On the Most Informative Boolean Functions of the Very Noisy Channel

Hengjie Yang and Richard D. Wesel

Department of Electrical and Computer Engineering
University of California, Los Angeles, Los Angeles, CA 90095, USA
Email: {hengjie.yang, wesel}@ucla.edu

***Abstract*—Let X^n be a uniformly distributed n-dimensional binary vector, and Y^n be the result of passing X^n through a binary symmetric channel (BSC) with crossover probability α. A recent conjecture postulated by Courtade and Kumar states that $I(f(X^n);Y^n) \leq 1 - H(\alpha)$. Although the conjecture has been proved to be true in the dimension-free high noise regime by Samorodnitsky, here we present a calculus-based approach to show a dimension-dependent result by examining the second derivative of $H(\alpha) - H(f(X^n)|Y^n)$ at $\alpha = 1/2$. Along the way, we show that the dictator function is the most informative function in the high noise regime.**

I. Introduction

A. Previous Work

In [1], Courtade and Kumar postulated the following maximum mutual information conjecture:

Conjecture 1 ([1]): Let $X^n = (X_1, \dots, X_n)$ be a sequence of n i.i.d. Bernoulli (1/2) random variables, and let Y^n be the result of passing X^n through a memoryless binary symmetric channel (BSC) with crossover probability α. For any Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have

$$I(f(X^n);Y^n) \leq 1 - H(\alpha). \tag{1}$$

Intuitively, Conjecture 1 asks the following question: "*What is the most informative bit X^n can provide about Y^n?*" It can be readily verified that the dictator function, $f(X^n) = X_i, i \in \{1, 2, \cdots, n\}$, can achieve the equality in (1), suggesting that the dictator function might be the most informative function that X^n can reveal about Y^n. However, the rigorous proof of showing that the dictator function is indeed the most informative function is still elusive. Courtade and Kumar [1] showed that Conjecture 1 holds when $\alpha \to 0$ using an edge-isoperimetric argument.

The first result pertaining to Conjecture 1 dates back to the work of Wyner and Ziv [2], known as Mrs. Gerber's Lemma.

Theorem 1 (Mrs. Gerber's Lemma, [2]): Let X^n, Y^n be binary random-n vectors, which are input and output, respectively, of a binary symmetric channel with crossover probability α. Let $h(x) \triangleq -x\log x - (1-x)\log(1-x), x \in [0,1]$ be the binary entropy function. Let $H(X^n), H(Y^n)$ be the entropies of X^n, Y^n, respectively, with $H(X^n)$ satisfying $\frac{1}{n}H(X^n) \geq h(\pi_0), 0 \leq \pi_0 \leq 1$. Then

$$\frac{1}{n}H(Y^n) \geq h(\pi_0(1-\alpha) + (1-\pi_0)\alpha), \tag{2}$$

with equality if and only if $X^n = (X_1, X_2, \cdots, X_n)$ are independent and with $H(X_i) = h(\pi_0), i \in \{1, 2, \cdots, n\}$.

With Mrs. Gerber's Lemma, Erkip [3] showed the following universal upper bound on $I(f(X^n);Y^n)$

$$I(f(X^n);Y^n) \leq (1-2\alpha)^2, \quad \forall \alpha \in [0,1]. \tag{3}$$

However, (3) is still strictly weaker than (1).

In [4], Ordentlich, Shayevitz, and Weinstein used Fourier analytic techniques and leveraged hypercontractivity to improve the upper bound on $I(f(X^n);Y^n)$ for all balanced Boolean functions, i.e., $\mathscr{P}\{f(X^n) = 0\} = \mathscr{P}\{f(X^n) = 1\} = \frac{1}{2}$, which beats Erkip's bound in (3) when $\alpha > \frac{1}{3}$.

Theorem 2 ([4]): For any balanced Boolean function $f : \{0,1\}^n \to \{0,1\}$, and any $\frac{1}{2}(1 - \frac{1}{\sqrt{3}}) \leq \alpha \leq \frac{1}{2}$, we have that

$$I(f(X^n);Y^n) \leq \frac{\log_2 e}{2}(1-2\alpha)^2 + 9(1 - \frac{\log_2 e}{2})(1-2\alpha)^4. \tag{4}$$

As a corollary, they also showed that the dictator function is the most informative balanced function in the high noise regime.

So far, the most promising result is due to Samorodnitsky [5] who proved that Conjecture 1 holds in the dimension-free high noise regime, i.e., for $\alpha \in (\frac{1}{2} - \delta, \frac{1}{2} + \delta)$ with $\delta > 0$ being a dimension independent number, by considering the entropy of the image of f under a noise operator.

Theorem 3 ([5]): There exists an absolute $\delta > 0$ such that for any noise $\alpha > 0$ with $(1-2\alpha)^2 < \delta$ and for any Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have

$$I(f(X^n);Y^n) \leq 1 - H(\alpha). \tag{5}$$

In addition to Conjecture 1, some related conjectures are also addressed in [1]. One conjecture is that, for Boolean functions f_1, f_2, does it hold that

$$I(f_1(X^n); f_2(X^n)) \leq 1 - H(\alpha)? \tag{6}$$

This conjecture is then positively resolved by Pichler, Matz, and Piantanida [6] using Fourier-analytic arguments. The Gaussian analogy of Conjecture 1 is proved by Kindler, O'Donnell, and Witmer [7]. Anantharam *et al.* [8] conjectured a result related to the chordal slope of the hypercontractivity ribbon of a pair of binary random variables, which would imply (6). However, this stronger result still remains open.

Recently, a complementary problem concerning Conjecture 1 is posed and proved by Huleihel and Ordentlich [9].

Theorem 4 ([9]): For any function $f : \{0,1\}^n \to \{0,1\}^{n-1}$, we have

$$I(f(X^n); Y^n) \leq (n-1)(1 - H(\alpha)), \tag{7}$$

and this bound is attained with equality by, e.g., $f(x^n) = (x_1, x_2, \cdots, x_{n-1})$.

Li and Médard [10] studied the problem of maximizing the p-th moment of the image of f under noise operator and discussed the connection between noise stability and Conjecture 1.

B. Main Contributions

In this paper, we prove that Conjecture 1 holds in high noise regime by applying a calculus-based approach directly to (1). Namely, by defining $F_f(\alpha) \triangleq H(\alpha) - H(f(X^n)|Y^n)$ and $T \triangleq 1 - H(f(X^n))$, we prove that given any n and $|f^{-1}(0)|$, $F_f(\alpha) \leq T$ for $\alpha \in (\frac{1}{2} - \delta_n, \frac{1}{2} + \delta_n)$, where $\delta_n > 0$ is some small constant dimensionally dependent on n.

Given $|f^{-1}(0)|$, we say "Boolean function f is most informative in the high noise regime" if $I(f(X^n); Y^n)$ is undominated in small interval centered at $\frac{1}{2}$, i.e. if $F_f(\alpha)$ for this choice of f is greater than or equal to $F_f(\alpha)$ for any other choice of f with the same $|f^{-1}(0)|$. Since $F_f(1/2) = T, F_f'(1/2) = 0$ for any f, the most informative condition reduces to finding f that maximizes $F_f''(1/2)$.

As pointed out in [1], Conjecture 1 inherently consists of two components, the *structure* of maximally-informative Boolean functions that asks which Boolean function maximizes $I(f(X^n); Y^n)$ for a given n and fixed $|f^{-1}(0)|$, and the *inequality component* of the conjecture that concerns how to establish the inequality given that f is the maximally-informative function for a given n and fixed $|f^{-1}(0)|$. Indeed, the maximization of $F_f''(1/2)$ requires the solution to the above two components. In this paper, we introduce the notion of *ratio spectrum* of $f^{-1}(0)$, an integer sequence that characterizes the structure of $f^{-1}(0)$, which uniquely determines $F_f''(1/2)$. We then show that the lex function (all notions will be defined shortly) is *a* maximizing Boolean function for a given n and fixed $|f^{-1}(0)| < 2^{n-1}$ and the dictator function (a special case $f(X^n) = X_1$ of which is lex) is *the* maximizing Boolean function for $|f^{-1}(0)| = 2^{n-1}$. Finally, we establish the nonpositivity of $F_f''(1/2)$ given that f is lex and show that the dictator function is the only type of functions that achieve $F_f''(1/2) = 0$, thus implying its optimality.

Due to the space limitation, we only present the proof sketches. We refer readers to [11] for complete proof details.

II. Reformulation and Main Results

A. Reformulation

Let $\mathcal{S}$ denote the universal set of all n-dimensional binary sequences with $S = |\mathcal{S}| = 2^n$. The logarithm base is 2, whereas in Sec. IV, we assume natural logarithm. For a scalar $p \in [0,1]$, $H(p) \triangleq -p \log p - (1-p) \log(1-p)$ denotes the binary entropy function.

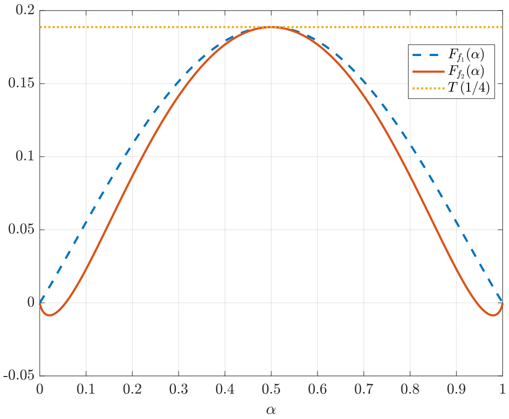

Fig. 1. An example of $F_f(\alpha)$ and $T(\mathscr{P}\{f(X^n) = 0\})$, with $n = 4, |f^{-1}(0)| = 4, \mathscr{P}\{f(X^n) = 0\} = 1/4$. We show two typical shapes of $F_f(\alpha)$: the quasi-concave shape for $F_{f_1}(\alpha)$ with $f_1^{-1}(0) = \{0,1,2,3\}$ and the "single-peak wave" shape for $F_{f_2}(\alpha)$ with $f_2^{-1}(0) = \{0,1,2,4\}$, where the element in the curly bracket represents the decimal representation of an n-dimensional binary sequence.

Define

$$\begin{aligned} F_f(\alpha) \triangleq& H(\alpha) - H(f(X^n)|Y^n) \\ =& H(\alpha) - \mathbb{E}_{Y^n} H(\mathscr{P}\{f(X^n) = 0|Y^n\}) \\ =& H(\alpha) - \frac{1}{S} \sum_{y^n \in \mathcal{S}} H(\mathscr{P}\{f(X^n) = 0|y^n, \alpha\}), \end{aligned} \tag{8}$$

and

$$T(\mathscr{P}\{f(X^n) = 0\}) \triangleq 1 - H(f(X^n)) = 1 - H\left(\frac{|f^{-1}(0)|}{2^n}\right) \tag{9}$$

where by letting $\overline{\alpha} = 1 - \alpha$,

$$\mathscr{P}\{f(X^n) = 0|y^n, \alpha\} = \sum_{x^n \in f^{-1}(0)} \alpha^{d(x^n, y^n)} \overline{\alpha}^{n - d(x^n, y^n)}. \tag{10}$$

With the above notation, Conjecture 1 translates to the following conjecture.

Conjecture 2: Given an n and fixed $|f^{-1}(0)|$, for any Boolean function f with the same $|f^{-1}(0)|$, we have

$$\max_{\alpha \in [0,1]} F_f(\alpha) = F_f\left(\frac{1}{2}\right) = T(\mathscr{P}\{f(X^n) = 0\}). \tag{11}$$

Note that it is trivial to show that $F_f(1/2) = T(\mathscr{P}\{f(X^n) = 0\})$ since when $\alpha = \frac{1}{2}$, X^n and Y^n are independent, which implies that $F_f(1/2) = H(1/2) - H(f(X^n)) = T(\mathscr{P}\{f(X^n) = 0\})$. Therefore, establishing that $\max_{\alpha \in [0,1]} F_f(\alpha) = F_f(1/2)$ for any Boolean function f is the key to solving Conjecture 1. Also worth mentioning is the dictator function, $f(X^n) = X_i, i \in \{1, 2, \cdots, n\}$, for which $\mathscr{P}\{f(X^n) = 0\} = 1/2$, $F_f(\alpha) = 0$ for $\alpha \in [0,1]$ and $T(\mathscr{P}\{f(X^n) = 0\}) = 0$.

As an example of the reformulation, Fig. 1 shows that $F_f(\alpha) \leq T(\mathscr{P}\{f(X^n) = 0\})$ for $n = 4$ and $|f^{-1}(0)| = 4$. Meanwhile, Fig. 1. also depicts two typical shapes of $F_f(\alpha)$: a quasi-concave shape as shown by $F_{f_1}(\alpha)$, and a "single-peak wave" shape as shown by $F_{f_2}(\alpha)$. In fact, we conjecture that these are the only two possible shapes of $F_f(\alpha), \alpha \in (0,1)$. Note that even for dictator function $f(X^n) = X_i, i \in \{1, 2, \cdots, n\}$, $F_f(\alpha) = 0$ is still quasi-concave.

B. Main Results

Our main result is the following theorem.

Theorem 5: Given an n and fixed $|f^{-1}(0)|$, for any Boolean function with the same $|f^{-1}(0)|$, there exists a small constant $\delta_n > 0$ independent of f such that

$$\max_{\alpha \in (\frac{1}{2}-\delta_n, \frac{1}{2}+\delta_n)} F_f(\alpha) = F_f\left(\frac{1}{2}\right) = T(\mathscr{P}\{f(X^n) = 0\}). \tag{12}$$

The entire paper is to establish Theorem 5 by proving the following lemmas.

Lemma 1: Given an n and fixed $|f^{-1}(0)|$, for any Boolean function f with the same $|f^{-1}(0)|$, $F_f(\alpha)$ is symmetric with respect to $\alpha = 1/2$.

As a corollary, combined with the fact that $F_f(\alpha)$ is differentiable, Lemma 1 also implies that

$$F'_f\left(\frac{1}{2}\right) = 0. \tag{13}$$

Lemma 2: Given an n and fixed $|f^{-1}(0)|$, for any Boolean function f with the same $|f^{-1}(0)|$, we have

$$F''_f\left(\frac{1}{2}\right) \leq 0, \tag{14}$$

where equality holds if and only if f is the dictator function.

Clearly, given an n and fixed $|f^{-1}(0)|$, the fact that $F_f(1/2) = T(\mathscr{P}\{f(X^n) = 0\})$, Lemma 1, and Lemma 2 lead to the existence of $\delta_{n,f} > 0$ such that $F_f(\alpha) \leq T(\mathscr{P}\{f(X^n) = 0\})$ for $\alpha \in (1/2 - \delta_{n,f}, 1/2 + \delta_{n,f})$. Letting $\delta_n = \min_{f:|f^{-1}(0)| \leq 2^n} \{\delta_{n,f}\}$ immediately implies Theorem 5.

By our previous most informative Boolean function argument, Lemma 2 clearly demonstrates that the dictator function is the most informative Boolean function in high noise regime.

III. Proof of Lemma 1

Consider $\alpha \in [0, 1/2]$ and its symmetric part $\overline{\alpha} = 1 - \alpha \in [1/2, 1]$. We have

$$\begin{aligned}\mathscr{P}\{f(X^n) = 0 | y^n, \alpha\} &= \sum_{x^n \in f^{-1}(0)} \alpha^{d(x^n, y^n)} \overline{\alpha}^{n - d(x^n, y^n)} \\ &= \sum_{x^n \in f^{-1}(0)} \alpha^{n - d(x^n, \bar{y}^n)} \overline{\alpha}^{d(x^n, \bar{y}^n)} \\ &= \mathscr{P}\{f(X^n) = 0 | \bar{y}^n, \overline{\alpha}\}\end{aligned} \tag{15}$$

where $\bar{y}_i = 1 - y_i, i = \{1, 2, \cdots, n\}$. Hence,

$$\begin{aligned}F_f(\alpha) &= H(\alpha) - \frac{1}{S} \sum_{y^n \in \mathcal{S}} H(\mathscr{P}\{f(X^n) = 0 | y^n, \alpha\}) \\ &= H(\overline{\alpha}) - \frac{1}{S} \sum_{y^n \in \mathcal{S}} H(\mathscr{P}\{f(X^n) = 0 | \bar{y}^n, \overline{\alpha}\}) \\ &= F_f(\overline{\alpha})\end{aligned} \tag{16}$$

which completes the proof of Lemma 1.

Similarly, an additional symmetry property for *complementary function* $1 - f$ with respect to f is presented as follows.

Theorem 6: Given n and M, $(0 \leq M \leq 2^n)$, for any Boolean function f, define its complementary function $f^c = 1 - f$ as $(f^c)^{-1}(0) = f^{-1}(1), (f^c)^{-1}(1) = f^{-1}(0)$. We have $F_f(\alpha) = F_{1-f}(\alpha)$.

Proof: By $H(p) = H(1-p), p \in [0, 1]$,

$$\begin{aligned}F_f(\alpha) &= H(\alpha) - \frac{1}{S} \sum_{y^n \in \mathcal{S}} H(1 - \mathscr{P}\{f(X^n) = 1 | y^n, \alpha\}) \\ &= H(\alpha) - \frac{1}{S} \sum_{y^n \in \mathcal{S}} H(\mathscr{P}\{f^c(X^n) = 1 | y^n, \alpha\}) \\ &= F_{1-f}(\alpha)\end{aligned} \tag{17}$$

■

The implication of Theorem 6 is that it suffices to focus on $1 \leq |f^{-1}(0)| \leq 2^{n-1}$.

IV. Proof of Lemma 2

The proof of Lemma 2 proceeds as follows: first, for a given n and Boolean function f with $|f^{-1}(0)| = M, 1 \leq M \leq 2^{n-1}$, we derive the general formula for $F''_f(1/2)$, which is uniquely determined by the *ratio spectrum* of $f^{-1}(0)$. Next, we prove that $\max_{f:|f^{-1}(0)|=M} F''_f(1/2) \leq 0$ for any $1 \leq M \leq 2^{n-1}$. Note that the entire proof now assumes natural logarithms unless otherwise specified.

During the proof, one can see that the *lex function* is *a* locally most informative function among all f with $|f^{-1}(0)| = M$, achieving $\max_{f:|f^{-1}(0)|=M} F''_f(1/2)$. In particular, if $M < 2^{n-1}$, we have $\max_{f:|f^{-1}(0)|=M} F''_f(1/2) < 0$. If $M = 2^{n-1}$, lex function becomes dictator function $f(X^n) = X_1$, achieving $\max_{f:|f^{-1}(0)|=2^{n-1}} F''_f(1/2) = 0$. By the uniqueness of the dictator function, we conclude that the dictator function is the most informative function in the high noise regime.

We first introduce several new definitions which will play an important role in proving Lemma 2. For brevity, let $M \triangleq |f^{-1}(0)|$ henceforth unless otherwise specified.

Definition 1: (*lex function*) Boolean function f is said to be lex when $f^{-1}(0)$ only contains the first M lexicographically ordered n-dimensional binary sequences.

Definition 2: ($0-1$ *ratio*) For $x^n = (x_1, x_2, \cdots, x_n)$, let $\gamma_k = \min\{\sum_{x^n \in f^{-1}(0)} \mathbb{1}_{\{x_k=0\}}, \sum_{x^n \in f^{-1}(0)} \mathbb{1}_{\{x_k=1\}}\}$, $k \in \{1, \ldots, n\}$, the $0-1$ ratio at k-th position is $\gamma_k/(M - \gamma_k)$. Clearly, $0 \leq \gamma_k \leq \lfloor \frac{M}{2} \rfloor$.

Definition 3: (*ratio spectrum*) The ratio spectrum of $f^{-1}(0)$ is defined by an integer sequence $\boldsymbol{R}_f = \{r_0, r_1, \cdots, r_{\lfloor \frac{M}{2} \rfloor}\}$, where $r_i = \sum_{k=1}^{n} \mathbb{1}_{\{\gamma_k = i\}}, i \in \{0, 1, \cdots, \lfloor \frac{M}{2} \rfloor\}$. Clearly, $\sum_{i=0}^{\lfloor \frac{M}{2} \rfloor} r_i = n$.

Definition 4: (*lexicographic ordering of ratio spectra*) The ratio spectrum $\boldsymbol{R}_f = \{r_0, r_1, \cdots, r_{\lfloor \frac{M}{2} \rfloor}\}$ is said to be (strictly) greater than $\boldsymbol{R}_{f'} = \{r'_0, r'_1, \cdots, r'_{\lfloor \frac{M}{2} \rfloor}\}$, denoted by $\boldsymbol{R}_f \succ \boldsymbol{R}_{f'}$, if and only if $r_j > r'_j$ for some j and $r_i = r'_i$ for all $i < j$.

First, we present the general formula of $F''_f(1/2)$ which is uniquely determined by ratio spectrum $\boldsymbol{R}_f$.

Theorem 7: Given n, for any Boolean function f, the ratio spectrum $\boldsymbol{R}_f = \{r_0, r_1, \cdots, r_{\lfloor \frac{M}{2} \rfloor}\}$ uniquely determines $F''_f(\alpha)$, that is,

$$F''_f\left(\frac{1}{2}\right) = -4 + \frac{4}{(S-M)M}\left(nM^2 - 4\sum_{t=0}^{\lfloor \frac{M}{2}\rfloor}(M-t)tr_t\right). \tag{18}$$

Proof: The derivation simply involves taking derivatives and simplifying combinatorial coefficients. See [11] for complete derivations. ■

Theorem 7 immediately implies the following corollary.

Corollary 1: For a given n, let $|f_1^{-1}(0)| = |f_2^{-1}(0)| = M, 1 \leq M \leq 2^{n-1}$, for Boolean functions f_1 and f_2. If $\boldsymbol{R}_{f_1} \succ \boldsymbol{R}_{f_2}$, then $F''_{f_1}(1/2) > F''_{f_2}(1/2)$.

Proof: Let $\boldsymbol{R}_{f_1} = \{r_0, \cdots, r_i, \cdots, r_j, \cdots, r_{\lfloor \frac{M}{2} \rfloor}\}$ and $\boldsymbol{R}_{f_2} = \{r_0, \cdots, r_i - 1, \cdots, r_j + 1, \cdots, r_{\lfloor \frac{M}{2} \rfloor}\}$, where $0 \leq i < j \leq \lfloor \frac{M}{2} \rfloor, r_i > 0, r_j > 0$. Obviously, $\boldsymbol{R}_{f_1} \succ \boldsymbol{R}_{f_2}$. By (18), we have

$$F''_{f_1}(1/2) - F''_{f_2}(1/2) = \frac{16(j-i)(M-i-j)}{(S-M)M} > 0. \tag{19}$$

In general, any other spectrum inequality can be established by successively constructing the above two "adjacent" ratio spectra. ■

Corollary 1 indicates that among Boolean functions with the same $|f^{-1}(0)|$, the one with the largest ratio spectrum maximizes $F''_f(1/2)$. The following theorem shows that the lex function is one type of functions with the largest ratio spectrum.

Theorem 8: For a given n, among all Boolean functions with the same $|f^{-1}(0)|$, the lex function is a function with the largest ratio spectrum $\boldsymbol{R}^*_f$.

Proof: It is enough to prove that when f is lex, $\boldsymbol{R}_f \succeq \boldsymbol{R}_{f'}$ always holds, where f' is any other Boolean function with the same $|f^{-1}(0)|$. This can be shown by contradiction.

Assume Boolean function f is lex and f' satisfies $\boldsymbol{R}_{f'} \succ \boldsymbol{R}_f$. This is only possible by first deleting bit 1's in $f^{-1}(0)$ and then performing any of the following operations:

- permutation: to permutate x_k's among all $x^n \in f^{-1}(0)$;
- flipping: to flip x_k for each $x^n \in f^{-1}(0)$;
- switching: to switch x_i and x_j for each $x^n \in f^{-1}(0)$.

However, only deletion of bit 1's can result in a $\boldsymbol{R}_{f'}$ greater than $\boldsymbol{R}_f$, which will also result in repetitive binary sequences in ${f'}^{-1}(0)$. Since the latter consequence contradicts the definition of Boolean functions, this means that the lex function has the largest ratio spectrum. ■

Note that, when $|f^{-1}(0)| < 2^{n-1}$ is fixed, there exist other Boolean functions that have the same largest ratio spectrum as the lex function. Therefore, a lex function can be referred to as *a locally most informative function*. However, when $|f^{-1}(0)| = 2^{n-1}$, the lex function reduces to dictator function $f(X^n) = X_1$. Since permutation and flipping still result in $f(X^n) = X_1$, and switching produces other dictator function $f(X^n) = X_i, 1 < i \leq n$, this implies that the dictator function is the only type of functions with the largest ratio spectrum. Next, we will show that the dictator function is also the only type of functions that can achieve $F''_f(1/2) = 0$. Functions other than dictator functions can only result in $F''_f(1/2) < 0$. Therefore, dictator functions are referred to as *the globally most informative function.*

Proposition 1: With Boolean function f being lex, $\sum_{t=0}^{\lfloor \frac{M}{2} \rfloor}(M-t)tr_t$ in (18) is constant if $n \geq \lceil \log_2 M \rceil$.

Proof: It is equivalent to examining the case when $0 \leq M \leq 2^n$. Since f is lex, it can be verified that

$$\sum_{x^n \in f^{-1}(0)} \mathbb{1}_{\{x_i=1\}} = \sum_{k=0}^{2^{n-i}-1} \left\lfloor \frac{M+k}{2^{n+1-i}} \right\rfloor. \tag{20}$$

for $i \in \{1, 2, \cdots, n\}$ and $x^n = (x_1, x_2, \cdots, x_n)$. Obviously, $\sum_{x^n \in f^{-1}(0)} \mathbb{1}_{\{x_i=1\}} = 0$ if $i \leq n - \lceil \log_2 M \rceil$.

Therefore, according to Definition 2 and Definition 3,

$$\begin{aligned}
&\sum_{t=0}^{\lfloor \frac{M}{2} \rfloor}(M-t)tr_t = \sum_{i=1}^{n}\sum_{t=0}^{\lfloor \frac{M}{2} \rfloor}(M-t)t \cdot \mathbb{1}_{\{\gamma_i = t\}} \\
&= \sum_{i=1}^{n}\left(M - \sum_{x^n \in f^{-1}(0)} \mathbb{1}_{\{x_i=1\}}\right)\left(\sum_{x^n \in f^{-1}(0)} \mathbb{1}_{\{x_i=1\}}\right) \\
&= \sum_{i=1}^{n}\left(M - \sum_{k=0}^{2^{n-i}-1}\left\lfloor \frac{M+k}{2^{n+1-i}} \right\rfloor\right)\left(\sum_{k=0}^{2^{n-i}-1}\left\lfloor \frac{M+k}{2^{n+1-i}} \right\rfloor\right),
\end{aligned} \tag{21}$$

which will be constant as long as $n \geq \lceil \log_2 M \rceil$. ■

For convenience, define for $n \geq \lceil \log_2 M \rceil$,

$$W_f(M) \triangleq \sum_{t=0}^{\lfloor \frac{M}{2} \rfloor}(M-t)tr_t \tag{22}$$

and

$$g(n) \triangleq -4 + \frac{4\Big(nM^2 - 4W_f(M)\Big)}{(2^n - M)M}, \tag{23}$$

which is $F''_f(\alpha)$ with $\alpha = 1/2$ and M fixed. We are now in a position to prove the last step of Lemma IV, the nonpositivity of $F''_f(1/2)$ for any choice of f.

Proposition 2: Assume f is lex and $n \geq \log_2 M + 1$ (since it suffices to examine $1 \leq M \leq 2^{n-1}$ by Theorem 6). We have

$$M^2 \log_2 M \leq 4W_f(M) < M^2\left(\log_2 M + \frac{2\ln 2 - 1}{2\ln 2}\right) \tag{24}$$

and

$$g(n) \leq g(\log_2 M + 1) \leq 0, \tag{25}$$

where in (25), the first equality holds if and only if $n = \log_2 M + 1$ and the second equality holds if and only if M is a power of 2.

Proof: We first prove (24). By replacing i with $n+1-i$ in (21), we have

$$(21) = \sum_{i=1}^{n}\left(M - \sum_{k=0}^{2^{i-1}-1}\left\lfloor\frac{M+k}{2^i}\right\rfloor\right)\left(\sum_{k=0}^{2^{i-1}-1}\left\lfloor\frac{M+k}{2^i}\right\rfloor\right). \tag{26}$$

It can be verified that for $i = 1, 2, \cdots, n$,

$$\sum_{k=0}^{2^{i-1}-1}\left\lfloor\frac{M+k}{2^i}\right\rfloor = \begin{cases} 2^{i-1}\left\lfloor\frac{1}{2}+\frac{M}{2^i}\right\rfloor, \left\lfloor\frac{M}{2^{i-1}}\right\rfloor \equiv 0; \\ M - 2^{i-1}\left\lfloor\frac{1}{2}+\frac{M}{2^i}\right\rfloor, \left\lfloor\frac{M}{2^{i-1}}\right\rfloor \equiv 1, \end{cases} \tag{27}$$

where $\equiv$ denotes modulo-2. Therefore, substituting (27) into (26) yields

$$\begin{aligned} W_f(M) =& \sum_{i=1}^{n}\left(M - 2^{i-1}\left\lfloor\frac{1}{2}+\frac{M}{2^i}\right\rfloor\right)\left(2^{i-1}\left\lfloor\frac{1}{2}+\frac{M}{2^i}\right\rfloor\right) \\ \triangleq& a(M-1). \end{aligned} \tag{28}$$

By [12], $a(M)$ can be computed recursively as follows.

$$\begin{aligned} a(0) &= 0 \\ a(2M) &= 2a(M) + 2a(M-1) + M(M+1) \\ a(2M+1) &= 4a(M) + (M+1)^2, \end{aligned}$$

for $M \in \mathbb{Z}^+$. Thus by induction, it can be shown that

$$M^2\log_2 M \le 4a(M-1) < M^2\left(\log_2 M + \frac{2\ln 2 - 1}{2\ln 2}\right), \tag{29}$$

where the equality holds if and only if M is a power of 2. Complete details of induction can be found in Appendix, [11].

Now we prove (25). Since f is lex and $n \ge \log_2 M + 1$, making n a continuous variable, $g'(n)$ is computed as follows

$$g'(n) = 4\cdot\frac{(2^n - M)M^2 - (nM^2 - 4W_f(M))2^n\ln 2}{(2^n - M)^2 M}. \tag{30}$$

In order to show that $g'(n) < 0$ for $n \ge \log M + 1$, we need to show

$$\begin{aligned} 4W_f(M) <& \min_{n\ge\log_2 M+1}\left\{\frac{M^3}{2^n\ln 2} + \left(n - \frac{1}{\ln 2}\right)M^2\right\} \\ =& M^2\left(\log_2 M + \frac{2\ln 2 - 1}{2\ln 2}\right), \end{aligned} \tag{31}$$

which has just been proved in (29). Therefore, $g(n)$ is a monotonically decreasing function and by (29)

$$\begin{aligned} g(n) \le& g(\log_2 M + 1) \\ =& -4 + \frac{4\Big((\log_2 M + 1)M^2 - 4W_f(M)\Big)}{(2^{\log_2 M+1} - M)M} \\ \le& -4 + \frac{4}{M^2}\Big[(\log_2 M + 1)M^2 - M^2\log_2 M\Big] \\ =& 0. \end{aligned} \tag{32}$$

The proof is completed. ■

Proposition 2 implies that for any Boolean function f being lex, $F_f''(1/2) \le 0$, where the equality holds if and only if M is a power of 2 and $n = \log_2 M + 1$, suggesting that f can only be the dictator function. By the aforementioned most informative Boolean function argument, we conclude that dictator function is the globally most informative function in the high noise regime, among all possible choices of f.

V. Discussion

In this paper, we establish Conjecture 1 in the high noise regime by looking at the derivatives of $F_f(\alpha)$. The limitations and future directions of this approach are as follows:

A limitation is that δ_n in Theorem 5 is dimensionally dependent on n, which weakens our result compared to Samorodnitsky's, where δ is a universal, dimension-free constant.

As a future direction, we note that to the best of our knowledge, current numerical exhaustive search indicates that there are only two possible shapes of $F_f(\alpha), \alpha \in (0,1)$, as depicted in Fig. 1. This seems promising for future investigation. The calculus-based approach presented in this paper could possibly facilitate a new approach to tackle Conjecture 1. Alternatively, the new function $F_f(\alpha)$ for $\alpha \in (0,1)$ might be analyzed using other techniques.

Acknowledgment

We are grateful to Jiange Li for valuable comments on an earlier version of this paper.

References

[1] T. A. Courtade and G. R. 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, November 1973.

[3] E. Erkip and T. M. Cover, "The efficiency of investment information," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1026–1040, May 1998.

[4] O. Ordentlich, O. Shayevitz, and O. Weinstein, "An improved upper bound for the most informative boolean function conjecture," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 500–504.

[5] A. Samorodnitsky, "On the entropy of a noisy function," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5446–5464, Oct 2016.

[6] G. Pichler, G. Matz, and P. Piantanida, "A tight upper bound on the mutual information of two boolean functions," in *2016 IEEE Information Theory Workshop (ITW)*, Sept 2016, pp. 16–20.

[7] G. Kindler, R. O'Donnell, and D. Witmer, "Remarks on the most informative function conjecture at fixed mean." [Online]. Available: http://arxiv.org/abs/1506.03167

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

[9] W. Huleihel and O. Ordentlich, "How to quantize n outputs of a binary symmetric channel to $n-1$ bits?" in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 91–95.

[10] J. Li and M. Médard, "Boolean functions: Noise stability, non-interactive correlation, and mutual information," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 266–270.

[11] H. Yang and R. D. Wesel, "On the most informative boolean functions of the very noisy channel." [Online]. Available: http://arxiv.org/abs/1807.11289

[12] A. Kundgen, "The on-line encyclopedia of integer sequences, 2003, sequence a022560."