–772, Nov 1973.

[3] E. Erkip, "The efficiency of information in investment," Ph.D. dissertation, Stanford University, CA, 1996.

[4] V. Chandar and A. Tchamkerten, "Most informative quantization functions," in *Proc. ITA Workshop, San Diego, CA, USA*, Feb. 2014, available online http://perso.telecom-paristech.fr/ tchamker/CTAT.pdf.

[5] R. Ahlswede and P. Gacs, "Spreading of sets in product spaces and hypercontraction of the markov operator," *The Annals of Probability*, vol. 4, no. 6, pp. 925–939, 1976.

[6] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On hypercontractivity and the mutual information between boolean functions," in *51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2013, pp. 13–19.

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

[8] A. Bonami, "tude des coefficients de fourier des fonctions de $l^p(g)$," *Annales de l'institut Fourier*, vol. 20, no. 2, pp. 335–402, 1970.

[9] W. Beckner, "Inequalities in fourier analysis on $\mathbb{R}^n$," *Proceedings of the National Academy of Sciences*, vol. 72, no. 2, pp. 638–641, 1975.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de