

# Which Boolean Functions are Most Informative?

Gowtham R. Kumar and Thomas A. Courtade

Department of Electrical Engineering

Stanford University

Stanford, California, USA

Email: {gowthamr, courtade}@stanford.edu

**Abstract**—We introduce a simply stated conjecture regarding the maximum mutual information a Boolean function can reveal about noisy inputs. Specifically, let  $X^n$  be i.i.d. Bernoulli(1/2), and let  $Y^n$  be the result of passing  $X^n$  through a memoryless binary symmetric channel with crossover probability  $\alpha$ . For any Boolean function  $b : \{0, 1\}^n \rightarrow \{0, 1\}$ , we conjecture that  $I(b(X^n); Y^n) \leq 1 - H(\alpha)$ . While the conjecture remains open, we provide substantial evidence supporting its validity.

## I. INTRODUCTION

This paper is inspired by the following conjecture:

**Conjecture 1.** *Let  $X^n$  be i.i.d. Bernoulli(1/2), and let  $Y^n$  be the result of passing  $X^n$  through a memoryless binary symmetric channel with crossover probability  $\alpha$ . For any Boolean function  $b : \{0, 1\}^n \rightarrow \{0, 1\}$ , we have*

$$I(b(X^n); Y^n) \leq 1 - H(\alpha). \quad (1)$$

At first sight, Conjecture 1 might appear suitable as a homework exercise for a first course on information theory. However, over the course of this paper, we hope to convince the reader that the conjecture is much deeper than it appears. Despite its apparent simplicity, standard information-theoretical manipulations appear incapable of establishing (1).

To the present authors, Conjecture 1 represents the simplest, nontrivial embodiment of Boolean functions in an information-theoretic context. In words, Conjecture 1 asks: “*What is the most significant bit that  $X^n$  can provide about  $Y^n$ ?*”

Despite their fundamental roles in computer science and digital computation, Boolean functions have received relatively little attention from the information theory community. The recent work [1] is perhaps most relevant to our Conjecture 1 and provides compelling motivation for its study. In [1], the authors prove that for  $n$  and  $\mathcal{P}\{b(X^n) = 0\} \geq 1/2$  fixed,  $I(b(X^n); X_1)$  is maximized by functions  $b$  which satisfy  $b(X^n) = 0$  whenever  $X_1 = 0$  (i.e., when  $b$  is *canalizing* in  $X_1$ ). The motivation for considering this problem comes from computational biology, where Boolean networks are used to model dependencies in various regulatory networks. We encourage the reader to refer to [1], [2] and the references therein for further information.

We remark that we have proven the weaker inequality

$$\sum_{i=1}^n I(b(X^n); Y_i) \leq 1 - H(\alpha) \quad (2)$$

using Fourier-analytic techniques similar to those employed in [1]. However, this Fourier-analytic approach appears incapable

of establishing the stronger statement of Conjecture 1. We omit the proof of (2) in this paper due to space constraints.

Conjecture 1 is also related to the *Information Bottleneck Method* [3], which attempts to solve the optimization problem

$$\min_{p(u|x^n)} I(X^n; U) - \lambda I(Y^n; U). \quad (3)$$

For a given  $\lambda > 0$ , the optimizing  $U$  is purportedly the best tradeoff between the accuracy of describing  $Y^n$  and the descriptive complexity of  $U$ . In our setting,  $b(X^n)$  plays the role of  $U$ , and we constrain the descriptive complexity to be at most one bit. It is relatively easy to show that randomized Boolean functions do not yield a higher mutual information. Thus, expressing Conjecture 1 in terms of deterministic Boolean functions comes without loss of generality.

A more concrete example comes in the context of gambling. To this end, suppose  $Y^n$  is a simple model for a market of  $n$  stocks, where each stock doubles in value or goes bankrupt with probability 1/2, independent of all other stocks. If an oracle has access to side information  $X^n$ , and we are allowed to ask *one* yes/no question of the oracle, which question should we ask to maximize the rate at which our wealth grows? The validity of Conjecture 1 would imply that we should only concern ourselves with the performance of a single stock; say  $Y_1$ . This is readily seen as a consequence of known results on gambling with side information [4, Theorem 6.2.1], since putting  $b(X^n) = X_1$  yields

$$I(b(X^n); Y^n) = I(X_1; Y^n) = I(X_1; Y_1) = 1 - H(\alpha), \quad (4)$$

hence the conjectured upper bound (1) is attainable and represents the maximum possible increase in doubling rate.

Finally, we point out that (1) is related in spirit to the notion of *average sensitivity* of Boolean functions. This topic has received a great deal of attention in the computer science literature (cf. [5]). To see the connection to sensitivity, note that (1) can be rewritten as

$$H(b(X^n)|Y^n) \geq H(b(X^n)) - 1 + H(\alpha). \quad (5)$$

For fixed  $\mathcal{P}\{b(X^n) = 0\}$ , the right hand side of (5) is constant. Hence, the conjecture essentially lower bounds the output uncertainty of Boolean functions with respect to noisy inputs.

This paper is organized as follows. Section II provides a summary of the main results and their implications. It includes

a refinement of Conjecture 1 by splitting it into two “sub-conjectures.” The following section deals with the proofs of the main results. Section IV delivers concluding remarks.

## II. RESULTS AND IMPLICATIONS

Let  $X^n$  be a sequence of i.i.d. Bernoulli  $(1/2)$  random variables,  $Z^n$  be a sequence of i.i.d. Bernoulli  $(\alpha)$  random variables independent of  $X^n$ ,  $0 \leq \alpha \leq 1/2$ . Let  $Y^n = X^n \oplus Z^n$ , where “ $\oplus$ ” denotes coordinate-wise XOR. Throughout, we let  $\Omega = \{0, 1\}$ ,  $\Omega_n = \{0, 1\}^n$ , and consider Boolean functions  $b : \Omega_n \rightarrow \Omega$ .

**Definition 1.** The lexicographical ordering  $\prec_L$  on  $\{0, 1\}^k$  is defined as follows:  $x^k \prec_L \tilde{x}^k$  iff  $x_j < \tilde{x}_j$  for some  $j$  and  $x_i = \tilde{x}_i$  for all  $i < j$ .

For example, if  $k = 3$ , we have  $000 \prec_L 001 \prec_L 010 \prec_L 011 \prec_L 100 \prec_L 101 \prec_L 110 \prec_L 111$ .

**Definition 2.** We define  $L_k(M)$  to be the initial segment of size  $M$  in the lexicographical ordering on  $\{0, 1\}^k$ . For example,  $L_3(4) = \{000, 001, 010, 011\}$ .

For a function  $b : \Omega_n \rightarrow \Omega$ , we say that “ $b$  is lex” when  $b^{-1}(0) = L_n(|b^{-1}(0)|)$ . In other words,  $b$  is lex when it maps an initial segment of the lexicographical order to 0, and the complement segment to 1.

Instead of dealing with Conjecture 1 directly, consider the following two conjectures:

**Conjecture 2.** For a given  $n$  and fixed cardinality  $|b^{-1}(0)|$ , the conditional entropy  $H(b(X^n)|Y^n)$  is minimized when  $b$  is lex.

**Conjecture 3.** If  $b : \Omega_n \rightarrow \Omega$  is lex, then

$$H(b(X^n)|Y^n) \geq H(b(X^n))H(\alpha). \quad (6)$$

Clearly, Conjecture 1 would follow as a corollary if Conjectures 2 and 3 were valid.

Referring to Conjecture 3 as a “conjecture” is perhaps too modest. Indeed, we derive a simple recursive algorithm capable of proving (6) for any fixed  $\alpha$ . With the assistance of a computer, this algorithm has verified (6) for  $\alpha$  ranging from 0 to  $1/2$  in increments of 0.001. Refer to Theorem 3 and the following discussion in Section II-B for details.

### A. Conjecture 2 and Isoperimetry

Conjecture 2 is reminiscent of a classical theorem in discrete mathematics originally due to Harper [6] that gives an exact edge-isoperimetric inequality for the hypercube. To state the theorem, we need a few basic notations. Let  $Q_n$  be the  $n$ -dimensional hypercube, and let  $V(Q_n) = \Omega_n$  be its set of vertices. For  $S \subseteq V(Q_n)$ , the edge boundary  $\partial(S)$  is the set of edges one has to delete to disconnect  $S$  from any vertex not in  $S$ .

**Theorem 1.** For  $S \subseteq V(Q_n)$  with  $|S| = k$ , we have  $|\partial(S)| \geq |\partial(L_n(k))|$ .

The simplest proofs of Theorem 1 rely on so-called *compression operators*, popularized by Bollobás and Leader [7]. These compression operators turn out to be useful in making progress towards Conjecture 2, so we introduce them now.

Let  $\mathcal{I}$  be subset of  $\{1, 2, \dots, n\}$  of cardinality  $k$ . To be concrete, let  $\mathcal{I} = \{i_1, i_2, \dots, i_k\}$ , where  $i_1 < i_2 < \dots < i_k$ . For a set  $B \subseteq \Omega_n$  and  $x^n$  having  $x_i = 0$  for all  $i \in \mathcal{I}$ , we define the  $\mathcal{I}$ -section of  $B$  at  $x^n$  by

$$B_{\mathcal{I}}(x^n) = \left\{ z^k : y^n \in B, y_i = \begin{cases} z_j & \text{if } i = i_j \in \mathcal{I} \\ x_i & \text{otherwise} \end{cases} \right\}. \quad (7)$$

For instance, if  $B = \{000, 001, 011, 101\}$ , then examples of  $\mathcal{I}$ -sections at different  $x^3 \in \Omega_3$  are given by:

$$B_{\{1\}}(001) = \{0, 1\}, \quad (8)$$

$$B_{\{2\}}(100) = \emptyset, \quad (9)$$

$$B_{\{1,2\}}(000) = \{00\}, \quad (10)$$

$$B_{\{1,2\}}(001) = \{00, 01, 10\}. \quad (11)$$

The  $\mathcal{I}$ -compression of  $B$ ,  $C_{\mathcal{I}}(B)$ , is defined in terms of its  $\mathcal{I}$ -sections

$$(C_{\mathcal{I}}(B))_{\mathcal{I}}(x^n) = L_k(|B_{\mathcal{I}}(x^n)|).$$

In other words,  $C_{\mathcal{I}}$  replaces each  $\mathcal{I}$ -section of  $B$  with an initial segment of the lexicographical order. We say that  $B$  is  $\mathcal{I}$ -compressed if  $C_{\mathcal{I}}(B) = B$ . Note that  $C_{\mathcal{I}}(B)$  is always  $\mathcal{I}$ -compressed.

Continuing the above example of  $B = \{000, 001, 011, 101\}$ , example  $\mathcal{I}$ -compressions are given by:

$$C_{\{1\}}(B) = \{000, 001, 011, 101\}, \quad (12)$$

$$C_{\{2,3\}}(B) = \{000, 001, 010, 100\}. \quad (13)$$

We pause to make two important observations. First,  $\mathcal{I}$ -compression preserves the size of the set on which it operates. That is,  $|C_{\mathcal{I}}(B)| = |B|$ . Second, if  $B$  is  $\mathcal{I}$ -compressed, then it is also  $\mathcal{J}$ -compressed for all  $\mathcal{J} \subset \mathcal{I}$ .

The following theorem states that when  $|\mathcal{I}| = 2$ , applying an  $\mathcal{I}$ -compression to  $b^{-1}(0)$  does not decrease the information  $b(X^n)$  reveals about  $Y^n$ . Thus, compression provides a method of modifying functions in a manner that does not adversely affect the mutual information  $I(b(X^n); Y^n)$ .

**Theorem 2.** Let  $b : \Omega_n \rightarrow \Omega$  and let  $\mathcal{I} \subset \{1, \dots, n\}$  satisfy  $|\mathcal{I}| = 2$ . If  $\hat{b} : \Omega_n \rightarrow \Omega$  is defined by its preimage  $\hat{b}^{-1}(0) = C_{\mathcal{I}}(b^{-1}(0))$ , then  $I(\hat{b}(X^n); Y^n) \geq I(b(X^n); Y^n)$ .

By definition, if  $C_{\mathcal{I}}(\cdot)$  changes an element of  $b^{-1}(0)$ , it moves it *lower* in the lexicographical ordering on  $\Omega_n$ . Therefore, one can repeatedly apply Theorem 2 for different subsets  $\mathcal{I}$  of cardinality 2, ultimately terminating at a function  $\hat{b}$  which is  $\mathcal{I}$ -compressed for all  $\mathcal{I}$  with  $|\mathcal{I}| \leq 2$ . Hence, we have the following corollary.

**Corollary 1.** Let  $\mathcal{S}_n$  be the set of functions  $b : \Omega_n \rightarrow \Omega$  for which  $b^{-1}(0)$  is  $\mathcal{I}$ -compressed for all  $\mathcal{I}$  with  $|\mathcal{I}| \leq 2$ . In

$n$	$ \mathcal{S}_n $	$ \mathcal{B}_n $
2	5	16
3	10	256
4	25	65,536
5	119	$4.3 \times 10^9$
6	1173	$1.8 \times 10^{19}$
7	44,315	$3.4 \times 10^{38}$

TABLE I: Reduction in number of candidate Boolean functions to be considered for verification of Conjecture 2.

maximizing  $I(b(X^n); Y^n)$ , it is sufficient to consider functions  $b \in \mathcal{S}_n$ .

The implications of Theorem 2 and its corollary are twofold. First, it allows the verification of Conjecture 2 for modest values of  $n$ . Indeed, we have numerically validated Conjectures 1 and 2 for  $n \leq 7$  by evaluating  $I(b(X^n); Y^n)$  for  $b \in \mathcal{S}_n$ . To appreciate the reduction afforded by Corollary 1, define  $\mathcal{B}_n$  to be the set of all  $2^{2^n}$  Boolean functions on  $n$  inputs. A comparison between  $|\mathcal{S}_n|$  and  $|\mathcal{B}_n|$  is given in Table I.

Second, Theorem 2 reinforces the intuition behind Conjecture 2. As we noted above, if  $C_{\mathcal{I}}(\cdot)$  changes an element of  $b^{-1}(0)$ , it moves it *lower* in the lexicographical ordering on  $\Omega_n$ . Thus, roughly speaking, applying  $\mathcal{I}$ -compression to  $b^{-1}(0)$  yields a function  $\hat{b}$  which is (i) closer to an initial segment of the lexicographical order, and (ii) for  $|\mathcal{I}| \leq 2$  satisfies  $H(\hat{b}(X^n)|Y^n) \leq H(b(X^n)|Y^n)$ .

Ideally, Theorem 2 should generalize to include  $\mathcal{I}$ -compressions for  $|\mathcal{I}| > 2$ . Indeed, if we could take  $|\mathcal{I}| = n$ , Conjecture 2 would be proved. However, we have found counterexamples where compression *increases*  $H(b(X^n)|Y^n)$  for  $|\mathcal{I}| > 2$  (but still reduces  $H(b(X^n)|Y^n)$  for  $|\mathcal{I}| = n$ ). We omit the details.

### B. An Algorithmic Proof of Conjecture 3

Now, we turn toward establishing Conjecture 3. Unless otherwise specified, all Boolean functions in this subsection are assumed to be lex.

Define  $f(x) = -x \log x$ . Note that if  $b$  is lex, then so is  $\neg b$  (i.e., the negation of  $b$ ) up to a relabeling of inputs. Therefore, to prove (6), it is sufficient to prove

$$\mathbb{E}_{Y^n} f(\mathcal{P}\{b(X^n) = 0|Y^n\}) \geq f(\mathcal{P}\{b(X^n) = 0\}) H(\alpha). \quad (14)$$

To simplify notation, for a dyadic rational  $p = k/2^n$ , define

$$T_\alpha(p) \triangleq \mathbb{E}_{Y^n} f(\mathcal{P}\{b(X^n) = 0|Y^n\}), \quad (15)$$

where  $b$  is the unique lex function on  $n$  inputs with  $\mathcal{P}\{b(X^n) = 0\} = p$ . Note that if  $k$  is even,  $b$  does not depend on its input bit  $x_n$ . Therefore,  $T_\alpha(p)$  is well-defined for all dyadic rationals  $p \in [0, 1]$ . It is a simple exercise to show that  $T_\alpha(\cdot)$  is continuous on the dyadic rationals (in fact, it is Hölder continuous with exponent  $1/2$ ). Therefore,  $T_\alpha(p)$

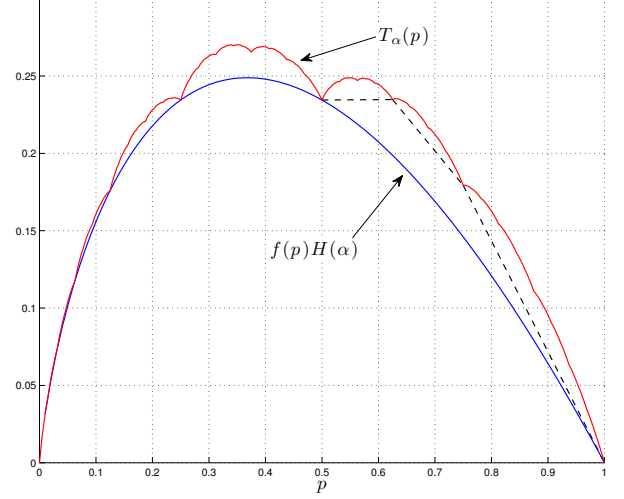


Fig. 1: A comparison of  $T_\alpha(p)$  and  $f(p)H(\alpha)$  for  $\alpha = 0.1$ . The broken line shows the three chords Algorithm II.1 constructs before terminating.

is also well-defined when  $p \in [0, 1]$  is not a dyadic rational, by considering its unique continuous extension to  $[0, 1]$ .

Thus, the validity of (14) for all lex  $b$  (and all  $n$ ) is equivalent to the inequality

$$T_\alpha(p) \geq f(p) H(\alpha) \quad \forall p \in [0, 1], \quad (16)$$

motivating the following theorem.

**Theorem 3.** Fix  $\alpha \in (0, 1/2)$ . If a call to Algorithm II.1 with arguments  $(p_-, p_+) = (1/2, 1)$  eventually terminates, then Conjecture 3 is true for the chosen  $\alpha$ .

#### Algorithm II.1: TESTINEQUALITY( $p_-, p_+$ )

```

main
  if CHECKCHORD( $p_-, p_+$ ) < 0
    then  $\begin{cases} p \leftarrow \frac{1}{2}(p_- + p_+) \\ \text{TESTINEQUALITY}(p_-, p) \\ \text{TESTINEQUALITY}(p, p_+) \end{cases}$ 

  procedure CHECKCHORD( $a, b$ )
    comment:  $C(x)$  is the chord connecting
               the points  $(a, T_\alpha(a))$  and  $(b, T_\alpha(b))$ .
     $C(x) := \frac{T_\alpha(b) - T_\alpha(a)}{b - a}(x - a) + T_\alpha(a)$ 
     $\nu \leftarrow \min_{x \in [a, b]} C(x) - f(x)H(\alpha)$ 
    return ( $\nu$ )

```

**Remark 1.** In the subroutine CHECKCHORD( $a, b$ ) of Algorithm II.1, the minimization has a closed form solution.

In words, Algorithm II.1 recursively constructs a piecewise linear function on the interval  $p \in [1/2, 1]$  which simultaneously upper bounds  $f(p)H(\alpha)$  and lower bounds  $T_\alpha(p)$ . As discussed in Section III-B, this is sufficient to prove (16). Figure 1 illustrates this procedure for  $\alpha = 0.1$ .

Using a Matlab implementation of Algorithm II.1, we have validated (16) for  $\alpha$  ranging from 0 to 1/2 in increments of 0.001. Hence, it is reasonable to believe that Conjecture 3 is true in general.

Despite the apparent gap between  $T_\alpha(p)$  and  $f(p)H(\alpha)$  for  $p \in (1/2, 1)$  (e.g., Fig. 1), the oscillatory behavior of  $T_\alpha(p)$  seems to render traditional analysis techniques ineffective in establishing (16). This was our motivation for pursuing an algorithmic proof. To get a sense for the strange behavior of  $T_\alpha(p)$ , we point out that it is possible to show that  $\lim_{\alpha \rightarrow 0} T_\alpha(p)/H(\alpha)$  is equal to the *Takagi function*, a classical construction of an everywhere-continuous, nowhere-differentiable function closely related to the edge-isoperimetric inequality given in Theorem 1 (cf. [8], [9]). We omit the details due to space constraints.

### III. PROOFS

#### A. Proof of Theorem 2

We begin the proof of Theorem 2 by first proving the following result for 1-dimensional compressions.

**Lemma 1.** *Let  $b : \Omega_n \rightarrow \Omega$  and  $i \in \{1, 2, \dots, n\}$ . If  $\hat{b} : \Omega_n \rightarrow \Omega$  is defined by its preimage  $\hat{b}^{-1}(0) = C_{\{i\}}(b^{-1}(0))$ , then  $I(\hat{b}(X^n); Y^n) \geq I(b(X^n); Y^n)$ .*

*Proof:* It suffices to consider the case where  $i = n$ , as any other case can be handled by first permuting coordinates.

Define  $B = b^{-1}(0)$  and  $\hat{b}^{-1}(0) = C_{\{n\}}(B)$ , and let

$$E_0 = \{x^{n-1} : B_{\{n\}}(x^{n-1}, 0) = \{0\}\} \quad (17)$$

$$E_1 = \{x^{n-1} : B_{\{n\}}(x^{n-1}, 0) = \{1\}\}, \quad (18)$$

where  $B_{\{n\}}$  is defined by (7) with  $\mathcal{I} = \{n\}$ . Define  $\check{b}(x^{n-1}, x_n) := \hat{b}(x^{n-1}, \neg x_n)$ , where  $\neg x_n$  is the negation of  $x_n$ .

By definition of  $\hat{b}$  and  $\check{b}$ , we have the identities

$$\begin{aligned} & \mathcal{P}\{b(X^n) = 0|y^{n-1}, 0\} \\ &= \mathcal{P}\{\hat{b}(X^n) = 0|y^{n-1}, 0\} - (1 - 2\alpha)\mathcal{P}\{X^{n-1} \in E_1|y^{n-1}\} \\ &= \mathcal{P}\{\check{b}(X^n) = 0|y^{n-1}, 0\} + (1 - 2\alpha)\mathcal{P}\{X^{n-1} \in E_0|y^{n-1}\}. \end{aligned}$$

Similar identities hold for  $\mathcal{P}\{b(X^n) = 0|y^{n-1}, 1\}$  with opposite signs on the  $(1 - 2\alpha)$  terms, giving

$$\begin{aligned} & \mathcal{P}\{b(X^n) = 0|y^n\} = \\ & \theta \mathcal{P}\{\hat{b}(X^n) = 0|y^n\} + (1 - \theta)\mathcal{P}\{\check{b}(X^n) = 0|y^n\}, \end{aligned} \quad (19)$$

where

$$\theta = \frac{\mathcal{P}\{X^{n-1} \in E_0|y^{n-1}\}}{\mathcal{P}\{X^{n-1} \in E_0|y^{n-1}\} + \mathcal{P}\{X^{n-1} \in E_1|y^{n-1}\}}. \quad (20)$$

Concavity of entropy implies that

$$\theta H(\hat{b}(X^n)|y^n) + (1 - \theta)H(\check{b}(X^n)|y^n) \leq H(b(X^n)|y^n).$$

Noting that  $\theta$  only depends on  $y^{n-1}$ , we average both sides over  $y_n \in \{0, 1\}$  to obtain

$$\begin{aligned} & \theta H(\hat{b}(X^n)|y^{n-1}, Y_n) + (1 - \theta)H(\check{b}(X^n)|y^{n-1}, Y_n) \\ & \leq H(b(X^n)|y^{n-1}, Y_n). \end{aligned} \quad (21)$$

By symmetry,  $H(\hat{b}(X^n)|y^{n-1}, Y_n) = H(\check{b}(X^n)|y^{n-1}, Y_n)$ . Therefore, averaging (21) over all values of  $y^{n-1}$ , we can conclude that  $H(\hat{b}(X^n)|Y^n) \leq H(b(X^n)|Y^n)$ . To complete the proof, we recall that  $|\hat{b}^{-1}(0)| = |C_{\{n\}}(b^{-1}(0))| = |b^{-1}(0)|$ . Combined with the fact that  $X^n$  is uniformly distributed on  $\Omega_n$ , this implies that  $H(\hat{b}(X^n)) = H(b(X^n))$ , as desired. ■

We are now in a position to finish the proof of Theorem 2, which is similar to the proof of Lemma 1.

*Proof of Theorem 2:* We assume that  $\mathcal{I} = \{n - 1, n\}$ , as all other cases follow by a permutation of coordinates. To simplify notation, we write  $B = b^{-1}(0)$ .

By a repeated application of Lemma 1, we can assume that  $B$  is  $\{n - 1\}$ - and  $\{n\}$ -compressed. Thus, the  $\mathcal{I}$ -sections  $B_{\mathcal{I}}(x^n)$  can only be one of the following:  $\emptyset$ ,  $\{00\}$ ,  $\{00, 01\}$ ,  $\{00, 10\}$ ,  $\{00, 01, 10\}$ , or  $\{00, 01, 10, 11\}$ . Note that all of these sets are initial segments of the lexicographical order on  $\Omega_2$  except  $\{00, 10\}$ . Hence, we aim to transform  $B$  so that  $B_{\mathcal{I}}(x^n) \neq \{00, 10\}$ . To this end, define

$$G_0 = \{x^n : x_{n-1} = x_n = 0, B_{\mathcal{I}}(x^n) = \{00, 01\}\} \quad (22)$$

$$G_1 = \{x^n : x_{n-1} = x_n = 0, B_{\mathcal{I}}(x^n) = \{00, 10\}\}. \quad (23)$$

Now, define  $\hat{b}$  by  $\hat{b}^{-1}(0) = C_{\mathcal{I}}(B)$  and the function  $\check{b}$  by permuting the last two coordinates:

$$\check{b}(x^{n-2}, x_n, x_{n-1}) = \hat{b}(x^{n-2}, x_{n-1}, x_n). \quad (24)$$

It is relatively straightforward to show that

$$\begin{aligned} & \mathcal{P}\{b(X^n) = 0|y^n\} = \\ & \theta \mathcal{P}\{\hat{b}(X^n) = 0|y^n\} + (1 - \theta)\mathcal{P}\{\check{b}(X^n) = 0|y^n\}, \end{aligned} \quad (25)$$

where

$$\theta = \frac{\mathcal{P}\{X^{n-2} \in G_0|y^{n-2}\}}{\mathcal{P}\{X^{n-2} \in G_0|y^{n-2}\} + \mathcal{P}\{X^{n-2} \in G_1|y^{n-2}\}}. \quad (26)$$

Concavity of entropy implies that

$$\theta H(\hat{b}(X^n)|y^n) + (1 - \theta)H(\check{b}(X^n)|y^n) \leq H(b(X^n)|y^n).$$

Noting that  $\theta$  only depends on  $y^{n-2}$ , we average both sides over  $y_{n-1}, y_n$  to obtain

$$\begin{aligned} & \theta H(\hat{b}(X^n)|y^{n-2}, Y_{n-1}^n) + (1 - \theta)H(\check{b}(X^n)|y^{n-2}, Y_{n-1}^n) \\ & \leq H(b(X^n)|y^{n-2}, Y_{n-1}^n). \end{aligned} \quad (27)$$

Crucially, the symmetry (24) implies that

$$H(\hat{b}(X^n)|y^{n-2}, Y_{n-1}^n) = H(\check{b}(X^n)|y^{n-2}, Y_{n-1}^n). \quad (28)$$

Combining this with (27) and averaging over  $y^{n-2}$  proves  $H(\hat{b}(X^n)|Y^n) \leq H(b(X^n)|Y^n)$ . Since  $H(\hat{b}(X^n)) = H(b(X^n))$ , the proof is complete. ■

#### B. Proof of Theorem 3

The proof of Theorem 3 requires the following lemmas.

**Lemma 2.** *In order to prove (16), it is sufficient to consider  $p \in [1/2, 1]$ .*

*Proof:* Suppose  $b$  and  $b'$  are both lex and satisfy<sup>1</sup>

$$p \triangleq \mathcal{P}\{b'(X^n) = 0\} = \frac{1}{2} \mathcal{P}\{b(X^{n-1}) = 0\}. \quad (29)$$

We have the identities  $f(pq) = pf(q) + qf(p)$  and

$$\mathcal{P}\{b'(X^n) = 0|y^n\} = \mathcal{P}\{X_1 = 0|y_1\} \mathcal{P}\{b(X_2^n) = 0|y_2^n\},$$

which imply the relation:  $2T_\alpha(p) = T_\alpha(2p) + 2pH(\alpha)$ . It follows easily that

$$\frac{1}{2} [T_\alpha(2p) - f(2p)H(\alpha)] = [T_\alpha(p) - f(p)H(\alpha)]. \quad (30)$$

Thus, the claim is proved.  $\blacksquare$

Although  $T_\alpha(p)$  is not concave, we are able to prove a pseudo-concavity characteristic of  $T_\alpha(p)$ . This is exploited in the following claim.

**Lemma 3.** *For  $k \leq 2^n$ , consider the lex functions  $b_-$  and  $b_+$  which satisfy*

$$p_- \triangleq \mathcal{P}\{b_-(X^n) = 0\} = \frac{k}{2^n} \quad (31)$$

$$p_+ \triangleq \mathcal{P}\{b_+(X^n) = 0\} = \frac{k+1}{2^n}. \quad (32)$$

For  $\theta \in [0, 1]$ , the following inequality holds:

$$T_\alpha(\theta p_- + (1-\theta)p_+) \geq \theta T_\alpha(p_-) + (1-\theta)T_\alpha(p_+).$$

*Proof:* First, observe that it suffices to prove

$$T_\alpha\left(\frac{p_- + p_+}{2}\right) \geq \frac{1}{2} [T_\alpha(p_-) + T_\alpha(p_+)]. \quad (33)$$

Indeed, from (33) an inductive argument proves the lemma when  $\theta p_- + (1-\theta)p_+$  is restricted to the set of dyadic rationals. Then, recalling the continuity of  $T_\alpha(\cdot)$  on  $[0, 1]$  completes the proof.

To this end, let  $b$  be the unique lex function on  $n+1$  inputs which satisfies

$$\mathcal{P}\{b(X^{n+1}) = 0\} = \frac{2k+1}{2^{n+1}} = \frac{1}{2} [p_- + p_+]. \quad (34)$$

By construction, we have

$$\begin{aligned} \mathcal{P}\{b(X^{n+1}) = 0|Y^{n+1}\} \\ = \mathcal{P}\{X_{n+1} = 0|Y_{n+1}\} \mathcal{P}\{b_+(X_1^n) = 0|Y_1^n\} \\ + \mathcal{P}\{X_{n+1} = 1|Y_{n+1}\} \mathcal{P}\{b_-(X_1^n) = 0|Y_1^n\}. \end{aligned} \quad (35)$$

Combining (34) and (35) with the fact that  $f(x)$  is concave, we have the desired inequality

$$\begin{aligned} T_\alpha\left(\frac{p_- + p_+}{2}\right) \\ \geq \mathbb{E}_{Y^{n+1}} \mathcal{P}\{X_{n+1} = 0|Y_{n+1}\} f(\mathcal{P}\{b_+(X_1^n) = 0|Y_1^n\}) \\ + \mathbb{E}_{Y^{n+1}} \mathcal{P}\{X_{n+1} = 1|Y_{n+1}\} f(\mathcal{P}\{b_-(X_1^n) = 0|Y_1^n\}) \\ = \frac{1}{2} [T_\alpha(p_-) + T_\alpha(p_+)]. \end{aligned} \quad (36)$$

<sup>1</sup>Any lex function with  $\mathcal{P}\{b(X^n) = 0\} = k/2^n$  can be reduced to a lex function on  $n-1$  inputs if  $k$  is even.

We are now in a position to prove Theorem 3.

*Proof of Theorem 3:* Lemmas 2 and 3 imply that, in order to prove (16), it is sufficient to construct a piecewise linear function  $g : [1/2, 1] \rightarrow [0, \infty)$  satisfying the following properties:

- 1) Each segment of  $g$  is a chord connecting the points  $(p_-, T_\alpha(p_-))$  and  $(p_+, T_\alpha(p_+))$ , where  $p_-$  and  $p_+$  are of the form

$$p_- = \frac{k}{2^n}, \quad p_+ = \frac{k+1}{2^n}. \quad (37)$$

for some integers  $k, n$ .

- 2) For  $p \in [1/2, 1]$ ,  $g(p) \geq f(p)H(\alpha)$ .

By definition, Algorithm II.1 terminates only if it constructs such a function.  $\blacksquare$

#### IV. CONCLUDING REMARKS

Although Conjecture 1 remains open, we have provided substantial evidence in support of its validity. Indeed, our results suggest that Conjecture 2 is valid and we have an algorithmic proof establishing Conjecture 3 for any given value of  $\alpha$ . Any complete proof of Conjectures 1 or 2 would be of significant interest, since it would likely require new methods which may be applicable in information theory and elsewhere (e.g., in proving discrete isoperimetric inequalities).

We leave the reader with a weak form of Conjecture 1 which could provide insight. For Boolean functions  $b, b'$ , does it hold that  $I(b(X^n); b'(Y^n)) \leq 1 - H(\alpha)$ ? While this problem appears difficult in general, it is a simple exercise to show this is true when  $b(X^n)$  and  $b'(Y^n)$  are both Bernoulli(1/2). Intuitively, this should be the case for  $b, b'$  which maximize  $I(b(X^n); b'(Y^n))$ .

#### ACKNOWLEDGMENT

The authors are grateful to Abbas El Gamal, Chandra Nair, and Yeow-Khiang Chia for many helpful discussions.

This work is supported in part by the Air Force grant FA9550-10-1-0124 and by the NSF Center for Science of Information under grant agreement CCF-0939370.

#### REFERENCES

- [1] J. G. Klotz, D. Kracht, M. Bossert, and S. Schöber, "Canalizing boolean functions maximize the mutual information," *arxiv:1207.7193*, 2012.
- [2] A. Samal and S. Jain, "The regulatory network of e. coli metabolism as a boolean dynamical system exhibits both homeostasis and flexibility of response," *BMC Systems Biology*, vol. 2, no. 1, p. 21, 2008.
- [3] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *The 37th Annual Allerton Conference on Communication, Control, and Computing*, September 1999, pp. 368 – 377.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.
- [5] R. O'Donnell, "Some topics in analysis of boolean functions," in *Proc. STOC '08*. New York, NY, USA: ACM, 2008, pp. 569–578.
- [6] L. H. Harper, "Optimal numberings and isoperimetric problems on graphs," *Journal of Combinatorial Theory*, no. 1, pp. 385 – 393, 1966.
- [7] B. Bollobás and I. Leader, "Compressions and isoperimetric inequalities," *J. Combinatorial Theory, Series A*, vol. 56, no. 1, pp. 47 – 62, 1991.
- [8] P. C. Allaart and K. Kawamura, "The Takagi function: a survey," *Real Analysis Exchange*, vol. 37, no. 1, pp. 1 – 54, 2011.
- [9] C. J. Guu, "The mcfunctor," *Discrete Mathematics*, vol. 213, no. 13, pp. 163 – 167, 2000.